

PROCESSING OF ENGLISH RELATIVE CLAUSES BY ADULT L2 LEARNERS

BY

SOONDO BAEK

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Assistant Professor Tania Ionin, Chair
Associate Professor Kiel Christianson
Associate Professor Susan M. Garnsey
Professor Hye Suk James Yoon

Abstract

The ability to use a second language learned after childhood is an important aspect of the human mind. A better understanding of the characteristics of the L2 sentence processing mechanism deployed by adult L2 learners would provide valuable insights into this important human ability. The present dissertation is an attempt to draw a detailed picture of the L2 sentence processing mechanism by comparing L1 speakers and adult L2 learners in terms of the use of different sources of information and the role of working memory capacity (WMC) in the course of sentence processing. In three self-paced reading experiments with L1 speakers of English (Chapter 6) and three parallel experiments with adult L1-Korean L2-English learners (Chapter 7), the processing of subject- and object-extracted relative clauses (SRs and ORs) in English were tested with the animacy of nouns systematically manipulated. The effect of individual differences in WMC was also explored.

The results from L1 speakers corroborated previous findings (i) that ORs are more difficult to comprehend than SRs, indicating greater syntactic complexity of the former, (ii) that ORs with inanimate head nouns (i.e., objects) and animate subjects are easier to comprehend than those with animate head nouns and inanimate subjects, reflecting the role of canonical noun animacy-grammatical role associations (i.e., animate nouns are typically subjects and inanimate nouns are typically objects), and (iii) that individual differences in WMC correlate with individual differences in the ability to incorporate noun animacy information into syntactic analyses but not with individual differences in syntactic complexity effects, suggesting a dissociation between processing resources recruited for structural computation and those recruited for semantic evaluation (e.g., Traxler, Williams, Blozis, & Morris, 2005).

The behaviors of L2 learners were similar to those of L1 speakers in several respects. They found ORs more difficult than SRs, and ORs with animate subjects and inanimate head nouns easier than those with inanimate subjects and animate head nouns. The results suggest that adult L2 learners are able to take into account both syntactic and semantic information in native-like ways during online sentence processing. In addition, L2 learners showed a potential dissociation between processing resources underlying syntactic computation and semantic evaluation, further suggesting a similarity between L1 and L2 sentence processing mechanisms. Unlike L1 speakers, however, they did not benefit from animate subjects in processing ORs when head nouns were also animate, suggesting a reduced ability to revise an initially

constructed, plausible interpretation (under the assumption that an animate head noun leads to the expectation of an SR, which must be discarded when the expectation turns out to be wrong).

The results taken together suggest that L1 speakers and adult L2 learners of a language employ similar kinds of processes and processing resources to comprehend sentences in the language, but L2 learners experience greater difficulty revising the currently built structure to which they have made a commitment due to its plausible interpretation. Implications of these findings are discussed for the theoretical understanding of L2 sentence processing by adult L2 learners.

to Do Hong Kim, my mom

Acknowledgements

I would like to give my deepest thanks to my advisor Tania Ionin. Her valuable advices and encouragements made possible much of what I could do in Illinois, including this dissertation. I also truly thank my committee members, James Yoon, Kiel Christianson and Susan Garnsey for their helpful comments on the dissertation and kind answers to my not-so-well articulated questions. I owe special thanks to Prof. James Yoon for giving me the opportunity to study at UIUC and supporting my graduate life in many important ways.

I owe much to other faculty members at UIUC. I especially thank Cindy Fisher for sharing so much of her time with me to discuss many things in psycholinguistics and for allowing me to work as her research assistant. I am also indebted to Silvina Montrul, who supervised one of my qualifying papers and allowed me to work for the SLATE program. I wish to thank Annie Tremblay and Duane Watson as well. Most of what I learned about L1 and L2 sentence processing came from their classes. I much enjoyed working on experimental phonology with Jennifer Cole, to whom I would like to extend my special thanks.

I wish to take the opportunity to thank Professor Oryang Kwon, who supervised my master's thesis in Korea and officiated my wedding. Both my academic and marital lives began with him.

I am grateful to my friends with whom I had the luck to share not a little time at Urbana and Champaign: Tae-jin Yoon, Young-Il Oh, Matt Garley, Ben Slade, Hee Youn Cho, Keun Young Shin, Young Sun Lee, Theeraporn Ratitamkul, Yoonsook Mo, Ju Hyeon Hwang and her husband Jungman Park, and Donghee Om and his wife Eunkyung Lee. They all greatly helped me and Eunah transition to Urbana and live through the graduate life.

My wholehearted thanks and love go to my wife Eunah Kim for her constant support and love. She has been my best friend and colleague who went through both good and bad days together at Urbana. My deepest gratitude goes to my parents, who made possible what I have been and will be, and what I have done and will do.

Finally I would like to thank National Science Foundation for providing me with a doctoral dissertation research grant (Award ID: 1021455, PI: Tania Ionin), which allowed me to run experiments for the present study and attend conferences for presentation of the results.

Table of Contents

Chapter 1 Introduction	1
Chapter 2 Sentence processing in native and second language	5
2.1. Sentence processing in native language	5
2.2. Issues and previous findings on second language sentence processing	9
Chapter 3 Working memory capacity in sentence processing	29
3.1. Measurements of WMC	30
3.2. Different views on the nature of cognitive resources underlying sentence processing	32
3.3. Previous findings on the role of WMC in second language sentence processing	36
Chapter 4 Previous findings on relative clause processing in L1 and L2	40
4.1. Relative clause processing in the native language: experimental findings and theoretical accounts	40
4.2. Processing relative clauses in L2	54
Chapter 5 Overview of the present study	60
5.1. Overview of the experiments	60
5.2. Results of WM span tests	68
Chapter 6 Experiments with native speakers of English	73
6.1. Experiment 1E	73
6.2. Experiment 2E	77
6.3. Experiment 3E	90
6.4. General discussion of L1 processing of English RCs	101
Chapter 7 Experiments with L2 learners of English	107
7.1. Experiment 1K	107
7.2. Experiment 2K	113
7.3. Experiment 3K	126
7.4. General discussion of L2 processing of English RCs	137
Chapter 8 Conclusion and future research	143
Appendix A Test sentences	148
Appendix B Cloze test	158
References	159

Chapter 1

Introduction

The ability to learn and use a second language (L2) is an important part of human cognitive ability and everyday life. It therefore has been one of the major focuses not only in pedagogical interests to help people learn a new language but also in scientific efforts to understand the nature of the human mind. Research on L2 acquisition offers valuable insights into the underlying forces that constrain the learning of an L2. More recent developments in research on L2 sentence processing complement the traditional study of L2 acquisition by adding evidence concerning how an L2 is put to use in real time.

A major issue in the L2 acquisition literature is whether the learning of an L2 is constrained by the same principles as those driving the acquisition of the native language (L1) and whether it ultimately results in the same kind of linguistic knowledge as the L1 (e.g., Schwartz, 1998; Bley-Vroman, 1989, 2009). In a similar vein, research on L2 sentence processing revolves around the question of whether it involves the same kinds of mental representations and processing mechanisms as those underlying L1 sentence processing. The main goal of this dissertation is to contribute to a better understanding of this issue.

Sentence processing is conceptualized as a set of dynamic processes through which various kinds of information are incorporated rapidly and efficiently to interpret a sentence as it unfolds. Words in a sentence are incrementally integrated based on their semantic properties and syntactic relationships with other words, leading to the extraction of the meaning of the sentence, which is also moderated by the comprehender's general knowledge about the world. Psycholinguistic study on human sentence processing investigates how these various sources of information constrain the real-time interpretation of sentences in order to model the mechanisms of the human ability to comprehend sentences.

From the perspective of L2 sentence processing research, L2 learners' sensitivity to the different kinds of constraints that have been identified in the L1 sentence processing literature could be taken as evidence for whether and how much L2 sentence processing is similar to or different from L1 sentence processing (e.g., Clahsen & Felser, 2006; Frenck-Mestre, 2002). Can L2 learners exploit the syntactic structure of a sentence in the L2 as efficiently as native speakers do? Or does some kind of domain-specific deficit in one or more aspects of the formal linguistic knowledge in the L2 lead to a greater reliance on extra-linguistic strategies and general world

knowledge as compared to native speakers of the language? The present dissertation addresses these questions by comparing L2 learners and L1 speakers in terms of their sensitivity to structural vs. non-structural (lexical-semantic) information during online sentence processing.

Human sentence processing has been shown to be incremental, with newly incoming input (e.g., a word) immediately interpreted in relation to the previous input (e.g., Just & Carpenter, 1980; Marslen-Wilson, 1973). Incremental processing of a sentence requires dynamic cognitive processes whereby incoming input is integrated with the product of previous computation carried out on the preceding input. Working memory (WM) is conceptualized as the locus of the maintenance and computation of various kinds of information required to interpret a sentence (e.g., Just & Carpenter, 1992; Waters & Caplan, 1996a). Individual differences in the availability of WM resources, which is often referred to as working memory capacity (WMC), have attracted researchers in psycholinguistics as a possible indicator of individual differences in sentence processing and in other cognitive behaviors. In the same light, the role of WMC in L2 sentence processing might provide important insights into individual differences in L2 sentence processing. Besides, it could also serve as a dimension along which L1 and L2 sentence processing can be compared in terms of the characteristics of underlying processing resources. The present dissertation pursues these possibilities by comparing the role of WMC in L1 and L2 sentence processing.

Relative clauses (RCs) appear to provide a rich test bed on which L1 and L2 sentence processing can be compared with each other. The sentences in (1) and (2) show two types of RCs enclosed in brackets, which differ in the understood grammatical roles of the modified nouns in relation to the verbs within the RCs. In (1), the noun phrase *the reporter* is the understood subject of the RC verb *attacked*, whereas in (2), the same noun phrase is the understood object of the RC verb. Accordingly, the RCs in these two sentences are called subject-extracted RC (SR) and object-extracted RC (OR), respectively.

- (1) The reporter [that attacked the senator] admitted the error in the public.
- (2) The reporter [that the senator attacked] admitted the error in the public.

One of the most reliable findings in the sentence processing literature is that ORs are more difficult to comprehend than SRs (e.g., King & Just, 1991; Wanner & Maratsos, 1978). The

processing difficulty involved with ORs, however, seems to be modulated by the semantic features of the relevant nouns (e.g., Gordon, Hendrick, & Johnson, 2001, 2004; Traxler, Morris, & Seely, 2002; Traxler, Williams, Blozis, & Morris, 2005). Animate entities are more likely to do an action than inanimate entities in real-life events and thus animate nouns are more likely to assume an agent/subject role than inanimate entities in sentences. When the animacy configurations of the nouns are consistent with this tendency as in (4) (i.e., an animate subject and an inanimate object), the usual processing difficulty involved with the OR is reduced as compared to when animacy information conflicts with the canonical animacy–semantic/syntactic role associations as shown in (3) (i.e., an inanimate subject and an animate object) (e.g., Traxler et al., 2002, 2005). Capitalizing on these findings in L1 sentence processing, we can test whether L2 learners are sensitive to these different kinds of information (i.e., syntactic structure and noun animacy) to a similar extent as L1 speakers as a way to compare sentence processing behaviors in the two groups of language users (e.g., Felser, Cunnings, Batterham, & Clahsen, in press; Omaki & Arijji, 2005).

(3) The director that the movie pleased received a prize at the film festival.

(4) The movie that the director watched received a prize at the film festival.

In addition, WMC seems to be related to RC processing in subtle ways. The results of the experiments on English speakers' RC processing in Traxler et al. (2005) showed that the ability to cope with different animacy configurations was modulated by the comprehender's WMC, whereas processing difficulty due to syntactic complexity was not. The results suggest a potential dissociation in the processing mechanisms underlying syntactic vs. semantic-pragmatic computation. If we see a similar distinction in L2 sentence processing, it could be taken as evidence for a qualitative similarity between L1 and L2 sentence processing in terms of the nature of underlying processing resources.

In summary, the present dissertation capitalizes on two major findings about RC processing in native speakers of English in order to investigate the corresponding aspects of L2 sentence processing. One finding concerns the roles of different kinds of information in determining the processing difficulty of ORs as compared to SRs. The other is a potential dissociation in the processing mechanisms responsible for syntactic computation and semantic-

pragmatic evaluation as revealed in the relationships between WMC and behavioral signs of processing difficulty.

To preview the findings of the present dissertation, adult L2 learners showed both similarities and differences compared to L1 speakers in sentence processing behaviors. L1-Korean L2-English learners processed different kinds of information available from English RCs in largely the same way as L1 speakers of the target language, and the two groups of speakers both showed evidence for a dissociation between the cognitive resources sub-serving the processing of syntactic information and those responsible for the processing of lexical-semantic information. Unlike native speakers, however, L2 learners experienced greater difficulty when a plausible initial interpretation of a sentence fragment must be revised to reach a correct final interpretation. The overall results suggest that L2 learners process sentences in the target language using quite similar mechanisms as those used by native speakers of the language, but the former are less capable than the latter of recovering from plausible initial analyses, reflecting limited processing efficiency and flexibility.

This dissertation is organized as follows.

Chapter 2 introduces major theories of L1 sentence processing, which are followed by major issues and previous findings in L2 sentence processing.

Chapter 3 discusses the issues related to the role of WM resources in L1 and L2 sentence processing. Various measurements of WMC are introduced and different views on its role in sentence processing are discussed. Previous findings on the role of individual differences in WMC in L2 sentence processing are then reviewed.

Chapter 4 presents previous findings on the processing of RCs in L1 and L2. Various accounts for RC processing are reviewed and discussed to establish a theoretical stance of the present dissertation and to elaborate predictions for the present experiments.

Chapter 5 presents a general overview of the present study. Tasks and overall procedures are presented. The results of WM span tests are reported.

Chapter 6 reports three experiments with L1 speakers of English.

Chapter 7 reports three parallel experiments with adult L1-Korean L2-English learners.

Chapter 8 concludes the present dissertation with a summary of the findings and suggestions for future research.

Chapter 2

Sentence processing in native and second language

This chapter begins with a brief discussion in Section 2.1 of major theories of sentence processing entertained in the psycholinguistics literature. In Section 2.2, issues and previous findings are reviewed on adult L2 learners' sentence processing in the L2, with a focus on the debate on the similarities and differences between L1 and L2 sentence processing.

2.1. Sentence processing in native language

Sentence processing refers to a cognitive event in which the human comprehender turns a perceived string of words in a sentence into a structured mental representation, and recovers the intended meaning of the sentence, as it unfolds in real time. Psycholinguistic research on sentence processing has identified several kinds of constraints that guide the online interpretation of sentences, including phrase-structure formation rules, lexical properties of words (e.g., semantic contents and structural requirements encoded in words), sentential and discourse contexts, frequency with which certain words or structures occur within certain environments, and limits on computational resources available to human comprehenders (Gibson & Pearlmuter, 1998).

Different theories have been proposed to explain how the different constraints affect online sentence processing. While most theories agree that human sentence processing is incremental in the sense that important decisions regarding the interpretation of a sentence are made as soon as relevant input is processed (i.e., it does not wait until the entire sentence is available), they disagree on the issue of which constraints take priority and how they interact with other constraints (see Pickering & van Gompel, 2006, for a review).

Incremental sentence processing implies that the human sentence processor tries to construct whatever interpretable representations from a set of words currently available and thus needs to continuously make decisions about how to integrate a newly incoming word into the currently built representation. A consequence of this incrementality is that the comprehender frequently encounters a situation where there are multiple ways to integrate new input into the structure built thus far. Major theories of sentence processing have developed as attempts to explain how the human sentence processor copes with such structural ambiguities arising during sentence processing.

Some influential views on ambiguity resolution prioritize structural considerations in making an initial parsing decision between multiple possible analyses. These views are represented by the garden-path model, according to which a limited number of processing principles or strategies guide initial decision making for assigning a syntactic tree to a given string of words (e.g., Clifton & Frazier, 1989; Clifton, Traxler, Mohamed, Williams, Morris, & Rayner, 2003; Ferreira & Clifton, 1986; Frazier, 1979, 1987a, 1987b; Frazier & Rayner, 1982; Pickering & Traxler, 1998; Rayner, Carlson, & Frazier, 1983). The processing strategies proposed in the garden-path model are *structure-based* in the sense that they operate exclusively based on the relative processing simplicity of alternative structures consistent with the current input, ignoring other kinds of potentially relevant information such as the plausibility of the resulting interpretation. Specifically, the Minimal Attachment principle dictates that among multiple possible syntactic trees, the one with the fewest syntactic nodes be selected, and the Late Closure principle dictates that the incoming input be attached to the constituent that has been processed most recently (e.g., Frazier, 1987a). The Active Filler strategy dictates that dislocated constituents such as a fronted wh-phrase be integrated into the current parse as soon as grammatically possible (e.g., Frazier, 1987b).

In the garden-path model, when an initial syntactic analysis is not tenable due to the information available in later stages of processing, it is revised to an alternative analysis to incorporate the new information (e.g., Ferreira & Clifton, 1986; Rayner et al., 1983). This process of reanalysis causes processing complexity as reflected in increased reading times and more frequent errors in final comprehension (when the reanalysis fails). Thus, the garden-path model assumes serial processing in which structural information is processed independently from semantic and pragmatic information that is available from lexical items and general knowledge about the world, at least in the initial stages of processing (Frazier, 1987a; cf. Fodor, 1985).

In contrast to the garden-path model, a group of theories that can be broadly categorized as constraint satisfaction models do not endorse the categorical priority of purely structural information and instead take the dynamic interaction of different kinds of information as the mechanism for drawing the most likely initial syntactic analysis of an input string (e.g., Garnsey, Pearlmutter, & Lotocky, 1997; MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, Tanenhaus, & Garnsey, 1994). Relevant sources of information include thematic fit between verbs and their arguments (e.g., Trueswell et al., 1994), discourse contexts (e.g., Altmann &

Steedman, 1988), relative frequency of alternative structures with which particular verbs are used (e.g., Garnsey et al, 1997; Trueswell, 1996) and relative frequency of different constructions (e.g., MacDonald, 1999). In the majority of these constraint-based models, multiple possible syntactic analyses are activated simultaneously with different strengths depending on their degree of fit with the currently available information. When the most strongly activated analysis at a given moment turns out to be incorrect later, the parser should activate an alternative analysis to a higher level, which incurs processing difficulty.

As an illustration of crucial differences between the garden-path and the constraint satisfaction models, consider the sentences in (5) and (6), which contain much studied type of temporary syntactic ambiguity (e.g., Ferreira & Clifton, 1986; Trueswell et al., 1984).

(5) The defendant examined by the lawyer turned out to be unreliable.

(6) The evidence examined by the lawyer turned out to be unreliable.

The two sentences differ only in their subject noun (i.e., *defendant* vs. *evidence*) and both contain a reduced relative clause (RC) *examined by the lawyer* that modifies the subject of the main clause, *the defendant* in (5) and *the evidence* in (6). The RCs in (5) and (6) are ‘reduced’ in the sense that they lack a relative pronoun and an auxiliary verb that would be expressed in the unreduced versions (e.g., *that was*). These reduced RCs lead to a temporary syntactic ambiguity because the past participle *examined* can be locally interpreted as part of the main predicate of the sentence until the following prepositional phrase *by the lawyer* makes it clear that it is a past participle comprising an RC.

The garden-path model predicts that the comprehender will initially analyze the past participle of the reduced RC as the main clause past tense verb in both (5) and (6), because it is a simpler analysis in terms of syntactic configuration than the correct reduced RC analysis (i.e., due to Minimal Attachment). The following prepositional phrase *by the lawyer* then forces a revision of the initial analysis, inducing processing difficulty. Crucially, the semantic implausibility of *evidence* examining something in (6) does not prevent the processor from being misled to the wrong main verb analysis because of the serial nature of sentence processing, resulting in comparable processing difficulty between (5) and (6) in the initial stages of processing. Processing difficulty associated with the later reanalysis, however, may vary

depending on the plausibility of the initial analysis, resulting in greater difficulty reanalyzing a plausible initial analysis in (5) compared to an implausible initial analysis in (6) (e.g., Clifton et al, 2003; Ferreira & Clifton 1986; Pickering and Traxler, 1998).

The prediction of the constraint satisfaction models is consistent with that of the garden-path model for (5) because both syntactic and semantic constraints initially promote the main verb analysis (i.e., The main verb analysis is more frequent than the reduced RC analysis; It is plausible for a human being to examine something). But in contrast with the garden-path model, the constraint satisfaction models do not predict unusual processing difficulty for (6) because the implausibility of the inanimate entity *evidence* examining something makes the main verb analysis unlikely and promotes the reduced RC analysis in which *evidence* is the theme of the examining action (e.g., Trueswell et al., 1994).

Rather different types of sentence processing theories have also been proposed to explain processing complexity that is not readily attributable to the presence of syntactic ambiguity. These theories mostly focus on the limits and properties of human memory resources responsible for sentence processing. For example, Gibson's (1998, 2000) memory-based processing complexity theory (i.e., Dependency Locality Theory) identifies two kinds of memory costs that lead to processing complexity. One is *storage cost* that varies as a function of the number of incomplete syntactic dependencies at a given moment, and the other is *integration cost* that increases with the distance between two syntactic sites to be integrated (e.g., a verb and its dependent). Processing complexity at a certain point in a sentence is hypothesized to increase with the sum of memory cost from these two sources, under the assumption that both storage and integration access the same limited pool of processing resources (e.g., Just & Carpenter, 1992).

Another group of memory-based sentence processing theories finds a major source of processing difficulty in interference between similar items activated in memory (e.g., Gordon et al., 2001, 2004; McElree, Foraker, & Dyer, 2003; Van Dyke & McElree, 2006). Interference occurs when there is intervening material between two non-adjacent syntactic sites to be integrated that has similar features to the preceding site. The dependency between two non-adjacent syntactic sites to be processed, the preceding site must be retrieved from memory to be integrated with the later coming site when the latter is encountered. At the point of retrieval, similar elements in memory are also activated, resulting in interference. For example, the OR *the reporter that the senator attacked* would be more difficult to process than the SR *the reporter*

that attacked the senator, because when the RC head noun phrase *the reporter* should be retrieved from memory and attached to the RC verb *attacked*, a similar noun phrase *the senator* intervenes between the RC verb and the RC head in the OR while there is no intervening material in the SR.

The sentence processing models described above provide rich theoretical backgrounds on which to compare L1 and L2 sentence processing as they identify and highlight different aspects of sentence processing. Different theories still compete for an adequate explanation of sentence processing and most of the behavioral data currently available on L2 sentence processing, including the findings of the present dissertation, seem to be readily accommodated by multiple models with slight modifications in their assumptions and perspectives. The present dissertation, therefore, was not intended to evaluate different theories in any serious way, but instead capitalizes on the insights offered by them. For the purposes of discussion, however, I will frequently refer to the terms and concepts derived from the garden-path model, because one of the most elaborated accounts for RC processing has been developed primarily based on the garden-path model (e.g., Traxler et al., 2005), and because a clear distinction between the roles of structural and non-structural information in the model allows a more vivid illustration of the similarities and differences between L1 and L2 processing, as will be discussed in the next section.

2.2. Issues and previous findings on second language sentence processing

In a broader context of L2 acquisition research, L2 processing in general has often been the object of interest as a constraint on the acquisition of linguistic knowledge in the L2 (e.g., Carroll, 1999; Pienemann, 1998; Truscott & Sharwood-Smith, 2004; Van Patten, 2002). Increasing interest in the psychological processes involved in L2 processing in and of themselves, however, has turned the attention of many researchers to the nature of representations and processes involved in L2 sentence processing. Several researchers submit that evidence from studies on L2 sentence processing may provide insights into the nature of L2 grammatical knowledge in ways that the traditional approaches to L2 acquisition research cannot (e.g., Clahsen & Felser, 2006; Juffs, 2004; Shacter & Yip, 1990). For example, an apparent deficit in grammatical knowledge as reflected in L2 learners' degraded performance in *offline* grammaticality judgments may be attributed to L2 learners' overall inefficiency in *online*

grammatical processing rather than some kind of genuine deficit in grammatical representations in the L2 (e.g., Juffs & Harrington, 1995; McDonald, 2006; Shacter & Yip, 1990).

The leading issue in the L2 sentence processing literature is whether adult L2 learners process sentences in the L2 in the same way as native speakers of the language do (e.g., Clahsen & Felser, 2006, 2009; Dallas & Kaan, 2008; Dussias & Pinar, 2009; Frenck-Mestre, 2002; Papadopoulou, 2005). This issue echoes the long-standing question in L2 acquisition research as to whether L2 learners past the critical period in language acquisition are constrained by the same principles as normally developing child L1 learners (e.g., Bley-Vroman, 1989, 2009; Schwartz, 1992, 1998). Accordingly, research on L2 sentence processing mainly concerns how adult L2 learners compare to native speakers in various aspects of online sentence processing.

Previous findings seem to generally suggest that adult L2 learners with a sufficient level of proficiency in the target language have no particular difficulty in incorporating lexically-based information and simple word-order information in native-like ways, whereas the findings concerning the use of complex structural information and syntactically-relevant morphological information are so far mixed (see for example, Dussias & Piñar, 2009; Papadopoulou, 2005, for reviews of the relevant findings). The divergent findings on the (morpho-)syntactic aspects of L2 sentence processing motivate two major contrasting views. Some researchers hold the view that there are some ‘qualitative’ differences between the L1 and L2 sentence processing mechanisms, which cannot be reduced to ‘quantitative’ factors such as L2 learners’ lower proficiency, slower processing speed and reduced experience in the target language (e.g., Clahsen & Felser, 2006; Felser, Cunnings, Batterham, & Clahsen, in press). This idea is represented by the shallow structure hypothesis (SSH) as proposed by Clahsen & Felser (2006) that adult L2 learners differ from native speakers in that they cannot construct as detailed syntactic representations as native speakers during online sentence processing, lacking syntactic markers such as phonologically null categories and hierarchical relations between constituents. This syntactic shallowness is putatively due to the characteristic of the L2 syntactic knowledge which is not immediately available in the early stages of sentence processing. As a result, adult L2 learners have to rely more on non-structural information such as lexical-semantic and pragmatic information available from (chunks of) lexical items, which are taken to be readily accessible during online sentence processing in the L2. Other researchers, on the other hand, propose that the syntactic components of L2 sentence processing can also become native-like as the L2 comprehender’s proficiency in

and exposure to the target language increase (e.g., Dussias & Sagarra, 2007; Frenck-Mestre, 2002; Rah & Adone, 2010). According to these researchers, the mechanisms underlying L2 sentence processing are qualitatively the same as those responsible for L1 sentence processing, and the apparent differences between them mostly come from L2 learners' limited processing efficiency and transfer of their L1 processing strategies to L2 sentence processing.

In the next two sections, previous findings are reviewed on adult L2 learners' use of structural vs. non-structural information during sentence processing in the L2, focusing on the debate concerning the similarities and differences between L1 and L2 sentence processing.

2.2.1. Referring to non-structural information in L2 sentence processing

Several studies suggest that adult L2 learners' commitment to the current syntactic analysis is modulated by the plausibility of the resulting interpretation (e.g., Dussias & Piñar, 2010; Felser & Roberts, 2011; Frenck-Mestre & Pynte, 1997; Williams, 2006; Williams, Möbius, & Kim, 2001). In a self-paced reading study by Felser & Roberts (2011), for example, advanced L1-Greek L2-English learners and native English speakers read sentences such as (7) and (8).

(7) The inspector warned the boss would destroy very many lives.

(8) The inspector warned the crimes would destroy very many lives.

The main clause verb *warned* can take either an NP or a sentence as its complement. Thus the following NPs, *the boss* in (7) and *the crimes* in (8), are temporarily ambiguous between the object of the preceding verb and the subject of the following sentence complement. This ambiguity is resolved (i.e., disambiguated) to the subject analysis when the auxiliary verb *would* is encountered in (7) and (8). What is crucial is that *the boss* in (7) is a plausible object of *warned* (i.e., a boss can readily be warned about something), whereas *the crimes* in (8) is not (i.e., crimes cannot be warned).

The results of Felser & Roberts (2011) showed that L2 learners experienced increased difficulty reading the ambiguous post-verbal NP when it was implausible as the object of the preceding verb, as reflected in the longer reading time at *crimes* in (8) as compared with *boss* in (7). This effect spilled over to the following auxiliary *would*. But the reverse pattern was observed at the embedded clause verb *destroy* and the following word *very*, with greater

difficulty in (7) where the ambiguous NP was a plausible object of the preceding verb than in (8) where the ambiguous NP was an implausible object of the preceding verb. The results suggest (i) that the L2 learners initially took the ambiguous NP as the direct object of the preceding verb, as reflected in longer reading times for implausible analyses (i.e., an implausible analysis takes longer to integrate than a plausible one), and (ii) that they later revised the initial analysis to incorporate the newly incoming information, as evidenced by longer reading times for the sentences with a plausible initial analysis as compared to those with an implausible one (i.e., a plausible analysis is harder to revise than an implausible analysis). Based on similar findings on L1 sentence processing (e.g., Frazier & Rayner, 1982; Pickering & Traxler, 1998), the authors interpreted the results as indicating that adult L2 learners incrementally incorporate plausibility information during sentence processing in the same way as native speakers.

A similar pattern of plausibility effects has also been found in adult L2 learners' processing of wh-questions in English (e.g., Dussias & Piñar, 2010; Williams, 2006; Williams et al., 2001; see Boland, Tanenhaus, Garnsey, & Carlson, 1995; Traxler & Pickering, 1996, for findings on native speakers). In typical English wh-questions, a fronted wh-phrase must be integrated with a verb that takes the wh-phrase as its argument or adjunct. In the L1 sentence processing literature, the relationship between a dislocated argument and its syntactic/thematic licenser is described as a filler-gap dependency because figuratively speaking, its processing involves filling the dislocated argument into a gap postulated at the position where it would appear in canonical declarative sentences. In Williams et al.'s (2001) experiments, English native speakers and adult L2 learners of English whose L1 was Chinese, German or Korean read sentences such as (9) and (10) on a word-by-word basis and pressed a button when they thought the sentences no longer made sense.

(9) Which girl did the man push the bike into late last night?

(10) Which river did the man push the bike into late last night?

The wh-phrases in (9) and (10) both must be analyzed as the object of the preposition *into* for the correct interpretation of the sentences. The crucial difference between the two sentences was that *which girl* in (9) was a plausible object of the verb *push* but *which river* in (10) was not. The L2 learners in the Williams et al. study pressed the button to indicate the implausibility of

sentences at the verb (e.g., *push*) more often when the wh-phrase was implausible as the verb's object as in (10) than when it was plausible as in (9), suggesting that they tried to immediately integrate the wh-phrase and the verb. The converse pattern was observed at the head noun of the overt object NP (e.g., *bike*), with increased implausibility detection rates for (9) than for (10). The result suggests that the L2 learners experienced greater difficulty revising the initial analysis when it was plausible than when it was implausible. The overall pattern of results was quite similar to that from the English-speaking control group. At the same time, however, the study also showed that in an offline comprehension task, the L2 learners often failed to revise the plausible initial association of the wh-phrase and the main verb. Based on these findings, the authors proposed that adult L2 learners, like native speakers, interpret a dislocated argument as soon as possible and are sensitive to plausibility information during online sentence processing, but they experience increased difficulty in reanalyzing an initially plausible analysis compared with native speakers (cf. Roberts & Felser, 2011).

Another type of non-structural information to which adult L2 learners have been shown to be sensitive during online sentence processing is the verb-specific properties about possible argument structures (i.e., number and syntactic categories of required arguments) and the probabilistic bias to one of the alternative argument structures (e.g., Dussias & Cramer Scaltz, 2008; Juffs, 1998, 2004; Frenck-Mestre & Pynte, 1997; Rah & Adone, 2010; cf. MacDonald, Pearlmutter, & Seidenberg, 1994, for discussions of native speakers' use of such information). In Dussias and Cramer Scaltz's (2008) self-paced reading study, for example, adult L1-Spanish L2-English learners read sentences such as (11) and (12) among others, in which the grammatical role of the NPs *the rumor* and *the mistake* was ambiguous between the object of the main clause verb and the subject of the embedded clause. This ambiguity is resolved to the object analysis at *when* in (11a) and (12a) and to the subject analysis at *could* and *might* in (11b) and (12b), respectively.

(11) a. The CIA director confirmed the rumor when he testified before Congress.

b. The CIA director confirmed the rumor could mean a security leak.

(12) a. The ticket agent admitted the mistake when he got caught.

b. The ticket agent admitted the mistake might not have been caught.

The crucial manipulation was that the main clause verb *confirmed* in (11) typically takes an NP complement while the verb *admitted* in (12) takes a sentence complement more frequently, as has been demonstrated by sentence completion tasks with native speakers (e.g., Wilson & Garnsey, 2009) and with L2 learners in the Dussias & Cramer Scaltz study. The reading time results showed that the L2 learners took longer to read the disambiguation region (i.e., *when he* in (11a) and (12a), *could mean/might not* in (11b) and (12b)) when the verb-specific bias and the actual continuation did not match as in (11b) and (12a), relative to when they did match as in (11a) and (12b), although this effect was restricted to the items for which the learners showed some knowledge about statistical biases of the verb in the sentence completion task administered after the self-paced reading task. The authors interpreted the results as indicating that the L2 learners initially adopted the more frequent argument structure for each verb, taking *the rumor* as the direct object of the verb *confirmed* in (11) and *the mistake* as the subject of the sentence complement of the verb *admitted* in (12), and they in turn experienced increased difficulty when the initial analysis turned out to be wrong and should be revised in (11b) and (12a).

Overall, the studies reviewed above generally suggest that adult L2 learners incrementally take into account various types of lexically-based information to keep the sentential representation updated according to the newly incoming input, though with less success in final interpretation compared with native speakers probably due to reduced processing efficiency. In the next section, I turn to syntactic aspects of L2 sentence processing upon which previous studies provided mixed results regarding the similarities and differences between L1 and L2 sentence processing.

2.2.2. Referring to structural information in L2 sentence processing

The resolution of syntactic ambiguity concerning which of the two potential head nouns is modified by the following relative clause (RC) has been extensively studied not only as a potential locus of cross-linguistic differences in the L1 sentence processing literature (e.g., Cuetos & Mitchell, 1988; Gibson, Schütze, & Salomon, 1996, among many others) but also as a potential test bed for the transfer of L1 processing routines to L2 processing (e.g., Dussias, 2003; Feler, Roberts, Marinis, & Gross, 2003; Fernández, 2002; Kim, 2008; Omaki, 2005; Papadopoulou & Clahsen, 2003). In (13a), for example, the complex NP *the servant of the actress* is followed by the RC *who was on the balcony*. The interpretation of the RC is

ambiguous because it can modify either the first NP *the servant* (i.e., ‘the servant was on the balcony’: high attachment) or the second NP *the actress* (i.e., ‘the actress was on the balcony’: low attachment). This attachment ambiguity can be resolved within the sentence by, for example, number agreement between the target head NP and the RC verb, which forces a high attachment in (13b) and a low attachment in (13c).

- (13) a. Someone shot the servant of the actress who was on the balcony.
b. Someone shot the servants of the actress who were on the balcony.
c. Someone shot the servant of the actresses who were on the balcony.

A large set of previous experiments tested how native speakers of diverse languages resolve the RC attachment ambiguity, using both offline questionnaires probing the final interpretations of sentences such as (13a) (e.g., ‘Who was on the balcony?’) and online reading time measures reflecting attachment preferences in sentences such as (13b) and (13c) (e.g., Carreiras & Clifton, 1999; Cuetos & Mitchell, 1988, among many others; see for example, Papadopoulou & Clahsen 2003, for a review). The overall results suggest that speakers of some languages prefer high attachment (e.g., Spanish, Greek, French, German and Dutch), while speakers of other languages prefer low attachment (e.g., English, Swedish, Norwegian, and Romanian), though with some contradictory reading time evidence as to the attachment preference in English (e.g., Felser, Marinis, & Clahsen, 2003; Traxler, 2007). This cross-linguistic difference in general attachment preferences has been attributed to various factors such as the frequency distribution of each type of attachment in language input (e.g., Mitchell, Cuetos, Corley, & Brysbaert, 1995), parametric rankings of a language-independent preference for more local attachment and a language-dependent preference for an attachment site that is structurally closer to the head of the predicate phrase (e.g., Gibson et al., 1996), and the general tendency of modifiers to attach to more salient referents in discourse (e.g., the syntactic head of the complex NP) which can sometimes be suppressed, for example in English, by the availability of an alternative unambiguous expression (e.g., the actress’s servant) (e.g., Frazier & Clifton, 1996).

Evidence for whether L2 learners ever transfer the attachment preference in the L1 to the L2 is so far inconclusive, and selected sets of experiments are taken to be evidence for different views on the nature of the L2 sentence processing mechanism. For example, a set of experiments

did not find any reliable preference for either high or low attachment in advanced L2 learners with sentences such as (13b) and (13c) in English and their equivalents in Greek, regardless of the convergence (or divergence) of the preference patterns shown by native speakers of the learners' L1 and L2 (e.g., Felser, Roberts et al., 2003; Kim, 2008; Papadopoulou & Clahsen, 2003). Adopting the account of Gibson et al. (1996), Felser, Roberts et al. (2003; Papadopoulou and Clahsen, 2003) interpret the lack of preference as indicating that adult L2 learners are not sensitive to structural recency or predicate proximity because they do not project detailed syntactic relations among constituents during sentence processing. In contrast to the insensitivity to the structural relations, L2 learners show a clear preference for low attachment when the two NPs within a complex NP are connected by a lexically meaningful preposition (e.g., *the actress with the servant who always smiles*), just like native speakers (e.g., Felser, Roberts et al., 2003; Papadopoulou & Clahsen, 2003; see for example, Gilboy, Sopena, Clifton, & Frazier, 1995, for native speaker results). This finding, together with the lack of reliable attachment preferences with NPs connected by a semantically light preposition *of*¹, has been taken by Clahsen and Felser (2006) as evidence for adult L2 learners' over-reliance on lexical-semantic information, which is a central claim of the shallow structure hypothesis (SSH).

Other studies on RC attachment ambiguity resolution in L2 suggest a rather different story, however. Some sets of previous experiments suggest that adult learners with limited exposure to the L2 do transfer the attachment preference in their L1 to the L2 while those with extensive exposure to the L2 may adopt the native-like preference pattern in the L2, lending support to the exposure-based view of L2 sentence processing (e.g., Dussias & Sagarra, 2007; Frenck-Mestre, 2002; cf. Dussias, 2003, on evidence for transfer from the more dominant L2 to the less dominant L1 in proficient bilinguals). Besides, the results of Omaki (2005) raise the possibility that the apparent lack of preference in ambiguity resolution in L2 learners could be due to averaging across individuals who show either L1- or L2-like preference. Finally, more recent studies on L1 sentence processing suggest that attachment preferences could be determined by how RCs are prosodically phrased (implicitly in case of reading) in relation to the head NPs, with variations in phrasing attributed either to cross-linguistic differences in syntax-

¹ The preposition 'of' in the complex NP *the servant of the actress* can be said to be semantically light in the sense that it does not assign a thematic role on its own, only assigning a case to the following NP. Here the NP *the servant* assigns a thematic role to its argument NP *the actress*. On the other hand, the preposition 'with' is assumed to assign its own thematic role to its argument NP, constituting an adjunct phrase.

prosody mappings (e.g., Fodor, 2002) or to individual differences in working memory capacity (WMC) (e.g., Felser, Marinis et al., 2003; Swets, Desmet, Hambrick, & Ferreira, 2007). If prosody really matters in RC attachment, then adult L2 learners' failure to show any preference in RC attachment in the L2 could be attributed to variable prosodic representations rather than to some deficit in syntactic representations. In the same light, if individual differences in WMC are associated with the variability in RC attachment, the L1-L2 difference in RC attachment preference patterns might be attributed to relatively limited WMC of L2 learners. These considerations undermine an important assumption of the SSH that RC attachment in L1 is guided by syntactically-based processing strategies.

Another type of evidence for the SSH concerns the psychological reality of phonologically null syntactic categories in L1 sentence processing (e.g., Bever & McElree, 1988; Nicol & Swinney, 1989; Roberts, Marinis, Felser, & Clahsen, 2007)² and the lack thereof in L2 sentence processing by adult L2 learners (Marinis, Roberts, Felser, & Clahsen, 2005; Felser & Roberts, 2007). Evidence comes from two studies on L2 processing of filler-gap dependency that used different methodologies. First, Felser and Roberts (2007) tested the reactivation of a preceding filler at the gap position with native English speakers and adult L1-Greek L2-English learners using a cross-modal priming task. In the task, participants heard sentences such as (14), which contained an RC whose head noun served as the indirect object of the RC verb. While the participants were hearing an experimental sentence, a picture was presented either at the gap position marked as [2] in (14), where a prepositional indirect object would canonically appear in English, or at the position 500ms earlier than the gap position as marked as [1]. The participants were instructed to respond to the picture as quickly as possible by indicating whether the entity described in the picture was alive or not.

(14) Fred chased the squirrel to which the nice monkey explained the game's [1] difficult rules [2] in the class last Wednesday.

The pictures described the referents of RC head nouns half of the time (e.g., a squirrel) and an unrelated object the other half of the time (e.g., a toothbrush). The results showed that the

² See for example, Pickering and Barry (1991) and Traxler and Pickering (1996) for an alternative account of filler-gap processing that does not depend on an empty category

English native speakers with relatively high WMC responded faster to the pictures describing the referent of the filler than to those describing an unrelated object at the gap position [2], but crucially, not at the earlier position [1]. The result suggests that the filler was reactivated at the gap position, lending support to the presence of an empty category that mediates the integration of the filler and the subcategorizing verb. The L2 learners, in contrast, showed a semantic priming effect at both positions, and the authors interpreted this lack of a *reactivation* effect as indicating that adult L2 learners do not postulate an empty category in the gap position. Instead, the authors say, they must directly associate the filler with its lexical licenser (cf. Gorrell, 1993; Pickering & Barry, 1991).

The second study that provides evidence for the lack of empty categories in L2 parse is Marinis et al.'s (2005) self-paced reading experiments on the role of a theoretically motivated intermediate gap intervening between a filler and its original gap position. Multiple syntactic theories posit the presence of a syntactic marker that mediates the structural association between a fronted phrase and its gap spanning one or more clausal boundaries. In (15), for example, an intermediate gap site is marked with *e'* at the beginning of the complement clause of the RC verb *claimed*, as would be postulated by, for example, a successive cyclic movement of a *wh*-phrase. In contrast to (15), there is no intervening gap between the filler (i.e., relative pronoun) and its subcategorizing verb (i.e., RC verb) in (16).

(15) The manager_i who_i the consultant claimed *e'* that the new proposal had pleased *e_i* will hire five workers tomorrow.

(16) The manager_i who_i the consultant's claim about the new proposal had pleased *e_i* will hire five workers tomorrow.

An earlier self-paced reading experiment on native speakers of English by Gibson and Warren (2004) compared the reading time for sentences such as (15) with that for sentences such as (16). The results showed that the RC verb region (i.e., *had pleased*) was read faster in (15) than in (16) despite the fact that the surface distance between the relative pronoun and its base gap position was the same for the two sentences. Gibson and Warren interpreted the results as suggesting that the presence of an intermediate gap between the filler and the original gap breaks the dependency into two shorter ones and reduces processing complexity that is assumed to

increase with the distance between the two elements to be integrated (e.g., Gibson, 2000). Modeled on the Gibson and Warren study, Marinis et al. (2005) tested adult L2 learners of English whose L1 was Chinese, Japanese, German or Greek on whether filler-gap processing in the L2 is also facilitated by the presence of a clause boundary between the filler and the gap (and the hypothesized intermediate gap therein). The results showed that the adult L2 learners did not benefit from the presence of a clause boundary, unlike native speakers of English, whether their L1 grammar incorporates overt dislocation of *wh*-phrases (as in German and Dutch) or not (as in Chinese and Japanese). This finding, according to the authors, lends support to the possibility that adult L2 learners directly associate the filler and its lexical licenser without the mediation of an empty category (cf. Pickering & Barry, 1991).

The findings in favor of the absence of empty categories in L2 syntactic representations, however, may need to be qualified by some inconsistent findings and possible alternative accounts. For example, other cross-modal priming studies on bilinguals' filler-gap processing showed that adult L2 learners did not show semantic priming effects in lexical decision or picture classification either at the gap position or at a pre-gap position (Love, Maas, & Swinney, 2009; Miller, 2011), which is not consistent with the maintained activation shown in Felser and Roberts (2007). On the other hand, an experiment by Baek (2010) using a visual lexical decision task following self-paced reading of sentences similar to (14) provides suggestive evidence that the activation state of a filler in the course of L2 sentence processing could be modulated by the canonical word order associated with dative verbs.

The results in favor of the lack of empty categories in L2 parses may be amenable to alternative explanations as well. For example, those results are mainly obtained through quite demanding tasks, either due to the nature of the tasks or the complexity of stimulus sentences. The cross-modal priming study by Felser and Roberts (2007) used auditory sentences that were produced at natural speed. Given generally slower processing of L2 learners, the rapid presentation of stimulus sentences could have overwhelmed L2 learners together with the additional processing demand incurred by the need to perform a secondary task in a different modality (i.e., a visual alive/not-alive decision task). In addition, the slower processing speed of L2 learners would make it harder to temporally map the presentation of auditory input and the cognitive processes invoked by the input. The self-paced reading study by Marinis et al. (2005) may be free from this speed-related problem but still may impose substantial processing

demands due to the complexity of the experimental sentences. They contained multiple embeddings and were presented in a non-cumulative way (i.e., the preceding segment disappeared when a segment appeared on the screen, so rereading the preceding segments was not allowed). These two factors might have increased overall processing complexity to the extent that normal online processing broke down for L2 learners.

However, two recent studies provide further evidence for L2 learners' over-reliance on lexical-semantic information in line with the SSH, one using eye-movement monitoring (Felser, Cunnings, Batterham, & Clahsen, in press) and the other using self-paced reading (Jackson, 2008). Felser et al. (in press) used two diagnostics to investigate the nature of L2 filler-gap processing involved in the comprehension of RCs. One was the effect of plausibility of the resulting filler-gap association as revealed while reading sentences such as (17) and (18).

(17) Everyone liked the magazine that the hairdresser read extensively and with such enormous enthusiasm about before going to the salon.

(18) Everyone liked the shampoo that the hairdresser read extensively and with such enormous enthusiasm about before going to the salon.

The RC head nouns in the two sentences, *magazine* and *shampoo*, are both the object of the preposition *about* in the RC, but they can be temporarily interpreted as the object of the RC verb *read* because the verb is ambiguous between transitive and intransitive use. Thus a temporary filler-gap integration at the RC verb results in a plausible proposition in (17) (i.e., 'the hairdresser read the magazine') but in an implausible proposition in (18) (i.e., 'the hairdresser read the shampoo'), with increased reading times expected in the latter case (e.g., Traxler & Pickering, 1996; Williams et al., 2001).

The other diagnostic was filled-gap effects as expected in reading sentences such as (19) and (20) (cf. Stowe, 1986). In both sentences, the filler *magazine* is the object of the preposition *about*. In (19), *magazine* can be temporarily interpreted as the object of *read*, whereas it cannot in (20) because of the overt object *articles*. Thus an immediate filler-gap association at *read* would result in increased reading times at the following overt object *articles* in (20) compared with the following adverb *quickly* in (19) because an overt object at the gap position signals that

the initial filler-gap association must be discarded, whereas an adverb is consistent with the initial analysis.

(19) Everyone liked the magazine that the hairdresser read quickly and yet extremely thoroughly about before going to the beauty salon.

(20) Everyone liked the magazine that the hairdresser read articles with such strong conclusions about before going to the beauty salon.

The results of Felser et al. (in press) showed that plausibility effects appeared at the RC verb region (i.e., *read extensively* in (17-18)) early in first-pass reading times for L1-German L2-English learners, while it appeared at the same region only in late measures such as re-reading times for L1 speakers of English. In contrast, filled-gap effects appeared at the post-verb region (i.e., *quickly* in (19) and *articles* in (20)) early in first-pass reading times for L1 speakers, whereas it only appeared at the subsequent region (i.e., *and yet* in (19) and *with such* in (20)) in re-reading times for L2 learners. Assuming that plausibility effects are the result of a lexically-based semantic integration and filled-gap effects are the result of the application of phrase structure information (i.e., the association of the filler with an unrealized argument slot), Felser et al. took the difference in the timing of the two effects between L1 and L2 sentence processing as suggesting that the latter primarily resorts to lexical-semantic relations between constituents, with structure-based processes coming into play only in later stages of processing.

The authors' interpretation of the finding as evidence for a general deficit in the structural processes in L2 sentence processing should be qualified by the extent to which the behavioral dissociation between plausibility effects and filled-gap effects reflects the hypothesized distinction between 'semantic' and 'structural' processes. Their findings, however, provide important insights into the differences between L1 and L2 processing by showing that L2 learners' sensitivity to (temporary) semantic anomaly can manifest earlier and/or more strongly in behavioral measures compared with their sensitivity to (temporary) ungrammaticality of a given sentence, which is consistent with the SSH.

Jackson (2008) also suggests that lexical-semantic information is crucial in L2 learners' online construction of a syntactic analysis, while they sometimes fail to utilize morpho-syntactic information such as case markers for the same purpose. In the study, intermediate and advanced

L2 learners and native speakers of German read simple wh-questions in German such as (21) and (22).

(21) a. Welche Ingenieurin traf den Chemiker gestern Nachmittag im Café?

Which_{NOM/ACC} engineer met the_{ACC} chemist yesterday afternoon in-the cafe?

"Which engineer met the chemist yesterday afternoon in the cafe?"

b. Welche Ingenieurin traf der Chemiker gestern Nachmittag im Café?

Which_{NOM/ACC} engineer met the_{NOM} chemist yesterday afternoon in-the cafe?

"Which engineer did the chemist meet yesterday afternoon in the cafe?"

(22) a. Welche Ingenieurin hat den Chemiker gestern Nachmittag getroffen?

Which_{NOM/ACC} engineer has the_{ACC} chemist yesterday afternoon met?

"Which engineer met the chemist yesterday afternoon?"

b. Welche Ingenieurin hat der Chemiker gestern Nachmittag getroffen?

Which_{NOM/ACC} engineer has the_{NOM} chemist yesterday afternoon met?

"Which engineer did the chemist meet yesterday afternoon?"

The wh-phrase *Welche* 'which' in (21) and (22) is ambiguous in its case, resulting in a temporary ambiguity on its syntactic role in the sentence. This ambiguity is resolved at the second NP *den/der Chemiker* 'the chemist', where case information is explicitly marked on the definite article *der* (i.e., nominative) or *den* (i.e., accusative). Thus the ultimately correct grammatical role of the wh-phrase is subject in (21a) and (22a), whereas it is object in (21b) and (22b). Given the general preference for subject analysis of a sentence-initial NP, Jackson predicted that increased reading times would be observed at the point when the comprehender recognized that the case of the second NP was nominative, in which case the first NP must assume a less preferred accusative case.

The difference between (21) and (22) lies in the location of the thematic-role assigning verbs. In (21), the past tense verb *traf* 'met' immediately follows the wh-phrase *Welche Ingenieurin* and precedes the second NP *den Chemiker*, with the lexical meaning of the verb available to the comprehender when assigning the thematic role to the second NP (and revising

that of the first NP, i.e., wh-phrase, when needed). In (22), however, the past participle *getroffen* ‘met’ appears at the end of the sentence and the auxiliary *hat* instead intervenes between the two NPs, due to a syntactic well-formedness requirement in German. Therefore, any evidence for processing difficulty at the nominative second NP in (22b) (arguably due to the reanalysis of the initially preferred subject-initial analysis) would suggest that the comprehender incrementally constructs syntactic representations solely based on the case marking on the NPs without the aid of lexical meaning available from the verb.

The results of Jackson (2008) showed that native speakers were slower in reading the second NP when it was marked as nominative than when it was marked as accusative regardless of the position of the lexical verbs, indicating that they preferred subject-first analysis and immediately reanalyzed the initial analysis when triggered by the case information marked on the second NP. L2 learners, however, showed different patterns of behavior from L1 speakers, as well as among themselves, depending on proficiency. Advanced learners showed evidence for an immediate reanalysis at the second NP in (21), but not in (22), suggesting that they are able to take into account case information on articles to incrementally compute syntactic relations between constituents but some substantial lexical-semantic information may be required for them to do so (cf. Jackson & Roberts, 2010, for a similar line of discussion³). Intermediate learners’ behavior further differed, with any evidence for the hypothesized reanalysis involved in object-extracted questions only appearing at the sentence-final word both in (21) and (22), which suggests that they postponed a determinate interpretation of a sentence until the end of the sentence.

On the other hand, there is also evidence that at least some part of the structural processes involved in advanced and near-native learners’ L2 sentence processing closely resembles those involved in L1 sentence processing, inconsistent with the SSH. Such evidence comes from previous studies that suggest that proficient L2 learners use arguably structure-based parsing principles (e.g., a preference for subject analysis of a clause-initial NP or a relative pronoun) and they are immediately sensitive to syntactically relevant morphological cues such as case and number marking during online sentence processing (e.g., Hopp, 2006, 2010; Jackson & Dussias, 2009; Jackson & Roberts, 2010; Omaki & Arijji, 2005; see Havik, Roberts, van Hout, Schreuder,

³ Jackson & Roberts’ (2010) study on RC processing will be discussed in Section 4.2 after I review previous findings and theoretical accounts for RC processing in L1.

& Haverkort, 2009, Hopp, 2010, Jackson & Bobb, 2009, for task demands effects). For example, Hopp (2010) tested how advanced L2 learners of German whose L1 was English, Dutch, or Russian dealt with word order variation in embedded clauses as exemplified in (23) and (24), using self-paced reading and speeded grammaticality judgments.

(23) a. Er denkt, dass der Hotelier im August den Gastwirt angezeigt hat.

He thinks that the_{NOM} hotel owner in August the_{ACC} landlord sued has

‘He thinks that the hotel owner sued the landlord in August.’

b. Er denkt, dass den Gastwirt im August der Hotelier angezeigt hat.

He thinks that the_{ACC} landlord in August the_{NOM} hotel owner sued has

‘He thinks that the landlord sued the hotel owner in August.’

(24) a. Sie sagt, dass die Radlerin am Freitag die Anwohner behindert hat.

She says that the cyclist on Friday the residents obstructed has

‘She says that the cyclist obstructed the residents on Friday.’

b. Sie sagt, dass die Anwohner am Freitag die Radlerin behindert hat.

She says that the residents on Friday the cyclist obstructed has

‘She says that the residents obstructed the cyclist on Friday.’

The sentences in (23) and those in (24) differ only in the surface order of the subject and object of the embedded clause, with a canonical SOV order displayed in (23a) and (24a) and a non-canonical OSV order displayed in (23b) and (24b). In (23), the correct syntactic functions of the embedded NPs are signaled by case information on the definite articles *der* (nominative) and *den* (accusative): the NP with nominative case *der Hotelier* ‘the hotel owner’ serves as the subject and the NP with accusative case *den Gastwirt* ‘the landlord’ as the object. In (24), on the other hand, the definite article *die* does not explicitly mark case due to morphological syncretism. Instead, the subject-verb agreement between the singular NP *die Radlerin* ‘the cyclist’ and the auxiliary verb *hat* indicates the appropriate syntactic roles of the NPs in the embedded clause.

The overall results of Hopp (2010) suggest that adult L2 learners can learn to process L2 morpho-syntactic information in a native-like way at least with a sufficiently high level of proficiency, and that non-native-like parsing patterns in less proficient L2 learners are attributable to increased processing demands rather than to representational deficits. In the self-

paced reading task, the native speakers of German slowed down at the accusative-marked first NP in (23b) *den Gastwirt* as compared with the nominative-marked first NP in (23a) *der Hotelier*, and they also slowed down at the sentence-final auxiliary verb *hat* that agreed in number with the second NP *die Radlerin* as in (24b), as compared to the auxiliary verb that agreed with the first NP as in (24a). The results suggest that native speakers of German initially prefer a subject-initial analysis in embedded clauses and initiate a revision as soon as they encounter information that conflicts this initial analysis, whether the relevant information is provided by case marking (23) or by subject-verb agreement (24). The near-native L2 learners showed the same pattern of reading times as native speakers, indicating their ability to immediately incorporate these kinds of formal information. Advanced learners, however, showed weaker evidence for processing difficulty at the initial accusative NP in (23b), while they showed a reliable sign of processing difficulty at the sentence-final auxiliary verb both in (23b) and (24b) compared with (23a) and (24a), respectively. The results suggest that L2 learners without a sufficient level of proficiency may not immediately incorporate formal information such as case marking and subject-verb agreement.

The results of the speeded grammaticality judgment task in the Hopp study suggest that the delayed reanalysis in advanced L2 learners may be primarily due to increased processing loads. When stimulus sentences were presented at so high a rate as to impose unusual processing loads even on native speakers, the pattern of native German speakers' performance in detecting ungrammatical case marking and subject-verb agreement came to resemble that of L2 learners evidenced at a slower presentation speed (cf., McDonald, 2006). Based on these findings, Hopp proposes that the observed differences between L1 speakers and adult L2 learners in online sentence processing is primarily due to increased processing loads for L2 learners and resulting difficulty in their integrating different levels of linguistic information (e.g., syntactic and morphological information), rather than fundamental differences in underlying grammatical representations⁴.

⁴ Hopp (2010) mainly attributes adult L2 learners' processing inefficiency to L1 influence, based on the modulating effects of morpho-syntactic characteristics of their L1 on reading times and grammaticality judgments in the L2. For example, Hopp's self-paced reading data suggest that L1-Russian learners are better able to correctly process case information during online sentence processing and grammaticality judgments than L1-English and L1-Dutch learners. This L1 influence is arguably due to the fact that Russian shows overt case markings on full NPs while English and Dutch do not, which heightens L1-Russian L2-German learners' sensitivity to case information in the target language.

To summarize, L2 learners do not seem to differ from native speakers in the ways and time course of using non-structural information such as lexical properties of words in online sentence processing. On the other hand, empirical findings and theoretical views diverge on the issue of whether adult L2 learners can incorporate detailed syntactic relationships among constituents in interpreting L2 sentences in the same ways and to the same extent as native speakers of the language would do. The SSH proposes that adult L2 learners cannot construct as detailed syntactic structures as native speakers in real-time sentence processing because their grammatical knowledge in the L2 or the access route to that knowledge is qualitatively different from native language, and thus they end up over-relying on lexical-semantic and pragmatic information (Clasen & Felser, 2006; Felser et al., in press). Alternative views propose that the sentence processing mechanisms responsible for an L2 acquired later in life are not fundamentally different in nature from those used for a native language and the apparent differences between them are attributed to slower or more inefficient processing due to reduced experience or proficiency in the L2 and/or transfer of L1 processing routines (e.g., Dussias & Sagarra, 2007; Frenck-Mestre, 2002; Hopp, 2010).

Several questions remain unresolved and open to further research for both the SSH and alternative exposure/experience-based views in order to provide adequate explanations of L2 sentence processing. As for the SSH, it seems unclear how complex a syntactic structure should be in order for the structure to qualify as a test case for the architectural difference between native speakers' normal processing and non-native speakers' 'shallow' processing. It is hardly surprising that L2 learners have greater difficulty parsing complex L2 sentences and thus often have no choice but to infer the meaning of a given sentence using some general-purpose heuristics that primarily resort to semantic and pragmatic information available from successfully processed words and common knowledge about the world (cf. Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, 2003; Ferreira & Patson, 2007). But this may not necessarily imply that there is some domain-specific deficit in syntactic processes involved in L2 sentence processing, especially given the previous findings that adult L2 learners' real-time sensitivity to structural information is modulated by various *quantitative* factors such as the complexity of the target structure, task demands and participants' proficiency/experience in the L2. Thus, as long as evidence for the SSH comes from tasks quite demanding in terms of either stimulus complexity or task demands, it may be hard to prevail over alternative

explanations appealing to increased processing loads and overall inefficiency in integrating various kinds of information. Conversely, if evidence for L1-L2 differences can be drawn from relatively simple structures that would not break down normal, incremental processing by L2 learners, then it may lend further support to the qualitative difference between L1 and L2 sentence processing mechanisms.

The proficiency effect as shown in previous studies is another issue to which the SSH has yet to pay due attention. Since the hypothesis as it is does not incorporate considerations of developmental changes in L2 sentence processing, it is unclear whether the hypothesis represents an attempt to best capture the average performance of L2 learners with different characteristics or an attempt to present a model of sentence processing mechanisms that adult L2 learners can ultimately achieve. This oversight, perhaps, is responsible for much of the debate on adult L2 sentence processing (e.g., Hopp, 2006, 2010). To clear out the obscurity and provide a more adequate explanation of L2 sentence processing, any model of L2 sentence processing would need to incorporate explanations of observed proficiency effects.

Experience-based accounts are free from these difficulties because increased processing loads and lower proficiency in the language on the part of L2 learners are exactly what explain the differences between L1 and L2 sentence processing. But what is not adequately explained by these accounts may be why the differences between L1 and L2 sentence processing appear to be localized to presumably (morpho)-syntactic processes. If the differences between L1 and L2 sentence processing are due to the overall inefficiency of adult L2 learners, we would expect them to experience comparable difficulties appreciating lexically- and structurally-based information alike, which is not consistent with the majority of the previous findings. Therefore, an adequate experience-based account for L2 sentence processing would need to explain the source of this apparent asymmetry in L2 learners' performance without recourse to some architectural differences between L1 and L2 sentence processing mechanisms.

The present dissertation is an attempt to contribute to resolving these unresolved questions. For present purposes, English relative clauses (RCs) were chosen as the target structure because the moderate complexity of their syntactic and semantic representations was considered appropriate for investigating various aspects of L2 sentence processing. RCs seem to be neither too complex for advanced L2 learners to perform incremental processing nor too simple for substantial behavioral differences between L2 learners and native speakers, if any, to

appear in online measures such as self-paced reading. In addition, the extensive coverage of RC processing in the L1 sentence processing literature provides a rich source of empirical and theoretical background based on which L2 learners' behavior could be understood in a systematic way. A few studies have already tested L2 processing of RCs in multiple languages and offered important insights into the nature of L2 sentence processing (e.g., L2 Dutch, Havik et al., 2009, Jackson & Roberts, 2010; L2 English, Hashimoto, 2007; Lim, 2011; Omaki & Arij, 2005; L2 Korean, Lee, 2010). The details of the studies relevant to the present dissertation will be discussed in Section 4.2 after previous studies on L1 RC processing are reviewed in Section 4.1.

Another issue of interest in the present dissertation is the role of individual differences in working memory capacity (WMC) in L2 sentence processing. Two potential uses of WMC as a tool to study L2 sentence processing motivated the present dissertation. Firstly, it could serve as one of the explanatory variables for a wide variability among L2 learners in the real-time use of the L2. Secondly, we can compare the role of WMC in L1 and L2 sentence processing as a way to investigate similarities and differences between L1 and L2 sentence processing mechanisms, which is the primary focus of the present dissertation. In the next chapter, the theoretical concepts and empirical implementations of WMC are introduced and then previous findings on its role in L1 and L2 sentence processing are reviewed.

Chapter 3

Working memory capacity in sentence processing

Working memory (WM) in its general sense refers to the memory system for temporary storage and processing of information that is required to perform mental activities (e.g., an arithmetic operation, $134 + 27 = ?$), throughout which the representations to be manipulated must be kept activated in memory (e.g., a set of digits) and a certain mental manipulation must be applied onto them (e.g., addition) (e.g., Gathercole, 2008)⁵. WM underlying language comprehension is thus conceptualized as a set of cognitive processes that are involved in maintaining linguistic input and extracting its meaning in real-time. When reading a text comprising multiple paragraphs, for example, the reader must interpret each sentence in a paragraph and then combine the meaning of a sentence to those of others to make sense of the overall idea of the paragraph, which in turn must be combined with the main ideas of the other paragraphs to reach a coherent understanding of the whole text (e.g., Daneman & Carpenter, 1980). The WM system operating at the level of sentence processing in turn amounts to a set of cognitive processes and resources that are responsible for constructing sentential representations based on the currently recognized string of words and storing the products of such computation to be integrated with later coming word(s) (e.g., Just & Carpenter, 1992).

Individual differences in the availability of WM resources have been shown to be associated with individual differences in the speed and/or accuracy in performing various kinds of higher order cognitive tasks such as reasoning, direction following, word learning and language comprehension (e.g., Engle 2010). In the same vein, individual differences in the available WM resources are hypothesized to result in individual differences in overall reading skills (e.g., Daneman & Carpenter, 1980) and the speed and/or accuracy of online sentence processing (e.g., Just & Carpenter, 1992).

The availability of WM resources is often referred to as working memory capacity (WMC). This term, however, does not necessarily mean that individuals with higher WMC possess a physically greater capacity of processing resources. Greater capacity (i.e., higher availability of WM resources) can be effectively achieved by efficient processing (e.g., Just &

⁵ WM is often contrasted with the traditional concept of simple short-term memory as passive storage of a limited amount of information (Miller, 1956), which has been usually operationalized as the maximum number of letters, words or digits that an individual can recall shortly after presentation (e.g., Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Turner & Engle, 1989).

Carpenter, 1992; Engle, 2010). If an individual is faster or more efficient than others in computing the syntactic structure of a given phrase, he or she will be also better in retaining the phrase in memory because structured input is easier to remember than unstructured input (e.g., Miller, 1956) and/or because additional processing resources can be devoted to maintaining information in memory (e.g., Just & Carpenter, 1992). The term WMC as used in the present dissertation thus should be taken as a cover term to refer to the degree of functional availability of WM resources rather than as referring to physical capacity of some sort.

3.1. Measurements of WMC

Various tests have been developed to measure an individual's WMC (cf. Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005; Waters & Caplan, 2003). They are designed to tap both storage and processing components of WM by combining some kind of processing task with some kind of memory storage task. In some tests, the processing task is unrelated to the to-be-remembered items and is interposed between them, while in other tests, the processing task should be performed directly on the memory items. The number of items to be recalled at a time varies generally from two to seven, with the number of processing tasks covarying with the number of memory items. An individual's WM measure is obtained from either the total number of correctly recalled memory items throughout a test or the maximum number of memory items that the individual can correctly recall at a time above a certain threshold level of accuracy. The latter type of measure has been traditionally called WM 'span', but the term will be used in the present dissertation to refer to either total score or maximum span. The accuracy in the processing task usually hovers at a ceiling level (mostly because it is simple), so it is often used as the criterion to detect unattentive participants (Conway et al., 2005; cf. Waters and Caplan, 1996c). Thus the degree of accuracy in the recall component of the WM tests (i.e., WM span) is taken to reflect one's WMC on the ground that it has been achieved using the processing resources available even when the individual is engaged in the processing task.

Widely used WMC tests include, for example, the reading span, alphabet span, operation span, and subtract-2 span tests⁶. The name of each test indicates what kind of processing task

⁶ Working memory span tasks can be verbal or visuospatial depending on the perceptual medium of the tasks. All of the given examples of span tasks are verbal in that they involve verbal elements. Visuospatial working memory span tasks, on the other hand, generally involve the processing of spatial information of some sort and the memorization of sequences of visual stimuli. In a symmetry span task (e.g., Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, 2004), for example, participants are presented with an 8×8 matrix with a certain number of squares filled in black.

must be performed while retaining some verbal stimuli. The reading span has been the most widely used test format in the sentence processing literature. In the original version(s) of the test as developed by Daneman & Carpenter (1980), participants read (and judged the grammaticality of) a set of sentences and then recalled the last word of each sentence. The number of sentences in each set increases from two to five. Participants were presented with three or five consecutive sets of the same size. Their reading span corresponds to the maximum set size on which they could correctly recall the last words of all the sentences for the majority of the sets of the same size, thus ranging from two to five. Later versions of the reading span test incorporate various changes in procedure. For example, (i) plausibility judgments have been used as the processing task, (ii) random words or alphabet letters have been used as recall items instead of the last word of the stimulus sentences to prevent participants from using some mnemonic strategies based on the content of the sentences, (iii) set size has been varied randomly to prevent participants from guessing how many sentences they will read for the current set, or (iv) the total number of correctly recalled memory items has been taken as reading span score (e.g., Conway et al., 2005; Friedman & Miyake, 2005).

In a typical alphabet span test, participants are presented with sets of English words that they should repeat in alphabetical order. In an operation span test, participants are required to perform simple math operations and later recall random set of letters or digits each presented after an operation was solved. In the subtract-2 span test, participants are presented with a set of digits of varying numbers, which they are required to repeat in order of presentation after subtracting two from each of the digits.

Although these WM span measures have been shown to be fairly reliable (e.g., Conway et al, 2005), it is recommended that multiple span tests be used simultaneously to draw a composite score to reduce sampling errors and increase reliability (e.g., Waters & Caplan, 2003).

They decide whether the black squares appear in symmetry against the vertical center of the matrix. After the decision, they are presented with a 4×4 matrix with a red square in random position. After repeating this procedure for a certain number of times, participants recall the position of the red square in each matrix in order. Participants' symmetry span is determined by for how many matrices they can correctly recall the position of the red square. The simple version of this task consists only of the recall component without the symmetry-decision task. Kane et al. (2004) show that verbal and visuospatial WMC as measured by multiple versions of working memory span tests share 70-85% of variance, suggesting that a significant part of WMC is domain-general. In this dissertation, I focus on verbal WMC (i) because sentence processing is obviously verbal and (ii) because L2 learners' behavior in sentence processing in the L2 would be constrained by their limited linguistic competence and thus the relationship between WMC and sentence processing would be better captured by verbal WMC which takes into account L2 learners' limited verbal ability in the L2.

In this light, the present dissertation obtained composite WM scores from two different span tests: a reading span and a subtract-2 span test. Waters & Caplan (2003) showed that these two span tests were quite reliable, especially when WM scores were obtained from the total number of correctly recalled items.

Despite the predictive power of various WMC measures for individual differences in performing various cognitive tasks, the specific role that the individual differences in WMC play in sentence processing is still a matter of debate. We turn to this issue in the next section.

3.2. Different views on the nature of cognitive resources underlying sentence processing

The role of a capacity-limited processing system has long been assumed in the sentence processing literature, as manifested in multiple theories of sentence processing that incorporate limited processing resources as a major constraint on the human sentence processor (e.g., Frazier & Fodor, 1978; Gibson, 1998, 2000). However, previous research leaves unresolved questions as to the architecture and mechanism of the processing system itself (e.g., Caplan & Waters, 1999). Researchers disagree on the nature of the WM system that carries out online sentence processing. The debate is represented by two major questions: (i) whether the reading span (and other WM spans) is a valid indicator of the capacity or efficiency of processing resources that are supposed to drive sentence processing (e.g., Just & Carpenter, 1992; Waters and Caplan, 1996a) and (ii) whether the concept of the capacity-limited processing system that is separate from the linguistic knowledge system is necessary to explain individual differences in sentence processing (e.g., MacDonald & Christiansen, 2002).

The question as to the validity of reading span (and other WM spans) as an indicator of the availability of cognitive resources responsible for sentence processing is related to the modularity of the cognitive processes involved in sentence processing. Some researchers make a distinction within language comprehension between ‘unconscious’, ‘obligatory’ processes involved in extracting the meaning of a sentence (i.e., interpretive processes) and ‘conscious’, ‘controlled’ processes involved in verbally-mediated cognitive tasks (i.e., post-interpretive processes) (e.g., Caplan & Waters, 1999; Waters & Caplan, 1996a; Waters & Caplan, 2004; cf. Fodor, 1983). Interpretive processes include a wide range of cognitive processes involved in online sentence comprehension such as lexical access, assignment of syntactic and thematic

structures to a given string of words, and extraction of the meanings of the resulting structures. Post-interpretive processes, on the other hand, refer to verbally-mediated general cognitive processes such as coherently integrating propositions abstracted from multiple sentences and explicitly searching memory to recall necessary information. Accordingly, a specialized WM system is postulated for the execution of interpretive processes, which is separate from the one responsible for post-interpretive processes. Under this *separate-resource* view, the reading span test merely combines two different tasks that draw on different pools of processing resources (i.e., processing of sentences and storage of memory items). The reading span (as well as other WM spans) as measured by the number of correctly recalled items thus primarily reflects explicit efforts to retain memory items, which is a controlled post-interpretive process rather than an obligatory interpretive process. Consequently, it is predicted that WMC as measured by WM span tests will not correlate with the efficiency of most aspects of online sentence processing. This prediction is consistent with a large set of experimental results that failed to find reliable correlations between WMC on one hand, and speed and accuracy in processing syntactically complex or ambiguous sentences on the other (e.g., Waters & Caplan, 1996b, 2004; see Caplan & Waters, 1999, for a thorough review of such findings).

The *separate-resource* view of sentence processing is contrasted with the *single-resource* view, according to which language comprehension in general is subserved by a single pool of processing resources (e.g., Fedorenko, Gibson, & Rohde, 2006; Just & Carpenter, 1992; Just, Carpenter, & Keller, 1996). This view does not posit a modular distinction between processes involved in the immediate interpretation of sentences and those involved in verbally-mediated general cognitive tasks, making it unnecessary to postulate specializations in processing resources responsible for language comprehension. Under this view, the processing and storage components of the reading span test (and other WM span tests) compete for the same finite pool of processing resources. Therefore, reading span and other WM spans are assumed to reflect the overall capacity or efficiency of the WM system because higher span would mean that the individual had more processing resources available for storage even after a significant amount of resources had been devoted to the processing task. The single-resource view therefore predicts that individual differences in WMC as measured by the traditional WM span tests will be associated with individual differences observed in online sentence processing as well as in general reading comprehension (e.g., Daneman & Carpenter, 1980; Just & Carpenter, 1992).

Consistent with this view, a set of previous experimental findings suggests that readers with lower reading span experience greater difficulty in processing object-extracted relative clauses (e.g., Fedorenko et al, 2006; King & Just, 1991) and those with higher reading span are sensitive to a wider range of different types of constraints in resolving syntactic ambiguities (e.g., Just & Carpenter, 1992; MacDonald, Just, & Carpenter, 1992; Pearlmutter & MacDonald, 1995).

Though in contrast in a crucial way, the separate- and single-resource views agree on the presence of a WM system that varies independently of the ‘linguistic’ system. A connectionist approach to sentence processing and language comprehension casts doubt on this view, however (MacDonald & Christiansen, 2002). According to a connectionist view of language system, there is no clear boundary between linguistic knowledge as symbolic representations, and linguistic processes as operations on those symbols (e.g., Elman, 1991). Instead, a network of processing units *represent* different concepts by means of different patterns and strengths of associations among the processing units and *process* input by activating the relevant processing units through weighted associations. Thus it becomes unnecessary to postulate a separate processing system whereby separately stored symbols are retrieved and manipulated. Consequently, individual differences in sentence processing cannot be attributed to a processing system of which the capacity varies across individuals, but instead arise from individual differences in the quantity and quality of prior experience with language (and biological differences in the quality of phonological representations), because what determines the patterns and strengths of associations among processing units within the language network is prior experience. According to this *experience-based view* of individual differences in sentence processing, performance in sentence processing tasks and that in traditional WM span tasks tend to correlate not because the two different types of tasks draw on the same pool of processing resources, but because both types of tasks are ‘linguistic’ in nature and thus carried out by the same processing network.

Somewhat orthogonal to these three different views focusing on sentence processing/language comprehension, more memory-oriented research suggests that individual differences in WMC do not predict performance on tasks that are largely based on automatic processes (e.g., Kane, Bleckley, Conway, & Engle, 2001; Unsworth, Schrock, & Engle, 2004). In Kane et al. (2001), for example, participants with varying operation spans identified visual stimuli (e.g., *B* or *I*) that were presented either to the left or right of where the participants originally fixated. The presentation of targets was signaled by a flickering visual cue (i.e., =). In

one condition, targets appeared at the same position as the visual cue, allowing the participants to naturally move their eyes to the attention-attracting visual cue (pro-saccade condition). In the other condition, targets appeared at the opposite side to the visual cue, forcing the participants to suppress the reflexive response toward the visual cue (anti-saccade condition). The results showed that the participants with high and low operation span did not differ in the speed and accuracy at identifying visual targets in the pro-saccade condition, whereas the participants with high span were reliably faster and more accurate than those with low span in the anti-saccade condition. This finding was interpreted by the authors as suggesting that the ability to allocate attention to task-relevant information (e.g., whether to pro- or anti-saccade) while suppressing the activation of task-irrelevant information (e.g., a flickering visual cue) is a crucial determinant of WMC. This view would submit that at least some processes involved in sentence processing that are so automatized as to not require attention control would not be subject to individual differences in WMC (e.g., phonological/orthographic decoding and lexical access by native speakers of a language).

As mentioned above, the single- and separate-resource views could be grouped together against the experience-based view in that the former two views postulate some pool of processing resources that varies independently of individuals' prior language experience. From a different perspective, the single-resource and experience-based views can be categorized together with the exclusion of the separate-resource view because the former two views agree in assuming a non-modular view of sentence processing and predicting some meaningful relationship between individual differences in WMC and those in sentence processing. The finding of the WM literature that automatic processes may not draw upon WM resources as measured by WM span seems to strike a chord with the separate-resource view in that both views postulate a distinction between sets of cognitive processes involved in language processing according to whether or not they draw upon WM resources as measured by WM span tests.

It is out of the scope of the present dissertation to systematically evaluate these different views. For the purpose of investigation, I take as a working hypothesis a version of the separate-resource view as formulated by Traxler et al. (2005), which will be elaborated in Section 4.1, because it highlights the potential distinctions between different processes involved in sentence processing as reflected in the pattern of their relationships with WMC. This characteristic of the view fits well with the intention of the present dissertation to test whether and to what extent L1

and L2 sentence processing share the same kind of processes by comparing the processes' relationship with WMC. In a similar light, the view that WMC does not predict automatic cognitive processes might also be instrumental for the present purpose because a predictive force of WMC could reveal the degree of automatization of cognitive processes involved in sentence processing in L1 and L2.

3.3. Previous findings on the role of WMC in second language sentence processing

It has been shown that L2 learners' reading span measured in the L2 correlates with their general reading comprehension skills in the L2 as reflected in various offline measures such as standardized reading comprehension tests (e.g., the reading section of TOEFL) and experimentally constructed reading comprehension questions (e.g., Alptekin & Erçtin, 2010; Harrington & Sawyer, 1992; Walter, 2004). This finding suggests that L2 reading span reflects WMC that is associated with individual differences in (some of the) higher order processes involved in the comprehension of L2 texts. There is also evidence that unlike reading span, passive short-term memory capacity as measured by simple word span in L2 does not predict individual differences in reading proficiency in L2 (Harrington & Sawyer, 1992), which is analogous to the similar finding in the L1 psycholinguistics literature (e.g., Daneman & Merikle, 1996).

In the L2 literature, questions have been raised as to which language should be used to measure L2 learners' WMC (e.g., Juffs, 2004; Havik et al., 2009). The present dissertation is primarily interested in learners' WMC that is measured in the L2 rather than in the L1. WMC as measured by a complex span test largely depends on, by definition, how efficient an individual is at the test's processing component and how familiar he or she is with the stimuli to be processed and recalled. Since L2 learners should experience increased difficulty in most aspects of language processing in the target language as compared with native speakers of the language, the effective WMC relevant to L2 sentence processing would be the one measured in the L2. Empirical evidence also suggests that L2 WMC is more closely associated with L2 reading comprehension than L1 WMC is (e.g., Alptekin & Erçtin, 2010; Walter, 2004).

However, using L2 to measure L2 learners' WMC leads to the concern that L2 WMC might be confounded to a certain extent with the level of learners' general proficiency in the L2. It is an important empirical question for the purposes of the present dissertation whether L2

WMC taps a different construct than what general L2 proficiency reflects. As a preliminary attempt to address this issue, L2 learners' proficiency in the L2 was measured by a cloze test and included in the statistical analysis of the present results. If L2 WMC reflects some aspects of the individual differences in L2 sentence processing that are not accounted for by individual differences in general L2 proficiency, it could be justified as a way to explore the nature of L2 sentence processing and test whether L2 sentence processing draws upon the same pool of processing resources as used by more explicit and controlled processes involved in L2 WM span tests.

Previous work is inconclusive about the relationship between WMC and online sentence processing in L2. Several studies fail to find reliable associations between them (e.g., Felser & Roberts, 2007; Juffs, 2004, 2005; Kim, 2008; Omaki, 2005), whereas the native speaker control groups in many of those studies show positive results, as is the case in the L1 literature. For example, the adult L2 learners in Felser and Roberts' (2007) cross-modal priming study did not show any reading-span-related difference in the activation pattern of wh-fillers during the course of sentence processing. In contrast, among child and adult native speakers of English tested under the same conditions as in Felser and Roberts (2007), only those with higher reading span were shown to locally 'reactivate' the extracted wh-filler at its canonical structural position (Roberts, Marinis, Felser, & Clahsen, 2007). In addition, Kim (2008) and Omaki (2005) suggest that L2 reading span does not constrain the resolution of RC attachment ambiguity in the L2, whereas L1 reading span does modulate the pattern of RC attachment preferences in the L1. These results might suggest that L2 WMC does not constrain online L2 sentence processing in the same way as L1 WMC constrains online L1 sentence processing or that a greater variability of performance among L2 learners both in the reading span test and sentence processing task may mask any effect of L2 WMC, if any.

Other studies, however, do report reliable associations between L2 learners' WMC in the L2 and their online processing of L2 filler-gap dependencies involved in English wh-questions and Dutch relative clauses (RCs) (e.g., Dussias & Piñar, 2010; Havik et al., 2009; cf. Williams, 2006). In Dussias and Piñar (2010), native English speakers were shown to immediately take into account plausibility information when a revision of the initial analysis is required in sentences such as (25) and (26).

(25) Who_i did the police declare *e_i* killed the pedestrian?

(26) Who_i did the police know *e_i* killed the pedestrian?

In the two sentences, the wh-filler *who* was hypothesized to be initially interpreted as the object of the main clause verb *declare/know* before the embedded clause verb *killed* was encountered, and must be reanalyzed to the subject of the embedded clause when *killed* was encountered in order to obtain a correct interpretation. An implausible initial analysis in (25) (i.e., ‘the police declared someone’) was easier for native speakers to give up when it turned out to be wrong, as compared with a plausible initial analysis in (26) (i.e., ‘the police knew someone’), suggesting that they did not make as strong a parsing commitment to the implausible analysis as to the plausible one (e.g., Pickering & Traxler, 1998; Traxler & Pickering, 1996). The L2 learners in the study, in contrast, showed this pattern of sensitivity to plausibility information only when their reading span in the L2 was high. Based on the results, Dussias and Piñar proposed that it is only with enough WM resources that L2 learners can utilize lexical-semantic information in a native-like way, which is not in perfect harmony with the view of the shallow structure hypothesis (SSH) that non-structural information is readily available to L2 learners as opposed to structural information. Similarly, Havik et al. (2009) observed that only L2 learners with high reading span showed behavioral signs of online preference for subject-extracted over object-extracted RCs in Dutch, whereas the native Dutch speakers in the study showed the same pattern regardless of their reading span. The result suggests that only with enough WM resources can L2 learners immediately build syntactic structures based on morpho-syntactic information. The results of Dussias and Piñar (2010) and Havik et al. (2009) together suggest the possibility that it may be a lack of sufficient processing resources that is responsible for adult L2 learners’ reduced ability to incorporate either lexical-semantic or morpho-syntactic information during online sentence processing in the L2.

To summarize, L2 WMC seems to reliably predict individual differences in general reading skills in L2 as measured by offline comprehension questions. However, previous work on the role of WMC in online sentence processing in L2 provides conflicting evidence. A group of studies suggests that L2 WMC may not constrain L2 sentence processing in the same way as L1 WMC modulates L1 sentence processing, suggesting some kind of qualitative differences in the sentence processing mechanism between L1 and L2. This set of results, however, should be

interpreted cautiously because they are based on null effects, which could be attributed to the possibility that the effect of WMC, if any, was masked by a greater variability of L2 learners' performance both in the reading span test and the sentence reading task. Another set of results, on the other hand, suggests that a certain threshold level of L2 WMC is required for L2 learners to show native-like behavioral signs in L2 sentence processing whether the behavioral signs are taken to reflect semantically-based or structurally-based processes. These results, however, also seem to have their own limitations in revealing the role of WMC in L2 sentence processing in that they do not directly speak to whether individual differences in L2 WMC are associated with different strategy use in L2 sentence processing while they do suggest that the successful incremental construction of syntactic structures requires a certain level of processing efficiency. The current state of affairs thus seems to call for further research on the role of WMC in L2 sentence processing, which is one of the main goals of the present dissertation.

Chapter 4

Previous findings on relative clause processing in L1 and L2

The present dissertation investigated the processing of English relative clauses (RCs) by native speakers and adult L2 learners of English in order to test the similarities and differences between L1 and L2 sentence processing. RCs were deemed to be a good test case for this purpose based on previous studies on L1 sentence processing that showed a complex interaction of syntactic complexity, semantic information and the effect of individual differences in working memory capacity (WMC) (e.g., Traxler, et al., 2005). In the next section, previous findings and theoretical accounts regarding L1 processing of RCs are reviewed. Previous findings on L2 processing of RCs are reviewed in Section 4.2.

4.1. Relative clause processing in the native language: experimental findings and theoretical accounts

One of the most reliable findings in the sentence processing literature on English is that object-extracted RCs (OR) are harder to comprehend than subject-extracted RCs (SR), as reflected in various behavioral measures such as reading times, comprehension errors, and response latencies in continuous lexical decision and acceptability judgments on subsequent words in a sentence, as well as in neural activities captured by event-related potentials and blood flow in the brain (e.g., Ford, 1983; Forster, Guerrera, & Elliot, 2009; Gennari & MacDonald, 2008; Gibson, Desmet, Grodner, Watson, & Ko, 2005; Gordon et al., 2001, 2004; Grodner & Gibson, 2005; Just, Carpenter, & Keller, 1996; King & Just, 1991; King & Kutas, 1995; Staub, 2010; Traxler et al., 2002, 2005; Wanner & Maratsos, 1978)⁷.

The sentences in (1) and (2), repeated as (27) and (28) below, each contain an SR and an OR. These RCs ‘modify’ their head NP *the reporter* in the sense that the former restrict the referent of the latter to a specific reporter with the characteristic described by the RCs. The head NP in turn serves as the subject of the main clause in the two sentences.

⁷ The same pattern of processing asymmetry between SRs and ORs has been found in several other languages including Dutch (e.g., Mak, Vonk, & Schriefers, 2002, 2006), German (e.g., Schriefers, Friederici, & Kühn, 1995), French (e.g., Holmes & O’Regan, 1981), Chinese (e.g., Lin & Bever, 2006), Japanese (e.g., Miyamoto & Nakamura, 2003; Ueno & Garnsey, 2008) and Korean (e.g., Kwon, Polinsky, & Kluender, 2006), suggesting that the processing consequences of different syntactic roles of the extracted phrases in RCs are universal to a large extent across languages (but see Gibson & Wu, in press; Lin & Garnsey, 2011; Carreiras, Duñabeitia, Vergara, de la Cruz-Pavía, & Laka, 2009 for different findings and accounts in Chinese and Basque).

- (27) The reporter_i that *e_i* attacked the senator admitted the error in the public.
 (28) The reporter_i that the senator attacked *e_i* admitted the error in the public.

In the sentence processing literature, RCs in English are conventionally described as containing a gap at the position where the head noun phrase would appear in canonical declarative sentences. The gaps in (27) and (28) are marked with *e*, which are coindexed with the RC head noun to indicate that they refer to the same entity⁸.

The account for SR-OR asymmetry in processing to be first discussed is based on the cross-linguistic generalization of permissible RC structures in 50 languages (Keenan & Comrie, 1977). According to the Noun Accessibility Hierarchy, the grammatical positions that can be relativized (i.e., the grammatical roles that the RC head can serve in the RC) are hierarchically ordered across languages such that if a certain grammatical position can be relativized in a language, then other positions higher in the hierarchy may be relativized but those lower in the hierarchy may not. The subject is placed at the top of the hierarchy, followed by the direct object, indirect object, oblique, genitive, and object of comparison, in that order. Keenan & Comrie suggest that the relative accessibility to different grammatical roles in RC formation reflects psychological ease of comprehension. As possible explanations of the relative ease of comprehension along the hierarchy, they proposed (i) that those NPs higher in the hierarchy are required in simple sentences more reliably (e.g., a subject is always required and a direct object is required more often than an indirect object, etc.) and (ii) that the grammatical requirement of subjects and RC heads to refer independently⁹ makes ORs psychologically more complex than SRs because the former involve two different cases of necessarily independent reference (i.e., RC head and RC subject), while SRs contain only one necessarily independently referring expression (i.e., RC head).

⁸ In syntactic theories that assume a movement of an argument in RCs (e.g., Chomsky, 1977), what bears a direct syntactic relationship with the gap resulting from the movement is the relative pronoun (e.g., *who*) or a null operator (e.g., in the RCs with the complementizer *that* as in (27-28)). In this dissertation, I simplify the relationship between the RC head and the corresponding gap within the RC by taking the RC head to be the filler, because it does not make any significant difference in terms of processing theories.

⁹ According to Keenan and Comrie (1977: p. 94), the subject of a sentence refers independently in the sense that whenever the subject co-refers with another NP in the sentence (e.g., object), what is marked is not the subject but the object (e.g., in the form of a reflexive pronoun). RC heads share this property of referring independently in that they cannot be 'pronominalized and coindexed' with an NP in the RC.

Another type of accounts for the SR-OR asymmetry in processing focuses on whether the grammatical role assigned to the head noun of an RC is the same in the RC and in the main clause (e.g., MacWhinney & Pleh, 1982; Sheldon, 1974). In the SR in (27), the RC head *the reporter* assumes a subject role both in the RC and in the main clause. In the OR in (28), however, the RC head is assigned a subject role in the main clause and an object role in the RC. According to the perspective shift account proposed by MacWhinney & Pleh (1982), comprehenders take the perspective of the subject of a sentence. Therefore, the sentence (28) that contains an OR involves two occasions of perspective shift during processing: (i) from the main clause subject *the reporter* to the RC subject *the senator* within the RC and (ii) from the RC subject back to the main clause subject after the RC. The sentence (27) that contains an SR, by contrast, keeps the perspective of the main clause subject all the way. Shifting perspectives, among other things such as syntactic complexity, imposes a processing load and thus results in increased difficulty with the OR in (28) compared with the SR in (27). Similarly, the parallel function hypothesis proposed by Sheldon (1974) claims that RCs are easier to comprehend when the RC head assumes the same grammatical function in the main clause and in the RC, which would predict the easier processing of the SR in (27) than the OR in (28)¹⁰.

Some accounts for the OR difficulty highlight limited processing resources available to human comprehenders and increased processing loads induced by a greater distance between the RC head and the RC verb in ORs compared with SRs (e.g., Gibson, 2000; Wanner & Maratsos, 1978). According to the *augmented transition network* model of sentence processing proposed by Wanner & Maratsos (1978), once the beginning of an RC is signaled by the relative pronoun, the head noun of the RC must be held in working memory without being assigned an appropriate grammatical function in the RC until it is integrated with the RC verb. With the available memory resources being limited, holding this unattached material in working memory consumes

¹⁰ Note that the perspective shift and parallel function accounts predict different patterns of relative processing difficulty between SRs and ORs when they modify the objects of main clauses as in *The editor criticized the reporter that attacked the senator*. The perspective shift account predicts a comparable degree of processing difficulty between SRs and ORs that modify an object because there is one instance of perspective change in both types of RCs (i.e., at the RC verb in an SR and at the RC subject in an OR). The parallel function hypothesis predicts the opposite direction of the traditional SR-OR asymmetry for RCs modifying an object because the head noun of an SR now assumes a double function (i.e., the object of the main clause and the subject of the RC) as opposed to the head noun of an OR that assumes the object role in both the main clause and RC. These predictions are not consistent with the findings of Gibson et al.'s (2005) self-paced reading study that showed that ORs were harder to comprehend than SRs whether the RCs modify a sentential subject or a sentential object, undermining the empirical validity of the two accounts. The effect of the grammatical position of RC head nouns, however, is not directly relevant to the questions of the present dissertation and thus is not discussed further.

processing resources and thus subsequent syntactic computation becomes harder. In English, the heads of ORs must be maintained for longer time than the heads of SRs in most cases due to the intervening RC subject in the ORs. Thus processing loads are greater for ORs than for SRs.

Based on similar insights into the relationship between limited processing resources and the processing demands imposed by complex sentences, Gibson's (2000) Dependency Locality Theory (DLT) offers a more specified complexity metrics that takes into account the number of newly introduced discourse referents (e.g., NPs and tensed verbs) between the two elements to be integrated (e.g., a syntactic head and its dependent). Since incorporating a new discourse referent into the current sentential representation taxes processing resources, it becomes harder (and takes longer) to integrate the RC verb with the RC head in ORs because of the intervening RC subject, compared with SRs in which there is no such intervening discourse referent. Note that this integration-based cost is hypothesized to occur at the point of integrating the RC head and RC verb, which is crucial for the purposes of the present dissertation as will be discussed later in this section. Supporting the DLT, processing difficulty with ORs is noticeably reduced when the RC subject is an indexical pronoun such as *I* and *you* as in (30) as compared to when the RC subject is a definite NP as in (29) (e.g., Warren & Gibson, 2002).

(29) The woman who *the boy* had accidentally pushed off the sidewalk got upset.

(30) The woman who *I/you* had accidentally pushed off the sidewalk got upset.

The indexical pronouns such as *I* and *you* are always present and central to the current discourse in normal communications, whereas a full NP is either newly introduced or peripheral to the current discourse (e.g., Gundel, Hedberg, & Zacharski, 1993). The indexical pronouns in (30) therefore do not incur a processing load, unlike the definite NP *the boy* in (29).

While the accounts offered by Wanner & Maratsos (1978) and Gibson (2000) highlight computational demands as the main source of the OR difficulty, another group of accounts emphasize the role of interference between similar items in memory (e.g., Gordon et al., 2001, 2004; Van Dyke, 2007; Van Dyke & Lewis, 2003). Gordon et al.'s (2001, 2004) account has been motivated by the observation that the usual processing difficulty with ORs was significantly reduced when the RC head was a definite NP and the RC subject was a proper noun (e.g., *Bob*), *you*, or *everyone* as in (32), relative to the standard type of materials in which both the RC head

and RC subject were definite NPs as in (31). By contrast, OR difficulty persists when the RC head was a definite NP and the RC subject was an indefinite NP (e.g., *an accountant*), generic expression (e.g., *accountants*), or a semantically light NP (i.e., *the person*) as in (33).

(31) The salesman that the accountant contacted spoke very rapidly.

(32) The salesman that Bob/you/everyone contacted spoke very rapidly.

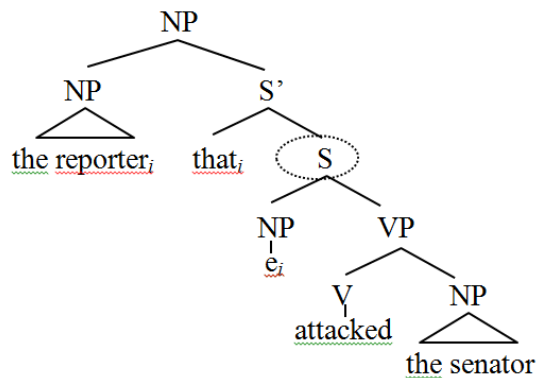
(33) The salesman that an accountant/accountants/the person contacted spoke very rapidly.

Based on this finding, Gordon et al. proposed that the main source of OR difficulty is the interference in working memory between two NPs that are similar at a psychologically relevant level of semantic properties. Specifically, both the first and second NPs in (31) and (33) (i.e., *the salesman*, *the accountant*, *an accountant*, *accountants*, and *the person*) contain a common noun and thus are assigned their semantic values in the same way (i.e., by restricting a set of entities denoted by a common noun), resulting in interference. On the other hand, the second NPs in (32) are interpreted in different ways than a common noun (i.e., by being bound by a linguistic antecedent or by a discourse referent), hence no interference. A cue-based retrieval theory of sentence processing as described in Van Dyke & Lewis (2003; Van Dyke, 2007) extends the range of linguistic features that can cause interference between similar elements, including morpho-syntactic (e.g., syntactic roles, number, etc.) and semantic properties (e.g., animacy) of NPs. A characteristic aspect of the theory of Van Dyke and her colleagues is that interference mainly arises when multiple items encoded in working memory are activated by partial matches to the syntactic/semantic cues provided by a newly integrated word (e.g., verb). This account readily applies to the SR-OR asymmetry. In the OR, the RC head and RC subject placed in working memory could both be activated by the cues provided by the RC verb, resulting in interference. In the SR, on the other hand, there is only one NP (i.e., RC head) in working memory when the RC verb is encountered, hence no interference.

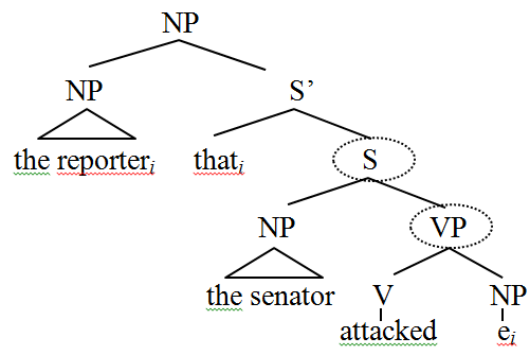
The accounts focusing on computational demands (e.g., Gibson, 2000) and interference (e.g., Gordon et al, 2001, 2004; Van Dyke, 2007) attribute processing difficulty to the presence of (certain types of) intervening material between the two elements to be integrated, which will often correlate with the linear distance between the elements. Other more syntactically-oriented accounts instead focus on ‘structural distance’ as measured by the number of intervening

maximal syntactic categories such as S and VP (e.g., O’Grady, 1997) or the size of the syntactic domain in which a filler-gap dependency is resolved (e.g., Hawkins, 1999). As shown in (34-35), there are more intervening maximal syntactic categories between the filler and the gap in an OR than in an SR, as illustrated with circles marked on the intervening maximal categories.

(34) Subject-extracted relative clause (SR)



(35) Object-extracted relative clause (OR)



The size of the relevant domain for filler-gap dependency resolution is determined by the number of ‘dependent’ arguments on the path that connects the filler (e.g., RC head) and its subcategorizer (e.g., RC verb). A dependent argument is defined as an argument on which the gap depends for being assigned its syntactic and semantic properties. The object is syntactically and semantically dependent on the subject because, for example, the syntactic license of the object requires the presence of the subject, and the object can be bound by the subject, whereas the opposites are not true (Hawkins, 1999). Therefore, the relevant filler-gap domain for an SR is smaller than for an OR as marked by the brackets in (36-37), because the OR has a dependent

argument (i.e., RC subject) between the RC head and RC verb but the SR does not. The larger the domain, the more difficult it is to process.

(36) [the reporter that *e* attacked] the senator

(37) [the reporter that the senator attacked *e*]

Another highly elaborated explanation of OR difficulty is the reanalysis-based account proposed by Traxler and colleagues (2002, 2005), which finds the main source of OR difficulty in the reanalysis process hypothesized to occur in ORs. This account is based on the garden-path model of sentence processing that assumes syntax-first serial processing (e.g., Ferreira & Clifton, 1986; Frazier, 1979, 1987a; Rayner et al., 1983). According to the account, a gap would be posited in the RC subject position as soon as the parser encounters the relative pronoun because it allows the earliest construction of the filler-gap dependency under the English grammar (*per* the Active Filler Strategy). This initial analysis is consistent with the ultimately correct analysis in an SR, causing no significant processing difficulty. By contrast, the presence of an overt subject NP in an OR forces the parser to discard the initial analysis and begin a process of reanalysis, increasing processing complexity.

The way in which the relative difficulty of a reanalysis process is determined naturally explains the effect of noun animacy on RC processing as shown in Traxler et al.'s eye-movement monitoring experiments (2002, 2005; see Betancort, Carreiras, & Sturt, 2009; Mak et al., 2002, 2006, for similar findings in Spanish and Dutch, respectively). In their study, the traditional OR difficulty as reflected in increased reading time and regressive eye movement was observed when the RC head was animate and the RC subject was inanimate as in (39). In contrast, the ORs with an inanimate RC head and an animate RC subject as in (41) were no more difficult than the corresponding SRs in (38) and (40).

(38) The director that watched the movie received a prize at the film festival.

(39) The director that the movie pleased received a prize at the film festival.

(40) The movie that pleased the director received a prize at the film festival.

(41) The movie that the director watched received a prize at the film festival.

Traxler et al. attribute this ‘animate subject advantage’ in OR processing to the general tendency for an animate entity to be a thematic agent/syntactic subject (and for an inanimate entity to be a thematic patient/syntactic object) (e.g., Clifton et al. 2003; Dowty, 1991). When the parser encounters the subject of an OR, it initiates a process of reanalysis, where it is hypothesized that the sentence processor evaluates various kinds of non-structural information such as the typical noun animacy – grammatical role association (e.g., Ferreira & Clifton, 1986). Thus it is hard to replace the mis-analyzed animate subject (e.g., director) with the correct inanimate subject (e.g., movie) in (39) because the former is a better subject by being animate. By contrast, it is relatively easy to replace the mis-analyzed inanimate subject (e.g., movie) with the correct animate subject (e.g., director) in (41) because the latter is a better subject.

The reanalysis-based account is further supported by a potential dissociation between processing resources that sub-serve initial syntactic decisions and those that sub-serve a later reanalysis process, which is evidenced in the role of working memory capacity (WMC) in RC processing as shown in Traxler et al. (2005). In the study, individual differences in WMC as measured by reading span did not correlate with individual differences in the processing difficulty of ORs relative to SRs when both head nouns and the nouns within the RCs were animate, which is consistent with the separate-resource view of WMC¹¹. On the other hand, the participants with higher reading span seemed to be better able to utilize helpful noun animacy information in processing the ORs with an inanimate head and an animate RC subject as in (41) than those with lower reading span did. Specifically, the higher reading span individuals showed a greater decrease of processing difficulty involved with ORs when the RC heads were inanimate and the RC subjects were animate, compared with the lower reading span individuals. Based on these findings, Traxler et al. (2005) proposed that earlier syntactic processes responsible for the OR difficulty draw on a separate pool of processing resources that are not measured by reading

¹¹ The dissociation between the individual differences in OR difficulty and those in WMC as shown in Traxler et al. (2005) is in contrast with the findings of King and Just (1991) and Fedorenko et al., (2006) that reading span is reliably correlated with OR penalty reflected in increased reading times as well as decreased comprehension accuracy (King & Just, 1991). As Traxler (2007) points out, it may be worthwhile to note here that many of the studies that showed a reliable correlation between reading span and online processing difficulty involved a dual task paradigm, in which participants are presented with items to remember while processing a given sentence. The two studies that showed reliable correlations between reading span and online OR difficulty also employed a dual-task paradigm. Therefore, it is a possibility that the processing requirement imposed by the need to switch attention between the two tasks might be responsible for the apparent association between reading span and online OR difficulty (e.g., Caplan & Waters, 1999). Given the ongoing debate on the nature of WM resources underlying sentence processing, the finding of the present dissertation might have some implications for this issue as discussed later.

span, whereas later considerations of semantic information in the course of reanalysis are subserved by the processing resources captured by reading span. This view could be seen as an extended version of the separate-resource view of WMC and sentence processing. The dissociation between the sets of processing resources underlying hypothetically different sub-processes involved in RC processing was in turn taken by the authors as evidence for their reanalysis-based account of RC processing.

Another group of accounts that can explain a wide range of empirical findings regarding SR-OR processing asymmetry is motivated in the context of the constraint satisfaction models of sentence processing (e.g., Gennari & MacDonald, 2008, 2009; MacDonald & Christiansen, 2002; Real & Christiansen, 2007; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). The general idea is that the surface order of major constituents in SRs (e.g., *the reporter that attached the senator*) resembles the canonical word order of simple declarative sentences (i.e., Subject–Verb–Object), which normal language users would encounter very frequently in everyday language use. However, the surface order of the major constituents in ORs (e.g., *the reporter that the senator attacked*) (i.e., Object–Verb–Subject) is much less frequent than that in SRs. Experience with many similar instances of the same word order pattern through overall input makes it easier to process SRs than ORs (MacDonald & Christiansen, 2002). Depending on the levels at which different types of RC constructions are described, this line of accounts also can explain the effect of semantic categories of RC subjects as shown in (31-33) and the effect of noun animacy as shown in (38-41). ORs with personal pronoun RC subjects (e.g., *the actor that you loved*) are shown in a corpus analysis to be more frequent than SRs with personal pronoun RC objects (e.g., *the actor that loved you*) (Real & Christiansen, 2007), which is consistent with the finding of reduced processing difficulty for ORs with personal pronoun RC subjects (e.g., Gordon et al., 2001, 2004). ORs with inanimate heads and animate subjects (e.g., *the movie that the director watched*) have been shown to be more common than those with animate heads and inanimate subjects (e.g., *the director that the movie pleased*) in corpus analyses and production tasks (Gennari & MacDonald, 2008, 2009; Roland, Dick, & Elman, 2007). These findings are consistent with the role of noun animacy in determining processing difficulty for ORs as shown in Traxler et al. (2002, 2005)

A recent account for the SR-OR processing asymmetry proposed by Staub (2010) capitalizes on the observation that there are at least two distinct sources of difficulty involved in

the processing of ORs, as identified by qualitatively different patterns of eye movements during sentence processing. One is the difficulty arising from the violation of expectations that have been entertained based on the input processed so far. According to expectation-based accounts of sentence processing, processing difficulty at a certain word in a sentence is an inverse function of the likelihood of encountering that word given the previous string of words (e.g., Hale, 2001; Levy, 2008). Since SRs are more frequent than ORs, an SR construction is expected when a relative pronoun is encountered. Thus processing difficulty arises when a comprehender encounters the RC subject following the relative pronoun in an OR, due to the violation of the expectation (e.g., Forster et al., 2009; Staub, 2010). The other source of processing difficulty is the cognitive cost due to integrating the RC verb with its arguments, as predicted by the integration component of Gibson's (2000) theory and the interference-based accounts of Gordon et al. (2001, 2004) and Van Dyke and Lewis (2003). Note that these accounts predict that the processing difficulty of ORs will mainly emerge at the RC verb, when the head nouns should be retrieved from memory and be integrated with the RC verbs. The results of Staub's (2010) eye-tracking experiments suggest that processing difficulty arises at both the RC subject and RC verb in ORs but in different eye movement measures at each point: the difficulty at the subject of the ORs took the form of increased regressive eye movements while the difficulty at the verb of the ORs was reflected in increased reading times. Staub interpreted the results as suggesting that the sources of difficulty at each point are attributable to different kinds of processes that are hypothesized to be related to the violation of expectations and the cost associated with integration.

Some accounts of the SR-OR asymmetry provide more adequate explanations of empirical findings than others. For example, earlier accounts such as the Noun Accessibility Hierarchy (Keenan & Comrie, 1977), the perspective shift account (MacWhinney & Pleh, 1982), the parallel function hypothesis (Sheldon, 1974), and the augmented transition network model of sentence processing (Wanner & Maratsos, 1978) explain the general processing asymmetry between SRs and ORs, but need to be further elaborated to account for the modulation of the OR difficulty by referential properties and animacy of the relevant NPs. The same is true of the accounts based on structural distance (O'Grady, 1977) and size of the filler-gap domain (Hawkins, 1999), which do not explicitly take into account the semantic properties of the relevant NPs.

The integration- and the interference-based accounts readily explain the effect of referential properties of the RC subject on RC processing as shown in (31-33). But they do not readily explain the noun animacy effect as shown in (38-41) because the two relevant NPs in these sentences are both new discourse referents and distinct in animacy, resulting in the prediction of comparable processing difficulty between the ORs with different animacy configurations. On the other hand, the reanalysis-based account based on the garden-path theory of sentence processing readily explains the noun animacy effect, as well as the effect of referential properties of the RC subject if we assume that, for example, indexical pronouns (i.e., *I* and *you*) are a more frequent and thus better subject than definite descriptions of animate entities (e.g., *the salesman*). Frequency-based accounts based on the constraint satisfaction model of sentence processing also seem to be able to explain a wide range of empirical findings as long as frequency counts can be obtained at appropriate levels of specificity. After all, these different accounts may provide better explanations for different aspects of the cognitive processes involved in RC processing, and we may need more than one account that focus on different sources of difficulty motivated independently by empirical evidence, as Staub's (2010) approach suggests.

A thorough evaluation of the various accounts of SR-OR asymmetry is beyond the scope of the present dissertation, although relevant findings from native speakers will be discussed to such effect. Following Staub (2010), the present dissertation predicts that there may be multiple sources of difficulty that operate in tandem to bring about the SR-OR asymmetry and other related effects such as the effect of semantic properties of nouns. Staub (2010) identified two distinct processes responsible for the OR difficulty based on their differential effects on eye movement patterns: (i) the violation of expectations at the RC subject resulting in processing failure that is associated with regressive eye movements and (ii) the memory cost incurred at the RC verb resulting in longer processing times. Although he invoked the expectation-based accounts of sentence processing to explain the difficulty arising at the RC subject of an OR (e.g., Hale, 2001; Levy, 2008), the phenomenon is also consistent with the predictions of the reanalysis-based (e.g., Traxler et al., 2002, 2005) and the frequency-based accounts (e.g., Gennary & MacDonald, 2008, 2009). The common aspect of these accounts regarding the difficulty at the subject of an OR is that they incorporate certain kinds of prediction-based processes that bind NPs to available syntactic positions in the current parse before a lexical

thematic role assigner (i.e., RC verb) is encountered. In this sense, the processes associated with processing difficulty at the RC subject could be taken to be structure-driven. On the other hand, the processing cost incurred by integration and interference would appear at the RC verb when it must be integrated with the RC head (e.g., Gibson, 2000; Gordon et al., 2001, 2004; Van Dyke & Lewis, 2003). The difficulty that is observable at the verb of an OR therefore could be attributed to largely lexically-driven processes in the sense that they are triggered by a lexical subcategorizer.

To summarize, the processing difficulty arising before the RC verb in an OR can be attributed to ‘structurally-based’ processes while the processing difficulty observed at and after the RC verb can be attributed to ‘lexically-triggered’ processes. This distinction between the sources of difficulty is important for the present purposes because the main question of the present dissertation is whether the structural computation in L2 sentence processing is carried out by the same principles as in L1 sentence processing. If adult L2 learners exclusively rely on the presence of lexical items in computing the syntactic structure of an L2 sentence as a strong version of the shallow structure hypothesis would predict, a logical prediction would be that the OR penalty in L2 sentence processing can only occur after the RC verb is encountered, because only then the preceding nouns can be integrated into the current parse and thus all the difficulty involved in that integration process can arise.

4.1.1. Processing relative clauses in Korean

The present dissertation targets L1-Korean L2-English learners processing English RCs. In this section, previous findings are reviewed on the processing of Korean RCs by native speakers of Korean, in order to consider potential influences of L1 processing routines on L2 sentence processing. A set of experiments conducted by Kwon (2008) on SR-OR asymmetry in RC processing in Korean suggests that native speakers of Korean find it more difficult to comprehend ORs than SRs. Although RCs in English and Korean both show an object penalty in processing, there are some structural differences between them. Since those differences could be a potential source of L1 influence and/or increased processing difficulty for Korean speakers reading English RCs, it will be helpful to review previous empirical findings and proposed theoretical accounts for L1 processing of Korean RCs.

Korean is a Subject-Object-Verb language with relatively free surface order of non-verbal constituents primarily due to a rich case marking system, whereas English is a Subject-Verb-Object language with relatively fixed word order. Related to the difference in the canonical word order, RCs in Korean precede the head noun, whereas those in English follow it. RCs in Korean are marked by an adnominalizer *-n/-nun* attached to the RC verb, whereas English uses relative pronouns for the same purpose. The Korean equivalents of English SRs and ORs in (42a) and (42b) are presented in (43a) and (43b), respectively.

(42) a. The reporter_{*i*} [that *e_i* attacked the senator] admitted the error in the public.

b. The reporter_{*i*} [that the senator attacked *e_i*] admitted the error in the public.

(43) a. [*e_i* uywen-ul kongkyekha-n] kica-ka_{*i*} kongkaycekulo silswulul incenghayssta
senator-acc attacked-adn reporter-nom in-the-public error-acc admitted

b. [uywen-i *e_i* kongkyekha-n] kica-ka_{*i*} kongkaycekulo silswulul incenghayssta
senator-nom attacked-adn reporter-nom in-the-public error-acc admitted

It has been under debate how to describe the syntax of Korean RCs (see, for example, Kwon, 2008, pp. 39–64, for a critical review of different syntactic accounts). Some accounts treat them as simple noun-modifying clauses that are the same in syntactic characteristics as normal clauses (e.g., Yoon, 1995). Other accounts acknowledge a non-trivial syntactic derivation by postulating the presence of a null operator that binds the unrealized constituent within the RC. These accounts further diverge on whether the null operator is hypothesized to move from the original syntactic position of the unrealized constituent (e.g., Han, 1992) or to be base-generated with the bound constituent being dropped (e.g., Kang, 1986).

Orthogonal to the theoretical debates on the syntactic description of the Korean RC, and despite the surface differences between the Korean and the English RCs, empirical findings on real-time processing of Korean RCs are consistent with the findings on English RCs in that ORs are harder to comprehend than SRs, as reflected in decreased accuracy in comprehension and increased reading times (e.g., Kwon, 2008). Kwon (2008) evaluated several accounts of the OR penalty against this finding in Korean. Since most of the accounts have been motivated based on English, some accounts make an empirically unsupported prediction that SRs would be more difficult than ORs (i.e., SR penalty) and other accounts cannot be readily applied to Korean RCs.

For example, it has been pointed out by several researchers that the linear-distance-based theories (e.g., the integration- and interference-based accounts) predict for languages with prenominal RCs that SRs will be more difficult to comprehend than ORs because the distance between the head noun and the gap in the RC is greater in SRs than in ORs as shown in (43). This prediction contradicts the empirical findings of, for example, Kwon (2008), Ueno and Garnsey (2008), and Lin and Bever (2006) on Korean, Japanese and Chinese, respectively (but see Ishizuka, Nakatani, & Gibson, 2006; Lin & Garnsey, 2010, for different results on Japanese and Chinese). Similarly, accounts based on the temporary ambiguity in the grammatical/thematic role of the head noun (e.g., the reanalysis-based account of Traxler et al., 2002, 2005, and the frequency-based account of Gennari and MacDonald, 2008) may not be directly applicable to Korean RCs because they precede the head noun and thus the syntactic role of the RC head is not ambiguous when it is encountered (but see Gibson & Wu, 2011, and Ishizuka, Nakatani, & Gibson, 2006, for evidence for a potential role of structure-level ambiguities in prenominal RC processing). For these reasons (and other considerations), Kwon (2008) concluded that the structural-distance account of O'Grady (1997) and the noun accessibility hierarchy of Keenan & Comrie (1979) fit best to the findings on Korean RC processing, both of which predict a cross-linguistic preference for subject extraction over object extraction in RC processing.

How would these observations relate to Korean speakers' L2 processing of English RCs? One possibility is that Korean-speaking learners of English may not have particular difficulty in learning to process English RCs in the same way as English native speakers process English RCs, because the same pattern of SR-OR asymmetry as shown in previous studies on English and Korean RCs might suggest that RC processing in English and Korean involves the same representations and processes at some stages of comprehension, which should be readily transferable between languages. But it is also true that the different surface structures of the RCs between the two languages require that the incremental construction of the relevant syntactic structures proceed in different ways in each language at some level(s), as could be illustrated by different predictions made for each language by those accounts that take into account language-specific surface structures (e.g., the linear-distance-based accounts). Therefore, it will be worthwhile to investigate whether adult L1-Korean L2-English learners process English RCs through the same processes as native English speakers do, which will ultimately contribute to understanding whether adult L2 learners use qualitatively different processes to comprehend

sentences in the L2 as compared with native speakers of the target language. In addition, testing Korean learners would also be relevant to the question of the role of typological distance between the native and target language of L2 learners in the acquisition of native-like processing strategies because the majority of the currently available data about online L2 processing of RCs comes from populations whose L1 and L2 are typologically closer than English and Korean, as will be reviewed in the next section.

4.2. Processing relative clauses in L2

Previous work on L2 comprehension of RCs in, for example, English, Dutch, Japanese and Korean suggests that L2 learners generally show a preference for SRs over ORs as the native speakers of those languages both in accuracy of final comprehension and in speed of online processing (e.g., Hashimoto, 2007; Izumi, 2003; Jackson & Roberts, 2010; Kanno, 2007; Lee, 2009; O'Grady, Lee, & Choo, 2003; Omaki & Arijj, 2005). Some studies further suggest that lexical-semantic information such as noun animacy can enhance L2 learners' sensitivity to the structural distinction between SRs and ORs (e.g., Jackson & Roberts, 2010) especially for those with limited proficiency (Kanno, 2007). Among these, Omaki and Arijj's study (2005) on L2 English and Jackson and Roberts' study (2010) on L2 Dutch are directly relevant to the present dissertation because they explicitly aimed to test the interaction of structural and non-structural information in L2 sentence processing.

Modeled on Traxler et al.'s (2002) study on the modulating effect of noun animacy on OR penalty, Omaki & Arijj (2005) had adult L1-Japanese L2-English learners with advanced L2 proficiency read and rate the perceived complexity of sentences containing SRs and ORs with different animacy configuration as in (38)–(41), repeated below as (44)–(47).

- (44) The director that watched the movie received a prize at the film festival.
- (45) The director that the movie pleased received a prize at the film festival.
- (46) The movie that pleased the director received a prize at the film festival.
- (47) The movie that the director watched received a prize at the film festival.

The results showed that the L2 learners found ORs more difficult than SRs when the heads were animate and the RC-internal NPs were inanimate as in (44) and (45), whereas they

found ORs with inanimate heads and animate subjects as in (47) as easy to comprehend as SRs with the same or different animacy configurations as in (44) and (46), as reflected in higher complexity ratings for sentences like (45) than for other sentences like (44), (46) and (47), which were rated to be comparable in complexity. Assuming that the OR-penalty is the result of employing a ‘structure-based’ parsing strategy (i.e., the active-filler strategy), and that the modulating effect of noun animacy (i.e., animate subject advantage) is evidence for the interaction of lexical-semantic information and syntactic structure, Omaki and Arijji interpreted the results as evidence that L2 learners’ sentence processing is guided not only by lexical-semantic information but also by structural information.

Modeled on Mak et al.’s (2002; 2006) study on L1 processing of Dutch RCs, Jackson & Roberts’ (2010) self-paced reading experiment with L1-German L2-Dutch learners also tested the interaction of syntactic structure and noun animacy in L2 processing of RCs using sentences as in (48)–(51). In addition, by capitalizing on the fact that the syntactic roles of the head noun and the RC-internal noun can be kept ambiguous until they are disambiguated by the number agreement marked on an auxiliary verb that precedes the clause-final RC verb, they also tested whether adult L2 learners immediately incorporate morpho-syntactic information (i.e., number agreement between the subject and the auxiliary verb) to make immediate parsing decisions before a lexical verb is encountered.

(48) Voor de kinderen is de clown, die de taarten heft gegooit, het hoogtepunt van de voorstelling. (SR, animate head–inanimate RC-internal noun)

For the children is the clown_{SG} that the pies_{PL} has_{SG} thrown the highlight of the performance

“For the children the clown, that threw the pies, was the highlight of the performance.”

(49) Voor de kinderen is de clown, die de taarten hebben geraakt, het hoogtepunt van de voorstelling. (OR, animate head–inanimate RC-internal noun)

For the children is the clown_{SG} that the pies_{PL} have_{PL} hit the highlight of the performance

“For the children the clown, that the pies hit, was the highlight of the performance.”

(50) Voor de kinderen zijn de taarten, die de clown heft gegoooid, het hoogtepunt van de voorstelling. (OR, inanimate head–animate RC-internal noun)

For the children are the pies_{PL} that the clown_{SG} has_{SG} thrown the highlight of the performance

“For the children the pies, that the clown threw, were the highlight of the performance.”

(51) Voor de kinderen zijn de taarten, die de clown hebben geraakt, het hoogtepunt van de voorstelling. (SR, inanimate head–animate RC-internal noun)

For the children are the pies_{PL} that the clown_{SG} have_{PL} hit the highlight of the performance

“For the children the pies, that hit the clown, were the highlight of the performance.”

The reading time results showed basically the same pattern of interaction between syntactic structure and noun animacy as observed in the complexity rating results in Omaki & Ariji (2005). The results further revealed a trend that the L2 learners showed a processing advantage to SRs over ORs as soon as the subject-verb agreement information was available at auxiliary verbs. The authors interpreted this finding as suggesting that adult L2 learners can make immediate syntactic decisions based on purely formal features such as subject-verb agreement at least when rich lexical-semantic information is available (e.g., helpful animacy information). This finding is not consistent with the shallow structure hypothesis (SSH), according to which adult L2 learners are not expected to transfer the structure-based parsing principles in the L1 to the L2 even when the two languages have many structural properties in common as in the case of the Jackson and Roberts (2010) study.

To summarize, both Omaki and Ariji (2005) and Jackson and Roberts (2010) point to the possibility that adult L2 learners are sensitive to structural features of different types of RCs (i.e., SRs and ORs) as well as to lexical-semantic properties of relevant nouns and verbs during online processing. Thus the results from the studies on L2 processing of RCs overall could be taken as evidence for the view that sentence processing in L1 and L2 is not qualitatively different from each other. However, a clearer picture is yet to be drawn on the issue (i) because Omaki and Ariji’s study is based on an offline measure of perceived complexity and thus does not allow a direct inspection of incremental processing of RCs by L2 learners of English and (ii) because the evidence from Jackson & Roberts’ study for L2 learners’ immediate sensitivity to agreement

information on auxiliary verbs was obtained from the learner population whose L1 and L2 are closely related (i.e., English and Dutch), leaving its generalizability yet to be empirically tested. The present study therefore further investigates the online processing of English RCs by adult L2 learners whose L1 is typologically dissimilar to the target language (i.e., Korean) using an online sentence processing task (i.e., self-paced reading), in order to produce additional evidence regarding the nature of L2 sentence processing.

For the purposes of the present dissertation, processing difficulties due to structurally- vs. lexically-driven processes, as discussed in Section 4.1, could be used as diagnostics on the nature of L2 processing of English RCs. If L2 sentence processing is carried out incrementally incorporating various kinds of information including possible syntactic structures at a given moment, we would expect L2 comprehenders to also show signs of processing difficulty that is associated with structural processes even before they encounter the verbs of ORs, as predicted by, for example, the reanalysis-based account. If, on the other hand, sentence processing in an L2 proceeds primarily by establishing lexical relations between lexical subcategorizers and their arguments without considering structural configurations, the OR difficulty in L2 sentence processing would mainly appear at the RC verb when the lexical relations can be established.

A potential criticism on interpreting the results from L2 processing of RCs as evidence for the qualitative similarity between L1 and L2 processing is that the sensitivity to the structural differences between SRs and ORs could be attributed to simple word order information. ORs in English and Dutch, which are post-nominal, involve a non-canonical word order in that the object precedes the subject, whereas SRs maintain the canonical SVO and SOV word order for English and Dutch, respectively (without considering the presence of the relative pronoun). It is easy to suppose that non-canonical word order would cause processing difficulty without the need to invoke complex syntactic-processes or structure-based parsing strategies (e.g., MacDonald et al., 1994). For this reason, it could be argued from the perspective of the SSH that L2 learners' sensitivity to the SR-OR asymmetry does not qualify as evidence for the ability to construct detailed, hierarchically organized syntactic structures.

It is possible that the investigation of RC processing may not directly speak to the question of whether L2 syntactic representations contain any hierarchical relationships in a strict sense, because the RC 'modifies' its head noun and thus they do not form such a syntactic relation as would be postulated to describe the relation between a lexical subcategorizer and (an

element within) its syntactic dependents (e.g., the relationship between the two NPs in a complex NP such as *the servant of the actress*, for which the SSH assumes a hierarchical configuration that L2 learners do not project). However, comprehension of RCs involves more than mere recovering of non-canonical word orders: The processing of the main clause should be temporarily suspended while processing the RC, and two different propositions from the main clause and the RC should be combined. And these processes incur a substantial degree of processing difficulty that is readily detectable through various behavioral measures even for native speakers. Therefore, we could assume that the processing of RCs involves a set of non-trivial, complex processes, at least part of which should be structural in nature. Especially if we assume an account of RC processing that acknowledges the primary role of syntactically-based processes such as an initial structural preference for SRs over ORs and a later reanalysis process in ORs (i.e., the reanalysis-based account), the processing difficulty observed during the course of RC processing could be taken to reflect these ‘complex’ syntactic processes.

But it is still unclear whether the structural processes involved in the processing of RCs are complex enough to qualify as a test case for the SSH, partly because the hypothesis is not clear enough about the issue. So the present dissertation also tests the role of WMC in L1 and L2 sentence processing as a complementary way to investigate the nature of the L2 sentence processing mechanisms. As mentioned in Section 4.1, Traxler et al.’s (2005) study with native speakers of English suggests a potential distinction between sets of processing resources responsible for largely syntactic processes as evidenced by the OR penalty and those responsible for semantically-based processes as evidenced by noun animacy effects, as reflected in the pattern of associations/dissociations between the two types of processes and individual differences in WMC. If we assume a modular distinction between structurally-based and semantically-triggered processes involved in L1 sentence processing, we can investigate the nature of different sub-processes involved in L2 sentence processing by testing whether and how they are modulated by L2 WMC. Note that this question is orthogonal to the issue of whether RCs represent a complex enough syntactic structure to test the fundamental similarities or differences between L1 and L2 sentence processing. If we see a dissociation between WMC and syntactic complexity effects on the one hand and a reliable association between WMC and noun animacy effects on the other during the course of L2 processing of RCs, it would suggest that some part of the L2 sentence processing mechanism is dedicated to structural processing as in L1

sentence processing. This would not be consistent with the view of the SSH that L2 learners over-rely on non-structural information due to a domain-specific deficit in syntactic components of the L2 sentence processing mechanism, regardless of the degree of complexity of the tested structure. Even if one does not embrace a modular view of sentence processing, dissociations between WMC and some sub-processes involved in L2 sentence processing still could be taken as evidence for a high level of automatization of those processes, as suggested by the WM literature (e.g., Conway et al., 2005). In contrast, if it were the case that individual differences in syntactic complexity effects are associated with individual differences in WMC, then it would suggest that L2 sentence processing is carried out largely by general-purpose cognitive processes, consistent with the view of the SSH that L2 learners rely more on lexical-semantic information and extra-linguistic comprehension strategies in L2 sentence processing.

Chapter 5

Overview of the present study

This chapter presents a general overview of the present experiments. In Section 5.1, the common aspects of the three experiments with native English speakers and the three experiments with L1-Korean L2-English learners are summarized. In Section 5.2, the results of the WM span tests are reported in order to explore the general characteristics of the measures as administered to native speakers and L2 learners of English.

5.1. Overview of the experiments

The present dissertation consists of three experiments with native English speakers and three parallel experiments with adult L1-Korean L2-English learners in which the participants read sentences containing an SR or an OR on a word-by-word basis in a self-paced non-cumulative moving-window paradigm. Animacy of the nouns relevant to the interpretation of RCs (i.e., RC head nouns and RC-internal nouns) was systematically manipulated across the experiments to test how adult L2 learners of English compare to native speakers of the language in terms of sensitivity to syntactic structure and lexical-semantic information during online sentence processing. Participants' WMC measures were also collected using a reading span and a subtract-2 span test in order to compare the pattern of associations/dissociations between individual differences in WMC and individual differences in the sensitivity to different kinds of information available during sentence processing.

The first experiments with native English speakers (Experiment 1E) and with L2 learners (Experiment 1K) were modeled on Traxler et al.'s (2005) Experiment 1, and aimed to test (i) processing difficulty due to syntactic complexity and (ii) relationships between WMC and syntactic complexity effects in L1 and L2 sentence processing.

Experiment 2E and 2K were modeled on Traxler et al.'s (2005) Experiment 2 and tested (i) how noun animacy information modulates syntactic complexity effects and (ii) whether individual differences in WMC predict individual differences in the sensitivity to semantic information during sentence processing. Experiment 1E and 2E were conducted together as were Experiment 1K and 2K, by incorporating the experimental sentences from each experiment into one presentation list. Consequently, the participants for Experiment 1 and 2 were the same for both native English speakers and L2 learners.

Experiment 3E and 3K were conducted separately from the other experiments and were designed to further test the role of noun animacy information and WMC in RC processing.

The materials and procedure of Experiment 1E, 2E and 3E were identical with those of Experiment 1K, 2K and 3K, respectively, except that there were additional tasks for L2 learners, including a cloze test, a vocabulary check, and a background questionnaire. The cloze test was administered to obtain preliminary measures on L2 learners' general proficiency in English (Brown, 1980). The vocabulary check was to make sure that the critical nouns and verbs in the test sentences are known to L2 learners. The background questionnaire included questions probing biographical information as well as learners' experience with the English language.

For ease of presentation and discussion, the experiments with native speakers (Experiment 1E, 2E and 3E) and those with L2 learners (Experiment 1K, 2K and 3K) are reported separately in Chapter 6 and 7, respectively. This is to first establish the native speaker norm, based on which to evaluate L2 sentence processing. The overall methods for the six experiments will be presented in the following sections because the experiments have much in common in their materials, methodologies and participants.

5.1.1. Participants

Adult native speakers of English and adult L1-Korean L2-English learners participated in the present experiments in exchange for extra credit in introductory courses in linguistics or for payment. Forty-eight college-aged native speakers of English participated in Experiments 1E and 2E, with four excluded from analysis due to incomplete data (i.e., no subtract 2-span score). A total of 46 college-aged native English speakers participated in Experiment 3, with one participant excluded from analysis due to low comprehension accuracy (i.e., lower-than-60% mean accuracy for the test sentences). Five participants in Experiment 3E also participated in Experiments 1E and 2E at least 1 month earlier. Therefore, a total of 84 native speakers of English contributed data to the present study.

Forty-four adult L1-Korean L2-English learners participated in Experiments 1K and 2K, with 10 of them excluded from the analysis of the self-paced reading time results due to low comprehension accuracy (i.e., lower-than-60% accuracy for the test sentences) and additional three participants were excluded due to incomplete data (i.e., no cloze test score for two participants and no WM span test scores for one participant), leaving 32 participants to contribute data to Experiments 1K and 2K. Five of the participants did not take the subtract-2

span test and thus their reading span scores were taken as representing their WMC. They were included to increase statistical power, but the analysis without these participants produced the same pattern of results. Note that the participants with no subtract-2 span score were excluded from analysis in the other experiments (i.e., four participants in Experiments 1E/2E and one participant in Experiment 3K). Thirty-seven L2 learners participated in Experiment 3K, among which 25 also participated in Experiments 1K and 2K at least 1 month earlier. Six participants were excluded from analysis due to low comprehension accuracy on the test sentences (N=4) and incomplete data set (N=2), leaving 31 participants to contribute data to Experiment 3K. Eighteen of the 31 participants in Experiment 3K also contributed data to Experiments 1K and 2K. After all, 32 participants were analyzed in Experiments 1K/2K and 31 in Experiment 3K, among which 18 participated in all the three experiments.

Table 1 summarizes the characteristics of the L2 learners who contributed data in Experiments 1K/2K and Experiment 3K. Note that those participants who participated in the both experimental sessions (Session 1 for Experiments 1K and 2K, Session 2 for Experiment 3K) took the cloze test and the background questionnaire only at the first visit, which resulted in the repeated cloze test score for Experiment 1K/2K and for Experiment 3K in Table 1. The learner groups in the two sessions of the experiments were quite similar along all the dimensions for which data was obtained, reflecting the fact that 18 out of 31 participants in Experiment 3K also participated in Experiment 1K and 2K. Though with a rather large variance in the immersion experience, the L2 learner participants can be roughly characterized as adult L2 learners in the sense that they all began to learn English as a foreign language in school settings in Korea mostly around 10 years of age and came to the US after the age of 13 to study abroad or to immigrate.

Table 1 Characteristics of L1-Korean L2-English learner participants

	Age (year)	Instruction began at (year)	Came to US at (year)	Length of Immersion (month)	Cloze test score (out of 40)
Experiment 1K/2K	26.1 (20~44)	10.8 (6-13)	20.3 (13-33)	62.3 (1~237)	27.8 (10~37)
Experiment 3K	25.6 (20~44)	10.5 (4~14)	20.4 (14~37)	63.5 (6~240)	28 (18~37)

5.1.2. Overall procedure and tasks

The participants were tested individually in a quiet room. They first performed the self-paced reading task and the WM span tests, which was followed by a post-hoc plausibility rating task on the experimental materials. The L2 learner participants also completed a cloze test, as well as a vocabulary check list on which they reported unknown words among the two nouns relevant to the interpretation of RCs, and the verbs in RCs and main clauses. The experiments ended with a background questionnaire. The entire procedure took 40-50 minutes for native English speakers and 70-90 minutes for L2 learners. The procedure for each of the major tasks is presented in the following sections.

5.1.2.1. Self-paced reading task

Participants read the test sentences in the self-paced non-cumulative word-by-word moving window paradigm (Just, Carpenter, & Wooley, 1982). The sentence reading task was implemented using the Linger software developed by Doug Rohde (available online at <http://tedlab.mit.edu/~dr/Linger/>). A row of dashes masking each letter in each word of a sentence initially appeared on the screen and each press of the spacebar on the keyboard revealed one word at a time from left to right with all the other words remaining masked by the dashes. The time lapse between the two adjacent events of key press was taken to be the reading time for the word on the screen. The key press at the last word of a sentence marked with a period led to a comprehension verification statement for every test sentence and about half of the filler sentences. The participants were instructed to read the sentences for meaning and informed that they would be checked for comprehension of some of the sentences. The self-paced reading task proper followed eight practice trials. Native speaker participants took around 20 minutes and L2 learner participants took around 30 minutes to complete the task.

5.1.2.2. Reading span test

The reading span test developed in Conway et al. (2005) was used for the present study. The test consisted of 42 sentence–letter pairs that are grouped into 12 sets of different sizes. About half of the sentences were semantically plausible (e.g., *When I get up in the morning, the first thing I do is feed my dog*) while the remaining sentences did not (e.g., *After yelling at the game, I knew I would have a tall voice*). Each set contained two to five sentence-letter pairs, with every three sets containing an equal number of the pairs. The sentence–letter pairs were presented on a computer screen one by one, with a question mark (?) appearing between the

sentence and the letter to prompt participants to respond to the semantic plausibility of the sentence. Participants read the sentence out loud, indicated whether it made sense or not by saying yes or no, and then read out loud the following letter. They pressed the spacebar to proceed to the next trial. A recall prompt (i.e., ???) was presented after participants finished reading all the sentence–letter pairs in a set, when they wrote down the preceding letters in order of presentation on a provided answer sheet. Three practice trial sets with two sentence–letter pairs preceded the task. The presentation order of the stimuli sets was randomized to prevent the participants from expecting how many trials they are going to encounter in a particular set. The total number of letters recalled in correct serial position was counted and translated into percentage to represent each participant’s reading span score.

The accuracy of semantic judgments on the stimulus sentences was recorded by the experimenter during the procedure to make sure that the participant paid attention to the processing component of the test. All the native speakers performed at ceiling on the processing task with the accuracy rate higher than 90%, indicating they indeed paid attention to interpreting the sentences. The L2 learner participants were not as accurate as native speakers with the accuracy rate ranging from 59.5% to 100% ($M = 85.1$, $SD = 9.9$), reflecting their lower proficiency in reading comprehension. Although a standard practice in the administration of a reading span test is to exclude participants with processing performance worse than a threshold level to make sure that sufficient attention has been paid to the processing task (e.g., 80% accuracy as adopted in Turner & Engle, 1989), any L2 learners were not excluded in the present dissertation based on such a criterion because it could be due to poor reading comprehension ability rather than a lack of attention. This possibility is supported by the fact that the L2 learners’ semantic judgment scores were correlated with their cloze test scores more strongly than their reading span scores, as will be shown in Section 5.2.

5.1.2.3. Subtract-2 span test

The subtract-2 span test consisted of 54 digits that were grouped into 12 sets. The number of digits in a set increased from two to seven and every two sets had the same number of digits. The digits in a set were automatically presented one by one on the center of a computer screen for one second for each digit, which the participant read out loud. The presentation order of the digit sets was randomized for each participant. After a set of digits was presented, blank boxes of the same number as the preceding digits appeared on the screen. Participants subtracted two from

each of the preceding digits and typed in the resulting digits into the blank boxes in order of presentation. Participants were instructed to either guess or skip the digit they could not recall for sure. Two practice trial sets with two digits preceded the task. The total number of the digits that were correctly subtracted and entered was counted and translated into percentage to represent each participant's subtract-2 span score.

5.1.2.4. *Post-hoc plausibility test*

As an attempt to control for potential differences in plausibility of the test sentences across conditions in Experiments 2E/K and 3E/K, participants rated the plausibility of the simple propositions derived from the RCs and the main clauses of the test sentences after they finished the self-paced reading task. From the test sentence *The director that didn't praise the movie received a prize at the film festival*, for example, the two propositions 'the director didn't praise the movie' and 'the director received a prize at the film festival' were derived. A seven-point scale was used for rating, with 1 representing "Least plausible" and 7 "Most plausible". The materials and results of the post-hoc plausibility ratings are reported in the materials section of each experiment. Experiment 1 did not administer a separate plausibility test because plausibility was controlled for by reversing the meaning of the propositions expressed by the RCs and main clauses within each set of test sentences, as will be detailed in the materials and design section of Experiments 1E/K).

5.1.2.5. *Cloze test*

The cloze test developed by Brown (1980) was administered with some modifications to obtain proficiency measures from L2 learner participants. The original test was adapted from a commercial reader for L2-English learners (Kurilecz, 1969), which consisted of a 406-word passage with every seventh word replaced with a blank. There are 50 blanks in the passage with two lead-in and one lead-out sentences left intact to provide a context. The testee is instructed to provide only one word for each blank to complete the passage. For the present study, the 11 blanks in the first paragraph of the passage were replaced with appropriate words and the main verb of the last sentence was replaced with a blank, leaving 40 blanks to fill, in order to reduce the time taken to complete the test (See Appendix II). A contextually appropriate and grammatically correct (i.e., acceptable) word provided for each blank was deemed correct and the percentage of correct answers was calculated as an individual's cloze test score. Spelling was not considered in scoring.

5.1.3. Data analysis

Independent variables in the present study were WMC and the experimentally manipulated factors in the respective experiments (i.e., RC type in Experiment 1E/1K, and RC type and noun animacy configuration in Experiments 2E/K and 3E/K). WMC for each individual was obtained by averaging his or her reading span and subtract-2 span scores. Thus WMC was treated as a continuous covariate while the other factors were categorical with two levels each. For L2 learners, cloze test scores were also treated as an independent variable.

The dependent measures analyzed in each experiment were accuracy for the comprehension verification statements and reading times at pre-determined critical regions of the test sentences. The data from the trials that contained unknown words to L2 learners as reported in the vocabulary check were excluded from analysis, removing seven trials in Experiment 1K, 5 trials in Experiment 2K and 6 trials in Experiment 3K. Responses to the comprehension verification statements were coded as a binary variable with the value of “correct” and “incorrect”. The reading times for all the test sentences were analyzed regardless of whether the responses to the comprehension verification statements were correct or not¹². The reading time data was analyzed at three critical regions. The RC region included all the words within the RC excluding the complementizer *that*. For this region, mean reading time per word was analyzed. The second critical region consisted of the word immediately following the RC region and was labeled as the main clause verb (MV) region. Note that the MV region did not always consist of the main lexical predicate of the main clause, often containing an auxiliary verb such as *be* or a negative contraction such as *didn't* and *couldn't* (See Appendix I). The last region was the word following the MV region, which was labeled as MV+1 region. This region was analyzed to detect any delayed effects. When the ORs with different animacy configurations were compared in Experiments 2E/K and 3E/K, reading times at each word within the RC were analyzed. Reading time data was trimmed by removing the data points below and above three standard deviations from the grand mean of all the data points available for each region under analysis, affecting about 2% of the data on average.

¹² It is conventional to remove such trials in the analysis of reading times. But the incorrectly responded sentences were also included for analysis under the assumption that those items would still reflect comprehension difficulty. The analysis only with the items whose comprehension statements were correctly verified yielded a similar pattern of results for both native speakers and L2 learners.

The statistical reliabilities of the comprehension accuracy and reading time data were tested based on logistic and linear mixed effects models, respectively (Baayen, Davidson, & Bates, 2008; Jaeger, 2008), as implemented in the lme4 statistical package (Bates, Maechler, & Dai, 2011) running on R, a programming language for statistical analysis (R Development Core Team, 2011). Mixed effects modeling can include both participants and items as random effects simultaneously, avoiding the problem of treating items or participants as a fixed variable in the course of aggregating data points across participants or items in the traditional statistical tests such as analysis of variance.

The main effect and interaction terms for WMC and sentence manipulations (i.e., RC type and noun animacy) were included as the fixed effects in all the models. RC type and noun animacy were coded using mean-centered contrast coding. The continuous covariates (i.e., WMC and cloze test scores) were mean-centered. This way of coding yields more interpretable statistical test results because the estimated coefficient for each factor represents the magnitude of its main effect when the model includes more than one factor. The maximal random effects structure for each model was determined based on the log-likelihood ratio test (Baayen et al., 2008). In the analysis of L2 learner data, the potential effects of proficiency in English was explored by testing whether including cloze test score as a predictor increased the fit of the models to the data using the log-likelihood ratio test. Specific structures of the fitted models are presented together with the coefficients for the respective models.

It is often acknowledged that it is sometimes problematic to derive exact p-values for the fixed effects estimated by mixed effects models because for logistic regression models, the Wald-Z statistic from which p-values are calculated is vulnerable to the presence of collinearity among predictors (e.g., Jaeger, 2008) and for linear models, it is hard to decide on the desirable degrees of freedom for the t-statistic based on the standard errors of the estimates (e.g., Baayen et al., 2008). Therefore, alternative ways of significance testing are recommended such as model comparison based on the log-likelihood test or, specifically for linear models, parameter estimation based on Markov chain Monte Carlo (mcmc) sampling as implemented in the pvals.fnc() function in the languageR package for R (Baayen et al., 2008). However, neither of these methods readily could serve the present purposes because the models in the present experiments included all the fixed effects of theoretical interest and because, for the linear models, the mcmc modeling method is not yet implemented for the models including random

slopes. For the present purposes, therefore, hypothesis testing was based on the p-values calculated based on the Wald-Z statistics for the logistic mixed effects models under the assumption that centering the factors will alleviate the collinearity problem (Jaeger, 2008). As for the linear models, following Baayen et al.'s (2008) suggestion that t-distribution converges to the standard normal distribution with several hundreds of observations (which is the case in the present experiments), the t-statistic exceeding 2 was considered indicative of significance at .05 level. To minimize errors in hypothesis testing, however, the log-likelihood test was also used to verify the significance of theoretically crucial interactions.

5.2. Results of WM span tests

In this section, the results from WM span tests are reported (i) to check whether the participants in different experiments were comparable in regard to WMC, (ii) to explore the relationship between the two WM span tests (i.e., reading span and subtract-2 span), and (iii) for L2 learners, to explore the relationship between WM span scores and cloze test scores. Additionally, since the L2 learners who participated in the both experimental sessions (i.e., Experiments 1/2K and Experiment 3K) yielded two instances of reading span and subtract 2-span scores with 1 to 4 months of time interval, a small-scale study of test-retest reliability of these WM span tests in L2 was possible.

Table 2 presents the summary of WM span test results for the native speakers and L2 learners who contributed data to the present experiments. The label cWM represents *composite* WM and refers to the average of the two WM span scores. This composite measure was used as the indicator of individual differences in WMC in the present dissertation. For the sake of completeness, however, the results of each span test are reported as well as the cWM scores in this section. Note that five native speakers and 18 L2 learners participated in all the three experiments, yielding two instances of each span test score. Also recall that five L2 learners in Experiments 1/2K lacked the subtract-2 span score and thus the reading span scores alone were taken as their cWM scores.

As Table 2 suggests, the two groups of native speakers who participated in Experiments 1/2E and in Experiment 3E did not differ from each other in reading span, subtract-2 span, and the cWM score as verified by a series of independent samples t-tests (all $ps > .1$), suggesting that the native speaker participants in the present study came from a fairly homogeneous population

in terms of WMC. As for the L2 learners, there was no significant difference in any of the WM measures between the two groups of L2 learners who participated in Experiments 1/2K and 3K (all $ps > .1$), which should be in large part due to the fact that the two groups shared the majority of the participants ($N=18$). The comparable level of WMC between the two experimental sessions within each language group makes it reasonable to compare any WM-related effects across experiments within each language group.

Table 2 Summary of WM measures for the participants in each experiment

		Native speakers			L2 learners		
		Reading span	Subtract-2 span	cWM ^a	Reading span	Subtract-2 span	cWM
Experiments 1E/K and 2E/K	M (SD) (N = 44)	77.8 (12.7)	84.0 (10.5)	80.9 (9.6) (N = 44)	64.4 (17.6) (N = 32)	73.1 (10.3) (N = 27)	68.0 (11.7) (N = 32)
Experiments 3E/K	M (SD) (N = 45)	77.5 (13.9)	80.5 (11.5)	79.0 (10.8) (N = 45)	62.4 (17.9) (N = 31)	72.0 (10.4) (N = 31)	67.2 (11.7) (N = 31)

^a cWM represents ‘composite working memory’ score, which is the mean of the two span test scores.

On the other hand, a series of independent samples t-tests between the native speakers and the L2 learners showed highly significant differences in all the WM measures (all $ps < .001$). The results indicate that the native speakers had reliably greater WMC than the L2 learners, which is hardly surprising. Thus we could expect that the L2 learners in the present experiments would experience greater difficulty than the native speakers in cognitive tasks that draw upon the same kind of processing resources as those captured by the WM span tests.

In order to further explore the relationship between different WM measures and, for L2 learners, the relationship between WMC and proficiency in the L2, a correlational analysis was conducted on all the available data from the native speakers ($N = 84$) and L2 learners ($N=49$), including those excluded for the analysis of the experimental results. For the participants who participated in all the three experiments and thus took the span tests twice, only the scores obtained in the first administration were included.

Table 3 presents the mean scores of the WM measures from each language group including the accuracy on the semantic judgment portion of the reading span test. The mean cloze test score for the L2 learners is also provided. Note that all the scores are in percentage.

Table 3 Mean WM span measures by language group based on the data available from all the participants in the present study (with mean close test score for L2 learners)

	Native speakers (N = 84)				L2 learners (N=49)				
	Reading span	Semantic judgment	Subtract 2 span	cWM	Reading span	Semantic judgment	Subtract 2 span	cWM	Cloze
M	77.6	98.9	82.1	79.8	61.4	85.6	71.6	66.5	67.6
(SD)	(13.5)	(1.8)	(11.2)	(10.4)	(18.5)	(9.7)	(10.4)	(11.7)	(14.9)

Table 4 presents correlations among the four different scores for the native speakers. The reading span and the subtract-2 span scores showed a moderate but highly significant correlation, indicating that the two span tests indeed share some common properties. The correlation of .41 as obtained in the present analysis is consistent with previous studies of a larger scale that showed comparable correlation coefficients typically ranging from .40 to .60 between different WM span tests such as the reading span, subtract-2 span, and operation span tests (e.g., Conway et al, 2005; Waters & Caplan, 2003). The high correlation of the cWM scores with each of the two span test scores reflects the obvious fact that the former measure was the average of the latter two measures.

Table 4 Correlations among span measures from 84 native English speaker participants

Measure	Reading span	Semantic judgments	Subtract-2 span	cWM
Reading span	1			
Semantic judgments	.15	1		
Subtract-2 span	.41***	.01	1	
cWM	.87***	.10	.81***	1

*** $p < .001$, ** $p < .01$, * $p < .05$, [†] $p < .1$

Table 5 Correlations among span measures and cloze test from 49 L2 learner participants

Measure	Reading span	Semantic judgments	Subtract-2 span	cWM	Cloze
Reading span	1				
Semantic judgments	.07	1			
Subtract 2 span	.26 [†]	.41**	1		
cWM	.90***	.24 [†]	.65***	1	
Cloze	.14	.63***	.27 [†]	.23	1

Table 5 presents the correlation coefficients from the L2 learner data. First of all, the reading span and subtract-2 span scores yielded quite a weak correlation, which was marginally significant. The weaker correlation between the two span tests as compared with native speakers could be due to the smaller number of participants and/or different processing demands imposed by the two span tests as administered in the L2. The reading span test requires the articulation and interpretation of sentences as its processing component, whereas the subtract-2 span test requires simple arithmetic and the articulation of limited number of digit names (i.e., ‘two’ to ‘nine’). Therefore, the reading span test would impose much greater ‘verbal’ processing demands than the subtract-2 span test especially for L2 learners, which would in turn reduce the common variance that is accounted for by the two span tests. The semantic judgment accuracy significantly correlated with both the subtract-2 span scores and the cloze test scores. The positive correlation between the semantic judgment accuracy and the subtract-2 span is consistent with earlier studies that showed a positive correlation between the processing and recall components of WM span tests (e.g., Conway et al., 2005). The highly significant correlation between semantic judgments and the cloze test is hardly surprising because the correct judgment of the semantic plausibility of a sentence would require a certain level of reading proficiency. The high correlation between cWM and the two span test scores reflects the composite nature of cWM. Importantly, WM measures do not seem to be closely associated with one’s general proficiency in the L2 as measured by a cloze test, as suggested by the generally weak correlations between the cloze test scores and the WM measures. Only the subtract-2 span showed a marginally significant correlation with the cloze test score. The cWM measure did not correlate with the cloze test score, however, suggesting that the cloze test and the WM span tests reflect at least partly different constructs for the present purposes.

Lastly, the test-retest reliability of the two span tests as administered in the L2 and the cWM measure derived from the two span tests was explored by calculating the correlation coefficients between the two sets of WM measures provided by the same L2 learners in the two experimental sessions. Note that the reading span data came from 24 participants and the subtract-2 span data (and the cWM measure) came from 20 participants, as summarized in Table 6. Both span tests showed quite a high reliability coefficient (for the reading span test, $r = .86$, $t(22) = 8.01$, $p < .001$; for the subtract-2 span test, $r = .71$, $t(18) = 4.26$, $p < .001$). The reliability coefficient for the cWM measure was slightly higher than that for the reading span ($r = .88$, $t(18) = 7.72$, $p < .001$). Despite a small amount of data for a correlational study, the high correlation coefficients suggest that WMC in the L2 as measured by the reading span and subtract-2 span tests may qualify as a reliable measure.

Table 6 Mean WM span measures from the L2 learners who took the span tests two times

		Reading span (N=24)	Subtract-2 span (N=20)	WM (N=20)
Session 1	M	64.8	44.4	53.6
	(SD)	(18.7)	(13.0)	(12.4)
Session 2	M	65.7	72.0	69.0
	(SD)	(17.9)	(10.0)	(11.8)

Another point that is worth noting in Table 6 is that the subtract-2 span scores increased at the second test by 27.6 % on average, which was a highly significant difference ($t(19) = -13.5$, $p < .001$), whereas the reading span scores did not show such a radical increase ($p > .1$). This may suggest that the subtract-2 span test is much more sensitive to practice effects, which would also be responsible for the lower test-retest reliability coefficient obtained from the same data. This seems to make sense because the subtract-2 span test involves simpler processing components than the reading span test, and thus it would be easier to develop some strategies to cope with the task demands for the former than for the latter.

Chapter 6

Experiments with native speakers of English

This chapter reports the three experiments with native English speakers on the processing of relative clauses (RCs) and the role of individual working memory capacity (MWC) differences thereof.

6.1. Experiment 1E

Experiment 1E tested the effect of syntactic complexity on the incremental interpretation and final comprehension of RCs with 44 native English speakers reading sentences that contained either a subject-extracted RC (SR) or an object-extracted RC (OR). Individual WMC was also included for analysis to explore whether and how individual differences in the availability of processing resources are related to sentence processing.

6.1.1. Materials and design

Twenty-four sets of four different versions of sentences such as (52) – (55) were taken from Experiment 1 of the Traxler et al. (2005) study, with some vocabulary items replaced with those that were judged by the experimenter to be more familiar to L2 learners (e.g., *despise* replaced with *hate*), and with most of the main clause predicates being rendered identical in the same set of sentences when resulting in plausible meaning (see Appendix I for the entire lists of test sentences). Two verbs were repeated two times as the relative clause verb and one verb was repeated two times as the main clause verb.

(52) The banker that irritated the lawyer played tennis every Saturday.

(53) The banker that the lawyer irritated played tennis every Saturday.

(54) The lawyer that irritated the banker played tennis every Saturday.

(55) The lawyer that the banker irritated played tennis every Saturday.

All the RC head and RC-internal nouns were animate, and plausible subjects/objects of the RC verbs. The test sentences like (52) contained an SR, whereas those like (53) contained an OR. The only difference on the surface between (52) and (53) is the order of the RC-internal noun and the RC verb. The other two sentences in this set, (54) and (55), controlled for any

differences in plausibility and word length/frequency across conditions by reversing the order of the RC head nouns and RC-internal nouns in (52) and (53).

Each test sentence was paired with a simple comprehension verification statement about the content of either the RC or the main clause (e.g., King & Just, 1991). Half of the statements were about the RC and the other half about the main clause. Half of the statements were true to the content of the target sentence and the other half were not. For example, a true statement about the SR *the banker that irritated the lawyer* in (52) and the OR *the lawyer that the banker irritated* in (55) would be *the banker irritated the lawyer*, while the statement *the lawyer irritated the banker* would not be true (See Appendix I for the entire lists of verification statements following the test sentences). The verification statement for each test sentence was identical across all the participants.

The test sentences were distributed across four different lists so that exactly one version of each set appeared in a list and an equal number of the sentences of the same version appeared across the lists. Each list thus contained six sentences from each version. Since the two different versions of SRs and ORs were combined for analysis, each participant read 12 SR and 12 OR sentences. The test sentences were presented together with 28 sentences from Experiment 2E and 58 filler sentences of various structures, in random order for each participant. Approximately half of the filler sentences were also followed by a comprehension verification statement.

6.1.2. Predictions

It was predicted that ORs would be harder to comprehend than SRs, following the findings of much previous work (e.g., Wanner & Maratsos, 1978). If the OR penalty can be characterized by the limitations of working memory (WM) resources, and if WMC as measured by WM span tests reflects the availability of those WM resources, we would expect the OR penalty to be modulated by WMC (e.g., King & Just, 1991). More specifically, individuals with higher WMC would be faster and/or more accurate in comprehending sentences than those with lower WMC, especially when the target structure is complex as is the case with ORs, while such individual differences would not be as pronounced for simpler structures such as SRs (cf. Just & Carpenter, 1992). If the OR penalty is instead due to the difficulty arising from the linguistic processes involved in assigning grammatical and thematic relations to non-canonically ordered constituents, and if (at least some of) these processes draw upon some kind of dedicated

processing resources that is not measured by WM span tests, OR penalty would not be modulated by WMC (e.g., Traxler et al., 2005; Waters & Caplan, 2004).

6.1.3. Results and discussion

In the results and discussion sections for this and the following experiments, comprehension and reading time data are reported in separate sub-sections in that order.

6.1.3.1. Comprehension accuracy

The mean comprehension accuracy for the SR condition was 86.0% (SD = 11.6), which was higher than the 77.7% accuracy (SD = 14.7) in the OR condition. Table 7 presents the logistic mixed effects model for the accuracy data. The fixed effects in the model are listed in the first column of the table. The selected random structure of the model is described at the bottom. The estimates given in the second column of the table represent the effects of the corresponding fixed effects in log odds. As the table shows, the main effects of RC type and cWM were reliable, but there was no interaction between the two factors. The negative sign of the estimate for RC type suggests that the odds of correct responses were lower for ORs than for SRs. The positive sign of the estimate for cWM indicates that individuals with higher WMC were more accurate than those with lower WMC in comprehension. The results suggest that SRs are easier to comprehend than ORs, and that greater WMC is associated with more accurate comprehension. The lack of interaction between the two factors suggests that the syntactic complexity effect on comprehension accuracy did not vary across individuals with different WM spans.

Table 7 Logistic mixed effects model for the comprehension accuracy in Experiment 1E

	Estimate	Standard Error (SE)	z
(Intercept)	1.85	0.21	8.77*** ^b
RC type	-0.67	0.18	-3.76***
cWM	0.04	0.01	3.30***
RC type:cWM ^a	-0.01	0.02	-0.44

The model included random intercepts for participants and items.

a. “:” indicates interaction.

b. *** $p < .001$, ** $p < .01$, * $p < .05$, [†] $p < .1$

6.1.3.2. Reading time results

Figure 1 presents mean reading times in milliseconds (ms) at each region across the conditions. The reading times at the RC region were averaged across the words within the RC excluding the complementizer *that*. The two words following the RC region constituted the MV and MV+1 region, respectively. Figure 1 suggests that ORs were read slower than SRs especially at the MV region.

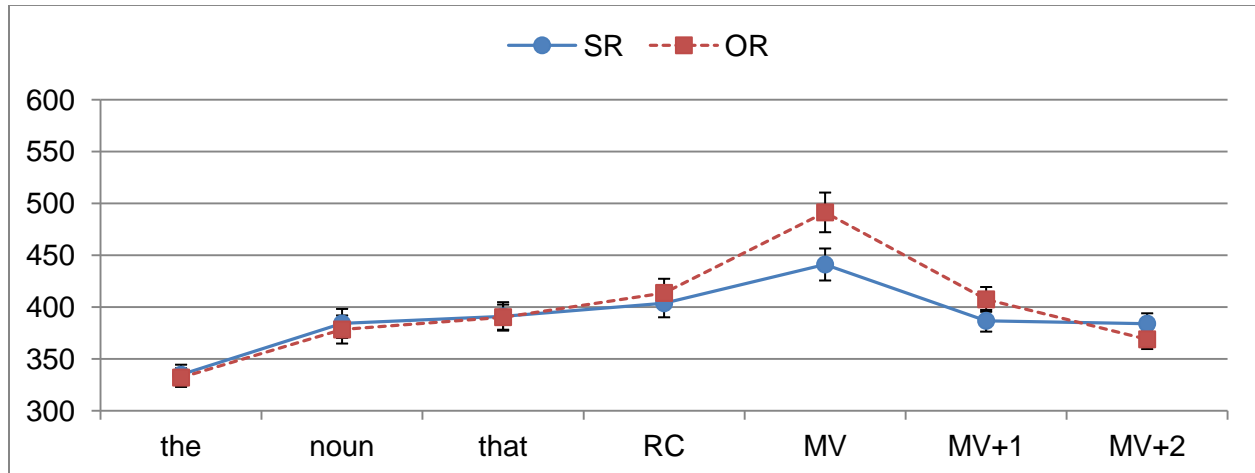


Figure 1 Mean reading times by region and condition in Experiment 1E (*Error bars represent two standard errors.*)

The effects of RC type and WMC were evaluated by a mixed effects model fitted to the data from each of the 3 critical regions. Table 8 presents the fitted models by region. At the RC region, there were no reliable main effects or interactions. At the following two regions, the main effect of RC type was reliable, indicating that ORs took longer to read than SRs by the estimated differences of 50ms and 19ms at each region. There was no reliable interaction of RC type and cWM at either of the two regions, suggesting that the syntactic complexity effect reflected as OR penalty in reading times was not modulated by individual differences in WMC.

Table 8 Linear mixed effects model for the reading times at each critical region in Experiment 1E

Region	RC		MV		MV+1	
Main effects	Estimate (SE)	t	Estimate (SE)	t	Estimate (SE)	t
(Intercept)	408.60 (13.71)	29.80***	465.88 (16.97)	27.46***	396.70 (10.99)	36.11***
RC type	9.32 (6.23)	1.50	49.52 (12.55)	3.95***	19.42 (8.00)	2.43*
cWM	-0.09 (1.42)	-0.06	-0.24 (1.73)	-0.14	-0.22 (1.11)	-0.20
RC type:cWM	0.44 (0.65)	0.67	-0.15 (1.32)	-0.13	0.98 (0.85)	1.16

Each model included random intercepts for participants and items.

Both comprehension and reading time results in Experiment 1E replicated the well-established finding that ORs are harder to comprehend than SRs. Individual differences in WMC were associated with overall comprehension accuracy, but not with syntactic complexity effects either in final comprehension or in online reading. The results are consistent with the dedicated-resource view of WMC and sentence processing (Traxler et al., 2005; Waters & Caplan, 2004), which does not predict an interaction of WMC as measured by WM span tests and interpretive processes involved in online sentence processing, but does allow post-interpretive processes such as the explicit recall of the content of a sentence to draw upon the WM resources that are tapped by WM span tests.

6.2. Experiment 2E

In this experiment, the interaction of syntactic structure and lexical-semantic information during the course of sentence interpretation was tested by varying the animacy of nouns relevant to the interpretation of SRs and ORs. The effect of WMC was also tested. Participants were the same as Experiment 1E.

6.2.1. Materials and design

Twenty-eight sets of four different versions of sentences such as (56)–(59) were taken from Experiment 2 of Traxler et al. (2005), with some vocabulary items replaced with those that

were judged by the experimenter to be more familiar to L2 learners (for Experiment 2K) and with all the main clause predicates rendered identical across the different versions (see Appendix I for the list of entire test sentences). Consequently, seven verbs were used two times and one verb was used three times as the RC verb in the entire set of test sentences. An additional element such as an adverb (e.g., *recently*) or a contracted form of negation (i.e., *didn't*) was placed in front of the RC verb as a spill-over region. This manipulation was critical for testing whether any effects due to the animacy of the subjects of ORs emerge before the processor encounters the RC verb in the present experiment. Without the spill-over region, any effects due to the characteristics of the subjects of ORs could spill over to the following verbs and become very hard to isolate from the effects due to the verbs.

(56) The director that didn't praise the movie received a prize at the film festival. *SR-A(I)*

(57) The director that the movie didn't please received a prize at the film festival. *OR-A(I)*

(58) The movie that didn't please the director received a prize at the film festival. *SR-I(A)*

(59) The movie that the director didn't praise received a prize at the film festival. *OR-I(A)*

Across the four sentences in (56) – (59), types of RCs (i.e., SR vs. OR) and animacy configuration of two nouns relevant to the interpretation of the RCs (i.e., animate RC head noun – inanimate RC-internal noun vs. inanimate RC head noun – animate RC-internal noun) were crossed. The sentence (56) contained an SR with an animate head and an inanimate object (i.e., *the director that didn't praise the movie*), while the sentence (57) contained an OR with an animate head and an inanimate subject (i.e., *the director that the movie didn't please*). In other words, the RC verb in (56) is associated with an animate subject and an inanimate object, whereas the RC verb in (57) is associated with an inanimate subject and an animate object. On the surface, however, the first noun is animate and the second noun is inanimate both in (56) and (57), hence the same animacy configuration. This animacy configuration is labeled as A(I) hereafter, representing the animate RC head noun and the inanimate RC-internal noun. By contrast, the sentences in (58) and (59) contained an SR and an OR with the reversed animacy configuration in which the first noun is inanimate and the second noun is animate. This animacy configuration is labeled as I(A) hereafter. Thus the sentences in (56) – (59) each represent SR-A(I), OR-A(I), SR-I(A), and OR-I(A) conditions, respectively.

Due to the manipulations of noun animacy and RC type, two different classes of nouns (e.g., *director* vs. *movie*) and verbs (e.g., *praise* vs. *please*) were used across conditions. So the lengths and frequencies of the nouns and those of the verbs were matched between the different classes, respectively. The length of a word was operationalized as the number of letters, and the mean printed frequency of a word was obtained from Francis & Kučera's (1982) counts of all the inflected forms of the word in a million-word corpus. There was no reliable difference in length between the animate nouns ($M = 6.1$, $SD = 1.8$) and the inanimate nouns ($M = 6$, $SD = 1.9$) ($t(27) = .35$, $p > .1$), nor in frequency (for the animate nouns, $M = 153.9$, $SD = 403.6$; for the inanimate nouns, $M = 111.5$, $SD = 150.9$) ($t(27) = .54$, $p > .1$)¹³. There were no reliable differences in length and frequency between the verbs associated with an animate subject (e.g., *praised*) (length: $M = 7.7$, $SD = 1.7$; frequency: $M = 51.0$, $SD = 42.7$) and those associated with an inanimate subject (e.g., *pleased*) (length, $M = 7.7$, $SD = 1.3$; frequency: $M = 42.3$, $SD = 28.4$) (both $ts < 1.1$, $ps > .1$).

Each set of test sentences contained two different propositions expressed by the RCs and two different propositions expressed by the main clauses as shown in (60)–(63), which are derived from the sentences (56)–(59).

(60) The director didn't praise the movie.

(61) The movie didn't please the director.

(62) The director received a prize.

(63) The movie received a prize.

The propositions (60) and (61) represent the RCs with different animacy configurations, and the propositions (62) and (63) represent the main clauses with different subjects. In order to check for potential differences in plausibility between the two different RC propositions and between the two different main clause propositions in each set of test sentences, 41 of the 44 native speaker participants analyzed in Experiment 2E provided plausibility ratings on the propositions after they had completed the self-paced reading task. The propositions were distributed through four lists in a Latin-square design so that a participant saw only one of the

¹³ The numerically higher mean frequency of the animate nouns as well as the large standard deviation was mainly due to an extremely frequent word *man* (2110/million). Without the word, the mean frequency and standard deviation for the animate nouns were 81.4 and 128, respectively.

propositions from a set. The participants marked the plausibility of each proposition on a seven-point scale with 1 representing “Least plausible” and 7 “Most plausible”. Table 9 presents the mean plausibility rating on each type of different propositions. The two different types of RC propositions did not differ from each other in terms of plausibility as indicated by a paired-samples t-tests ($t(27) = .074, p > .1$), as was the case with the two types of main clause propositions ($t(27) = .382, p > .1$).

Table 9 Mean plausibility ratings on different propositions included in the test sentences in experiment 2E

	Mean plausibility rating (SD)
RC propositions with an animate subject and an animate object (e.g., The director didn’t praise the movie.)	5.30 (1.11)
RC propositions with an inanimate subject and an inanimate object (e.g., The movie didn’t please the director.)	5.28 (1.08)
Main clause propositions with an animate subject (e.g., The director received a prize.)	5.73 (0.82)
Main clause propositions with an inanimate subject (e.g., The movie received a prize.)	5.64 (0.94)

The result of the post-hoc plausibility ratings does not guarantee that the test sentences were equally plausible across conditions because the test sentences in a set contained different combinations of different RC and main clause propositions, which could be different in plausibility. However, this problem may not be directly relevant to the present purposes for at least two reasons. First, the interpretation of any effects observable within the RCs (i.e., before the comprehender encounters the main clause verb) will be free from this problem because those effects are obtained before any plausibility effect due to different combinations of propositions can appear. Second, although the interpretation of reading times at and after the main clause verb, and the interpretation of comprehension accuracy might be confounded with the potential differences in plausibility across conditions, these effects will still reflect participants’ sensitivity to plausibility information which is a type of lexical-semantic information after all. Thus any modulation of the traditional OR penalty could be ultimately interpreted as the effect of lexical-semantic information, the sensitivity to which is one of the main question of interest for the present dissertation.

Each test sentence was paired with a comprehension verification statement. Due to the manipulation of animacy configuration, the propositions expressed by the RCs and the main clauses were not semantically reversible¹⁴. So it was impossible to perfectly balance the number of true and not-true statements about the RCs and the main clauses as in Experiment 1E. The result was that 23 statements were about the main cause and five were about the RC. Half of the statements were true and the other half were not true. For this reason, the comprehension accuracy results should be interpreted as somewhat indirect evidence for the relative comprehension difficulty associated with different types of RCs.

The test sentences were presented together with the 24 test sentences from Experiment 1E and 58 filler sentences of various structures, in random order for each participant.

6.2.2. Predictions

Following Traxler et al. (2002, 2005; also Mak et al, 2002, 2006), an interaction of RC type and noun animacy configuration is expected in such a way that ORs with animate head nouns and inanimate RC-internal nouns (e.g., *the director that the movie didn't please*) will be more difficult to comprehend than SRs with the same animacy configuration (e.g., *the director that didn't praise the movie*), whereas ORs with inanimate head nouns and animate RC-internal nouns (e.g., *the movie that the director didn't praise*) will be no more difficult than SRs with the same animacy configuration (e.g., *the movie that didn't please the director*). Furthermore, if the asymmetrically large difficulty with the OR-A(I) condition is due to the *structural* processes involved in reanalyzing the grammatical roles of the noun phrases before the processor encounters the RC verb as the reanalysis-based accounts would predict, the effect will arise before the RC verb. If the difficulty arises during the *lexical* integration between the RC verb and its arguments as predicted by the integration- and interference-based accounts of sentence processing, the effect should appear at or after the RC verb. In this case, however, it will be practically impossible to distinguish between the two potential sources of difficulty because the behavioral consequence due to the difficulty with structural computation can be temporally delayed and appear later in the sentence at or after the RC verb. On the other hand, any effect

¹⁴ For example, the test sentence *The movie that didn't please the director received a prize* could be not paired with a not-true statement such as *The director didn't please the movie* because it is simply impossible for a director to (not) please a movie.

that appears earlier than the RC verb could not be attributed to the purely lexical integration between the RC verb and its arguments.

6.2.3. Results and discussion

6.2.3.1. Comprehension accuracy results

It was found that the comprehension statements associated with one set of test sentences had an error after running the experiment and they were excluded from the analysis of comprehension accuracy. Thus, 27 items were analyzed for each participant. Table 10 summarizes the comprehension accuracy for each condition. The results show that the accuracy on the comprehension verification statements is particularly low in the OR-A(I) condition as compared with the other three conditions.

Table 10 Mean comprehension accuracy by condition in Experiment 2E (in percent)

	A(I)		I(A)	
	SR	OR	SR	OR
M	85.1	68.2	89.9	87.0
(SD)	(16.0)	(20.9)	(13.3)	(15.0)

The logistic mixed effects model is presented in Table 11 and shows that the interaction of RC type and animacy configuration was reliable as were the respective main effects of the two factors, indicating that the OR penalty was reliably larger in the A(I) condition than in the I(A) condition. The result suggests that the typicality of the semantic properties associated with certain grammatical roles (e.g., animate nouns in subject position) indeed modulates syntactic complexity effects, replicating Traxler et al.'s (2002, 2005) results. The main effect of cWM was also reliable, with the positive coefficient indicating that higher WMC was associated with generally higher accuracy in comprehension as was the case in Experiment 1E. Any interactions involving cWM were not reliable.

Table 11 Logistic mixed effects model for the comprehension accuracy in Experiment 2E

	Estimate	SE	z
(Intercept)	2.02	0.18	11.35***
RC type	-0.59	0.19	-3.19**
Animacy	1.00	0.18	5.40***
cWM	0.03	0.01	2.26*
RC type:Animacy	1.06	0.39	2.88**
RC type:cWM	-0.02	-1.06	-1.06
Animacy:cWM	-0.02	-1.01	-1.01
RC:Ani.:cWM ^a	-0.02	-0.47	-0.47

The model included random intercepts for participants and items.

a. Predictor labels are abbreviated.

6.2.3.2. Reading time results

Figure 2 presents the mean reading times at each region by condition. The reading times at the RC region were averaged across the four words in that region. The figure suggests that the OR-A(I) condition was read more slowly than the other three conditions both at the MV and MV+1 region. To assess the observed pattern, a mixed effects model was fitted to the reading time data at each of the three critical regions, and the estimates for the fixed effects at each region are presented in Table 12.

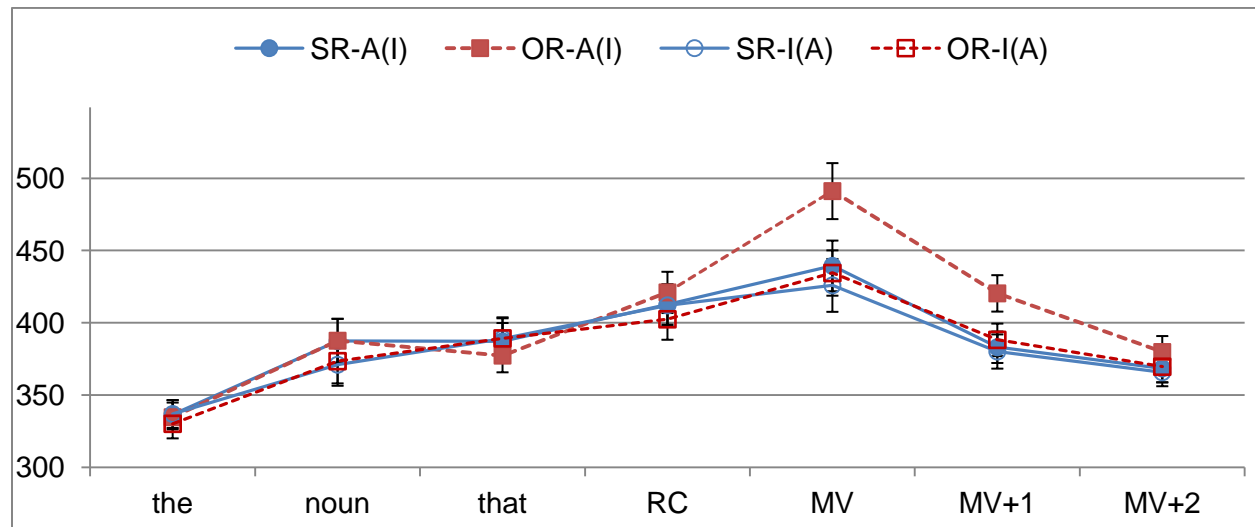
**Figure 2 Mean reading times by region and condition in Experiment 2E**

Table 12 Linear mixed effects model for the reading times at each critical region in Experiment 2E

Region	RC		MV		MV+1	
Fixed effects	Estimate (SE)	t	Estimate (SE)	t	Estimate (SE)	t
(Intercept)	411.79 (13.64)	30.18***	447.84 (16.04)	27.93***	392.74 (10.48)	31.47***
RC type	-0.73 (5.63)	-0.13	31.07 (9.46)	3.29**	22.44 (6.51)	3.44**
Animacy	-9.19 (5.61)	-1.64	-35.69 (9.43)	-3.78***	-16.92 (6.50)	-2.60*
cWM	-1.20 (1.42)	-0.85	-1.82 (1.66)	-1.10	-0.01 (1.09)	-0.01
RC:Ani.	-17.09 (11.22)	-1.52	-44.58 (18.87)	-2.36*	-28.04 (13.00)	-2.16*
RC:cWM	1.03 (0.59)	1.75 [†]	-0.51 (0.99)	-0.51	0.16 (0.69)	0.24
Ani.:cWM	-0.49 (0.59)	-0.83	-0.71 (0.99)	-0.72	-0.70 (0.69)	-1.02
RC:Ani.:cWM	3.69 (1.18)	3.123**	3.06 (1.98)	1.54	2.31 (1.38)	1.68 [†]

Each model included random intercepts for participants and items.

At the RC region, there was a marginally significant interaction of RC type and WMC. But it was further moderated by a significant 3-way interaction of RC type, animacy configuration and cWM, indicating that the effects of RC type and animacy configuration differed across participants with different WMC. The main effects and the other interactions were not reliable. In order to explore the effect of WMC on the interaction of RC type and animacy configuration, the participants were divided into the higher and lower WMC groups at the median cWM score (those with the median cWM score was categorized as higher WMC participants). Figure 3 presents the mean reading times in each condition by WMC group. The pattern suggests that the 3-way interaction is mainly due to the fact that the lower WMC participants read the SR-I(A) condition more slowly than the OR-I(A) condition, which was not the case with the higher WMC participants. The result might suggest that the lower WMC participants experience substantial processing difficulty even with simple SR constructions when noun animacy conflicts with grammatical roles, unlike the higher WMC participants.

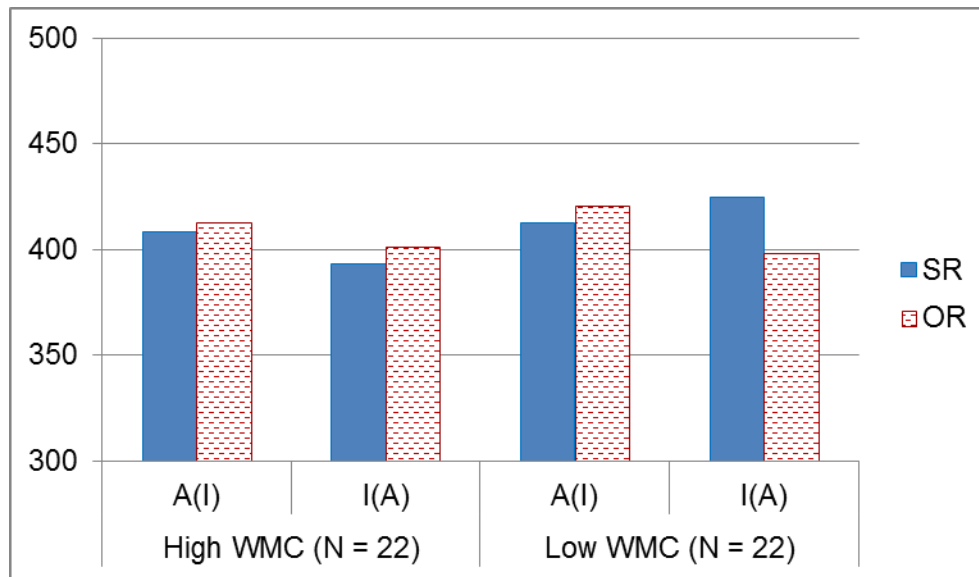


Figure 3 Mean reading times at the RC region by WMC group Experiment 2E

At the MV region, main effects of RC type and animacy configuration were both reliable, which were further moderated by their reliable interaction. This interaction was due to the fact that, as shown in Figure 2, the OR-A(I) condition took longer to read than the other three conditions, which were very similar to one another in mean reading time. The result suggests that the native speakers of English, regardless of their WMC, benefited from helpful semantic information available in the OR-I(A) condition, as was shown by Traxler et al. (2002, 2005). Thus the result is consistent with the reanalysis-based account of Traxler et al. that proposes that syntactic complexity and noun animacy jointly determine the relative ease of processing RCs. More specifically, in the OR-I(A) condition, the initial misanalysis of the head noun as the subject of the RC seems to be easily revised because the head noun is not a good subject in the first place and the later coming RC subject is a good subject by being animate.

The same statistical pattern was maintained in the MV+1 region, except that the 3-way interaction of RC type, animacy configuration, and cWM became marginal. To explore the pattern of this trend of interaction, the mean reading time for each condition are presented by WMC group in Figure 4. The figure suggests that the marginal 3-way interaction was mainly due to the lower WMC participants still experiencing greater difficulty with the OR-A(I) condition relative to the SR-A(I) condition as was the case in the MV region, whereas the higher WMC participants no longer showed an asymmetry in the magnitude of the OR penalty across the two

animacy configuration conditions. The result seems to suggest that individuals with lower WMC experience particular difficulty in processing ORs when semantic information is not helpful to carrying out a reanalysis process.

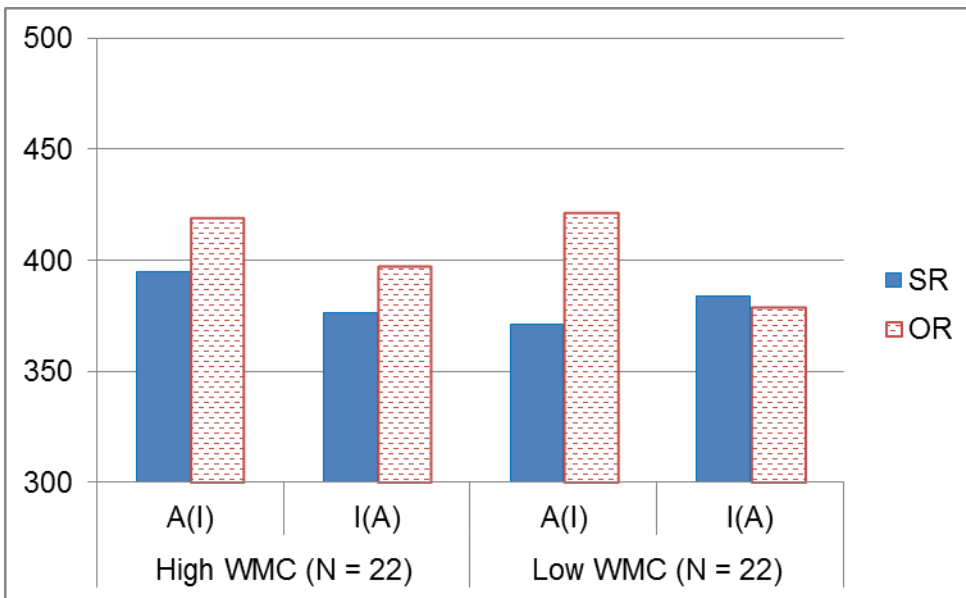


Figure 4 Mean reading times at the MV+1 region by WM group in Experiment 2E

As a further analysis, the mean reading times at each word within the RC region was compared between the two OR conditions to explore the time course of noun animacy effects on reanalysis. Only ORs were analyzed because the main interest of the present dissertation is how animacy configuration modulates syntactic complexity effects as expected in the ORs. Figure 5 presents the mean reading times at each word by condition. A mixed effects model with animacy configuration as its only fixed effect was fitted to the reading times at each word. The animacy effect was not reliable at the first two words (all $|t|s < 1$). However, the ORs with animate heads and inanimate subjects (e.g., *the director that the movie didn't please*) were read reliably more slowly than those with inanimate heads and animate subjects (e.g., *the movie that the director didn't praise*) both at the adverb and at the RC verb (at the adverb, Beta = -23.24, SE = 10.26, $t = -2.266$, $p < .05$; at the RC verb, Beta = -32.41, SE = 16.41, $t = -1.975$, $p < .05$)¹⁵. The result indicates that the effect of helpful animacy information emerged *before* the processor encountered the RC verb.

¹⁵ Both models included random intercepts for participants and items as the random effects.

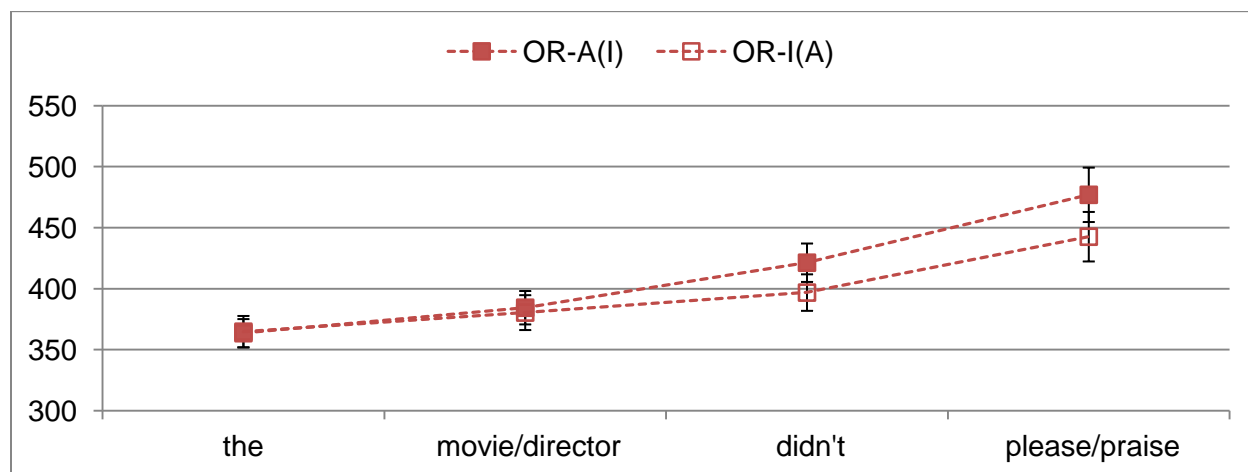


Figure 5 Mean reading times at each word within the ORs in Experiment 2E

To summarize, Experiment 2E tested the effects of syntactic complexity, noun animacy configuration, and WMC on the processing of RCs with native speakers of English. The general pattern of results across all the participants will be first discussed and then the effect of individual WMC differences. The results show that the participants experienced greater difficulty in processing ORs than SRs when the head nouns were animate and the RC-internal nouns were inanimate, but this traditional OR penalty disappeared when the head nouns were inanimate and the RC-internal nouns were animate, as reflected in final comprehension accuracy and in the reading times at the two words following the RC region. All of these findings are consistent with the previous studies that showed the same pattern of results (e.g., Gennari & MacDonald, 2008; Mak et al, 2002, 2006; Traxler et al., 2002, 2005).

The word-by-word comparisons within the RC region between the OR conditions with the opposite animacy configurations revealed that the relative difficulty due to different animacy configurations began before the RC verb was encountered, by showing a reliable difference in reading time between the two conditions at the region preceding the RC verb. The result suggests that native English speakers incrementally build sentential representations based on the syntactic structure and semantic information available at a given moment without resorting to lexical subcategorizers such as verbs. This finding can easily be explained by the accounts highlighting incremental structure-building processes such as the reanalysis-based and frequency-based accounts (e.g., Gennari & MacDonald, 2008; Staub, 2010; Traxler et al., 2002, 2005).

Conversely, the result points to an empirical inadequacy of the accounts of OR penalty that exclusively consider the processing load incurred by lexical integration between syntactic dependents or memory interference appearing at the point of retrieval (e.g., Gibson, 2000; Gordon et al., 2001, 2004; Van Dyke, 2007).

The effect of WMC appeared not only in final comprehension but also in reading speed in Experiment 2E. Consistent with Experiment 1E, higher WMC was associated with generally higher accuracy in final comprehension regardless of syntactic complexity and animacy configurations. Unlike Experiment 1E, in which there were no reliable WMC-related effects on syntactic complexity effects in online reading, the results of Experiment 2E showed that the ability to incorporate noun animacy information to process RC constructions might be modulated by individual differences in the availability of WM resources, as reflected in the reliable 3-way interaction of RC type, noun animacy configuration, and WMC at the RC region, as well as in the same trend of interaction two words after the RC region.

The pattern of WMC effects on reading times differed across the regions. At the RC region, the lower WMC participants seemed to experience particular difficulty in processing the SRs with inanimate heads and animate RC-internal nouns, which were read more slowly than the ORs with the same animacy configuration, whereas the higher WMC participants did not distinguish the SRs and ORs in processing difficulty in either animacy configuration. At the second word after the RC region, the lower WMC participants showed a tendency to experience particular difficulty when they read the ORs with animate heads and inanimate RC-internal nouns, which was not the case with the higher WMC participants. In short, the lower WMC participants seemed to experience more steeply increased difficulty as compared to the higher WMC participants with the RCs that contained non-canonical noun animacy – grammatical role associations (i.e., an inanimate subject and an animate object), though the corresponding behavioral signs appeared at the RC region for SRs and at a later region for ORs. This pattern of results might suggest that WMC is associated with the ability to resolve the conflicting information to reach an ultimately correct analysis (cf. Pearlmutter & MacDonald, 1995).

There seem to be at least two possibilities for why WMC effects occurred earlier for SRs than for ORs. One possibility is that the different locations of RC verbs in SRs and ORs may have determined at which point in a sentence the processing complexity reached a threshold level to be modulated by WMC and to be detected by reading time measures. Assuming that

encountering verbs triggers multiple processes related to syntactic/thematic role assignments and thus causes a generally higher level of processing cost than nouns, processing difficulty involved with SRs may reach the threshold earlier in the sentence than processing difficulty involved with ORs because RC verbs appear earlier in the former. Processing difficulty arising from verbs could have spilled over to the following word(s), making the individual differences in reading time for SRs detectable at the RC region and those for ORs detectable one word after the RC region.

An alternative possibility is more concerned with the difference in the nature of difficulty between with SRs with inanimate heads and ORs with animate heads, as is hypothesized by the reanalysis-based account. According to the account, SRs do not induce reanalysis because the structure is initially preferred. Then any difficulty with SRs with inanimate heads should be due to the non-canonical animacy configuration in the arguments of the RC verbs. It is possible that the non-canonical animacy configuration alone induced substantial difficulty for the lower WMC participants, but not for the higher WMC participants. On the other hand, the processing difficulty involved with ORs with animate heads is due to a required reanalysis process plus non-canonical animacy configuration, which explains the increased difficulty that the participants of both higher and lower WMC experienced in the OR-A(I) condition at the MV region. With the severe processing difficulty in that condition, it is possible that the lower WMC participants took longer to initiate/complete the reanalysis process, hence the WMC effect shown in the MV+1 region.

The two possibilities discussed above are not mutually exclusive. If we assume that the processing difficulty for RCs comes from multiple sources of different kinds, as suggested by Staub (2010), both the location of the RC verb and the unusual difficulty of reanalysis would be responsible for the time course of WMC effects as shown in the present experiment. What is important for the present purposes is that WMC seems not only to predict the overall comprehension accuracy but also to modulate readers' ability to incorporate different types of information during the course of online sentence processing, especially when the different types of information conflict with each other.

Recall that in Experiment 1E, WMC was associated with overall comprehension accuracy but not with syntactic complexity effects either in final comprehension or in online reading. With Experiment 1E and 2E combined, the results seem to be consistent with Traxler et al.'s (2005)

view on the relationship between WM resources and sentence processing, which is an elaborated version of the separate-resource view. The general implications of the present findings about the debate on the nature of WM resources underlying sentence processing will be further discussed in Section 6.4 after presenting the results of Experiment 3E in the following section.

6.3. Experiment 3E

Experiment 3E tested different animacy configurations from Experiment 1E, as an attempt to further investigate the processing of RCs and the effect of individual differences in WMC on sentence processing. Forty-five native English speakers contributed data.

6.3.1. Materials and design

For Experiment 3E (and 3K), 28 sets of four sentences were constructed as exemplified in (64) – (67). The crucial manipulation was that the head nouns of the RCs were kept animate and identical across conditions. The conditions represented by the four sentences in (64) –(67) are labeled as SR-A(A), OR-A(A), SR-A(I), and OR-A(I), in that order.

- | | |
|---|----------------|
| (64) The teacher that applauded the actor was invited to the film festival. | <i>SR-A(A)</i> |
| (65) The teacher that the actor surprised was invited to the film festival. | <i>OR-A(A)</i> |
| (66) The teacher that applauded the movie was invited to the film festival. | <i>SR-A(I)</i> |
| (67) The teacher that the movie surprised was invited to the film festival. | <i>OR-A(I)</i> |

An effort was made to restrict the vocabulary items to those that are judged by the experimenter to be relatively familiar to L2 learners. Consequently, 18 verbs were used twice as the RC verb across the entire set of test sentences. As in Experiment 2E, different classes of nouns (e.g., *actor* vs. *movie*) and verbs (*applaud* vs. *surprise*) across conditions were matched in mean length and printed frequency. The lengths of nouns in the number of letters were not significantly different between the animate ($M = 6.8$, $SD = 1.6$) and the inanimate nouns ($M = 6.6$, $SD = 1.5$), as also were the case with the printed frequencies (animate nouns, $M = 84.4$, $SD = 125.4$; inanimate nouns, $M = 87.5$, $SD = 69.2$). The same was true between the verbs that allow only animate subjects such as *applauded* (length, $M = 7.9$, $SD = 2.0$; frequency, $M = 42.9$, $SD = 33.3$) and the verbs that allow both animate and inanimate subjects such as *surprised* (length, $M = 7.9$, $SD = 1.5$; frequency, $M = 36.9$, $SD = 29.7$) (all $ps > .1$).

The propositions expressed by the RCs differed across conditions as exemplified in the first two columns of Table 13. In order to check for potential differences in plausibility of the propositions across conditions, 30 of the 45 participants in Experiment 3E rated the plausibility of the propositions derived from the RCs of the test sentences after they had finished the self-paced reading task, in the same way as described in the Materials and design section for Experiment 2E. The third column of Table 13 presents the result of the plausibility ratings. A 2 (SR vs. OR) \times 2 (A(A) vs. A(I)) repeated measures ANOVA on the item means aggregated across participants showed that there was no significant difference in plausibility between the conditions (all F s < 1).

The test sentences were distributed across four lists so that each participant read only one version of a set of test sentences and equal number of test sentences of the same version. Each list of test sentences was presented together with 32 sentences from an unrelated experiment and 60 filler sentences. Presentation order was randomized for each participant. Each test sentence was followed by a comprehension verification statement. Half of them were about the RCs and the other half were about the main clauses, with half of them being true and the other being not-true. Half of the unrelated experimental sentences and half of the filler sentences were also followed by a comprehension statement.

Table 13 Mean plausibility ratings on the propositions derived from the RCs in Experiment 3E

Condition	Example of RC proposition	Mean plausibility rating (SD)
SR-A(A)	The teacher applauded the actor.	6.01 (0.74)
OR-A(A)	The actor surprised the teacher.	6.00 (0.70)
SR-A(I)	The teacher applauded the movie.	5.98 (0.84)
OR-A(I)	The movie surprised the teacher.	5.94 (1.08)

6.3.2. Predictions

Because the head nouns of the RCs were the same across conditions and were all animate, they were equally good subjects of the RC verbs. Thus the magnitude of reanalysis difficulty due to animate RC heads will also be equal between the two OR conditions with different animacy configurations. By comparing the two OR conditions, therefore, we can isolate the effect of the animacy of RC subjects on the relative ease of reanalysis hypothesized to occur in ORs.

The reanalysis-based account predicts that ORs will be easier to process when the RC subjects are animate (e.g., *the teacher that the actor surprised*) than when they are inanimate (e.g., *the teacher that the movie surprised*), because an animate noun is a good subject so it should help to discard the initial misanalysis of the RC head noun as the RC subject (Traxler, et al., 2005: p.218). Other accounts highlighting incremental structure-building such as the expectation- and frequency-based accounts would also accommodate the same prediction as long as the OR-A(A) construction can be shown to be more expectable or frequent than the OR-A(I) construction.

The results of Experiment 2E suggest that the difficulty involved with ORs may not be exclusively attributed to processing loads or memory interference arising at the point of integrating the RC verb with the distant RC head, because the modulation of syntactic complexity effects by noun animacy had begun before the processor encountered the RC verb. However, the results do not speak directly to the role of the processes triggered by the RC verbs because any relevant effects could have been masked by the difficulty that had emerged earlier. By manipulating animacy configurations in a different way, Experiment 3E was also able to test at least one type of primarily lexically-based account. According to the interference-based account, animate RC subjects should induce greater processing difficulty than inanimate RC subjects in ORs, because the former are fully confusable with RC head nouns in terms of animacy while the latter are not, which is exactly the opposite to what is predicted by the reanalysis-based account.

6.3.3. Results and discussion

6.3.3.1. Comprehension accuracy

Table 14 presents the mean accuracy in percentage by condition, which suggests that the ORs were overall more difficult to comprehend than the SRs and that the RCs with animate RC-internal nouns (i.e., A(A) condition) were overall more difficult to comprehend than the RCs with inanimate RC-internal nouns (i.e., A(I) condition). To assess this pattern of results and the effect of WMC, a logistic mixed effects model was fitted to the comprehension accuracy data, which is presented in Table 15.

Table 14 Mean comprehension accuracy by condition in Experiment 3E (in percent)

	A(A)		A(I)	
	SR	OR	SR	OR
M	85.7	72.1	90.5	80.0
(SD)	(14.6)	(20.6)	(9.1)	(16.2)

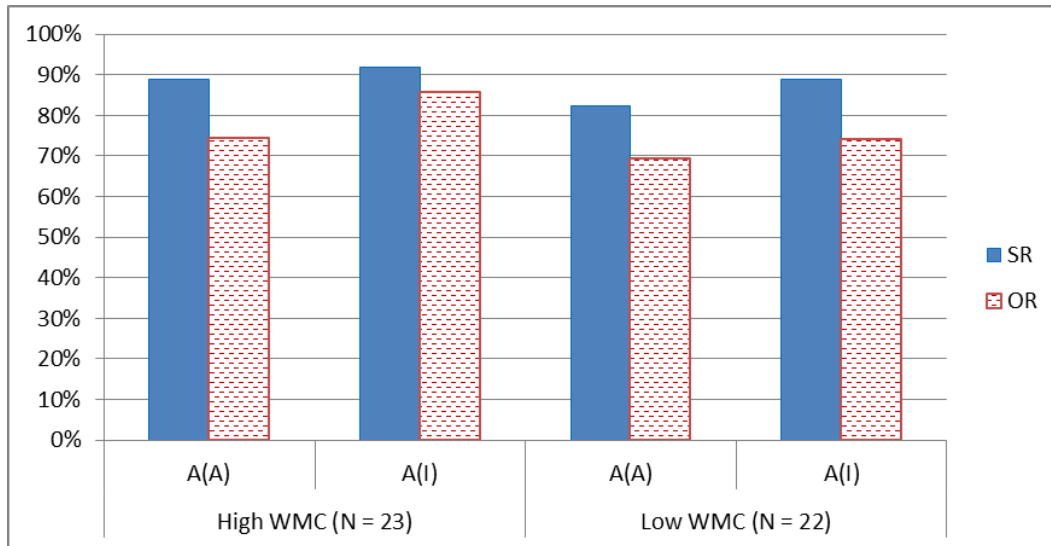
The main effects of RC type and animacy configuration were reliable, indicating that the OR condition was more difficult to comprehend than the SR condition, and that the A(A) condition was more difficult than the A(I) condition. The main effect of cWM was also reliable, indicating that the participants with higher WMC were overall more accurate in final comprehension, as was also the case in Experiment 1E and 2E. Unlike Experiment 2E, however, the interaction of RC type and noun animacy configuration was not reliable. There was a trend toward a 3-way interaction of RC type, animacy, and cWM, suggesting modulation by WMC of the effects of syntactic complexity and semantic information on comprehension accuracy. To explore this trend, the participants were divided at the median cWM score, and the mean accuracy for each condition is presented by WMC group in Figure 6.

The figure suggests that the marginal effect of WMC on the interaction of RC type and animacy configuration is primarily due to the higher WMC participants comprehending the A(I) condition more accurately than the lower WMC participants did, especially in the OR condition. This result is consistent with the pattern of reading times in the parallel condition (i.e., the A(I) condition) at the MV+1 region in Experiment 2E, where the lower WMC participants took particularly longer to read the ORs with animate heads and inanimate RC-internal nouns than the SRs with the same animacy configuration, as compared with the higher WMC participants. The results may suggest that the lower WMC participants experience particular difficulty when the available information from different sources is conflicting.

Table 15 Logistic mixed effects model for the comprehension accuracy in Experiment 3E

	Estimate	SE	z
(Intercept)	1.88	0.17	11.196***
RC type	-0.97	0.17	-5.732***
Animacy	0.74	0.26	2.899**
cWM	0.02	0.01	2.373*
RCtype \times Animacy	-0.06	0.34	-0.190
RC type \times cWM	-0.00	0.02	-0.302
Animacy \times cWM	0.01	0.02	0.826
RC type \times Animacy \times cWM	0.05	0.03	1.708 [†]

The model included the random intercepts for participants and items, and the by-items random slope for animacy configuration.

**Figure 6 Mean comprehension accuracy by WMC group in Experiment 3E**

6.3.3.2. Reading time results

Figure 7 presents the mean reading times at each region by condition. The general pattern shows a clear OR penalty, but not as large animacy configuration effects as were shown in Experiment 2E, although there was a hint of increased difficulty for the ORs in the A(I) conditions as compared with those in the A(A) condition at the RC region, and for the SRs in the A(I) condition as compared with those in the A(A) condition at the MV region. To assess these patterns, a linear mixed effect model was fitted to the reading time data from each of the three critical regions, the results of which are presented in Table 16.

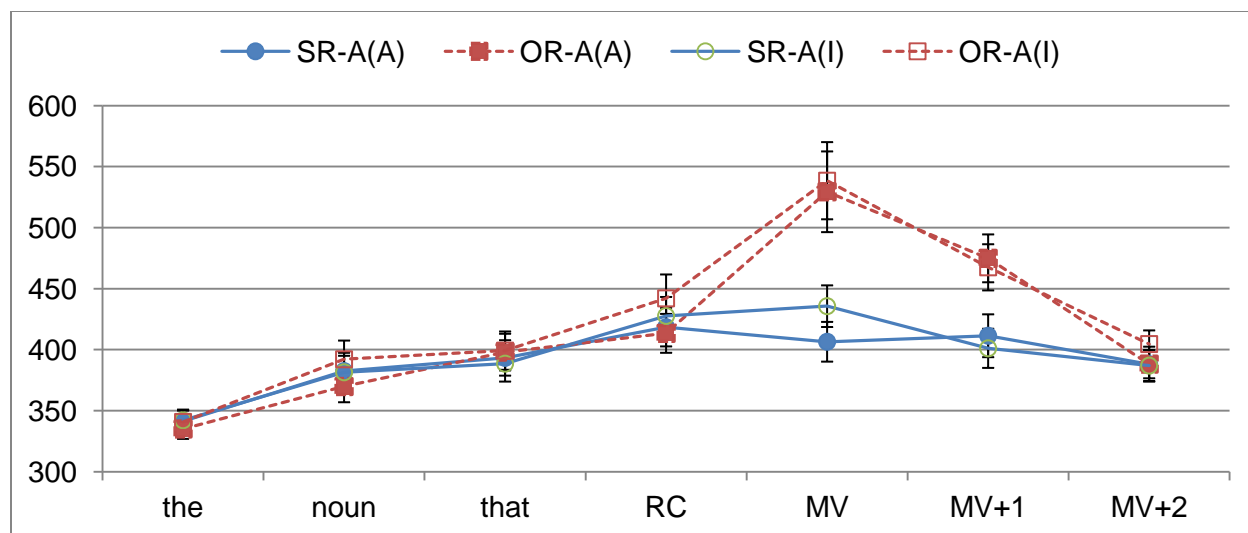


Figure 7 Mean reading times by region and condition in Experiment 3E

At the RC region, there was a reliable main effect of animacy configuration, indicating the A(I) conditions were read more slowly than the A(A) conditions. The main effect of RC type was not reliable, nor was the interaction of RC type and animacy configuration. The result suggests that both SRs and ORs were easier to read when the RC head and the RC-internal noun were both animate than when the RC head was animate and the RC-internal noun was inanimate. There was also a reliable main effect of WMC, with higher WMC associated with generally faster reading. Finally, the 3-way interaction of RC type, animacy, and WMC reached significance. To explore the pattern of this interaction, the participants were divided at the median cWM score, with each group's mean reading times presented in Figure 8. First, the figure shows that the lower WMC participants generally took longer to read the test sentences as indicated by the reliable main effect of WMC in the main analysis. Second, the 3-way interaction seems to be mainly due to the lower WMC participants being affected by the animacy manipulation to a greater extent than the higher WMC participants. The lower WMC participants showed comparable difficulty for the two SR conditions with different animacy configurations, while they found the ORs with the inanimate subjects harder than the ORs with animate subjects. The higher WMC participants, however, did not show a noticeable degree of asymmetry in the relative difficulty of the SRs and ORs across the two animacy conditions, merely showing an overall reading advantage for the A(A) conditions over the A(I) conditions.

At the MV and MV+1 region, the main effect of RC type alone was reliable with all the other main effects and their interactions not reaching significance. The result indicates that the ORs were harder to read than the SRs in a similar magnitude across the two animacy configuration conditions and across individuals with different WMC at the two words following the RCs.

Table 16 Linear mixed effects model for the reading times at each critical region in Experiment 3E

Region	RC		MV		MV+1	
Fixed effects	Estimate (SE)	t	Estimate (SE)	t	Estimate (SE)	t
(Intercept)	424.99 (14.74)	28.84	476.93 (20.90)	22.816***	437.97 (15.94)	27.47***
RC type	1.99 (7.36)	0.27	111.21 (22.84)	4.869***	63.67 (10.53)	6.05***
Animacy	18.26 (7.36)	2.48*	18.68 (15.36)	1.216	-8.48 (10.53)	-0.81
cWM	-3.15 (1.33)	-2.73**	-2.64 (1.91)	-1.379	-1.85 (1.46)	-1.27
RC:Ani.	18.34 (14.71)	1.25	-19.88 (10.72)	-0.647	2.90 (21.10)	0.14
RC:cWM	0.24 (0.67)	0.35	-0.05 (2.09)	-0.024	1.46 (0.96)	1.52
Ani.:cWM	-0.68 (0.67)	-1.02	0.31 (1.40)	0.028	0.48 (0.96)	0.50
RC:Ani.:cWM	-2.67 (1.34)	-1.99*	-1.59 (2.80)	-0.569	0.55 1.92	0.29

The models for the RC region and the MV+1 region included random intercepts for participants and items. The model for the MV region included a by-participants random slope for RC type as well as random intercepts for participants and items.

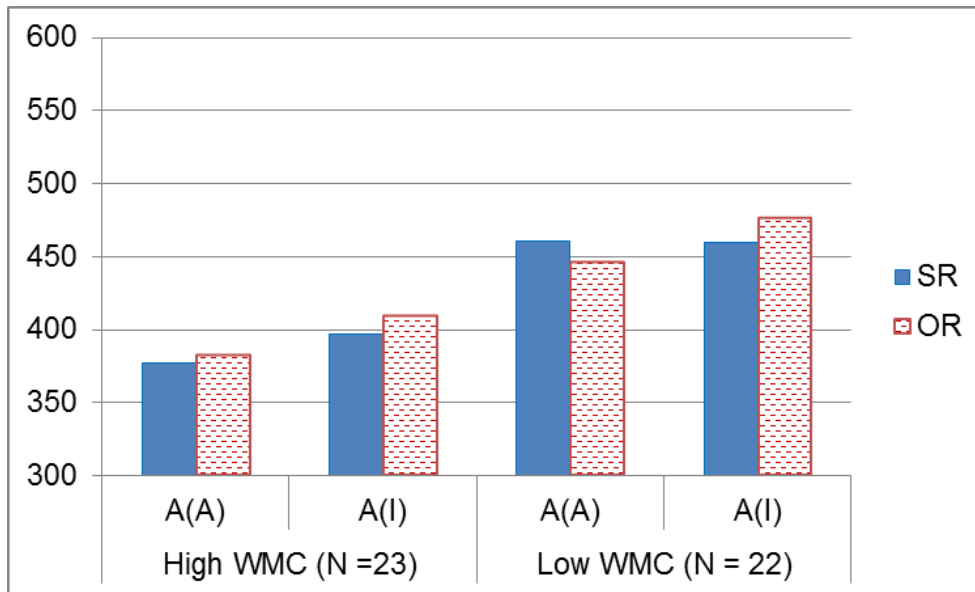


Figure 8 Mean reading times at the RC region by WMC group in Experiment 3E

As a further analysis, the RC region of the two OR conditions was compared on a word-by-word basis in order to explore when the effect of noun animacy emerges within the OR. Figure 9 shows the mean reading times at each word by condition. A mixed effects model was fitted for each word. At the first word of the RC region (i.e., *the*), the reading time in the OR-A(A) condition was marginally shorter than that in the OR-A(I) condition (Beta = 15.07, SE = 8.45, $t = 1.783$, $p < .1$), which may be a spurious result since the material up to this point in the two OR conditions were identical. The effect of animacy configuration was not significant at the RC subject (e.g., *actor* or *movie*) ($t = 1.581$, $p > .1$). It became reliable at the RC verb (Beta = 41.56, SE = 20.03, $t = 2.075$, $p < .05$), however, indicating that the ORs with animate head nouns were easier to read when the subjects were animate than when they were inanimate. The result suggests that processing difficulty due to the animacy configuration emerges within the RC region, which seems to be quite short-lived as shown in the null effect of animacy in the following regions.

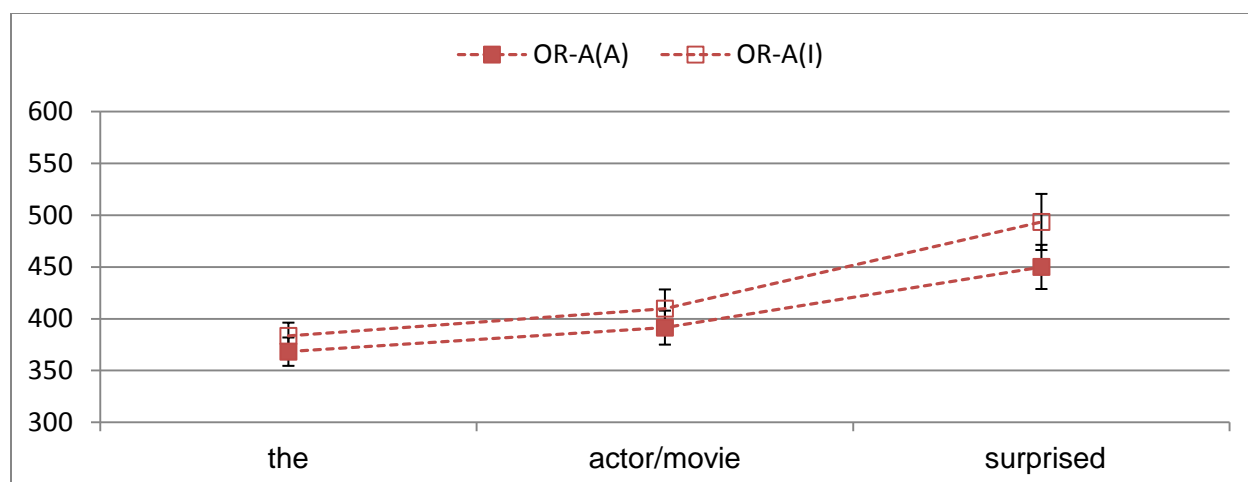


Figure 9 Mean reading times at each word within the ORs in Experiment 3E

Experiment 3E aimed to further investigate the sources of processing difficulty involved with ORs and the role of WMC in sentence processing with native speakers of English reading RCs in the L1. The head nouns of the RCs were kept animate to isolate the effect of the animacy of RC subjects on the relative difficulty of reanalysis required in the processing of ORs. The reanalysis-based account predicts that the ORs with an animate subject will be easier to process than those with an inanimate subject, because animate nouns are better subjects than inanimate nouns and thus should facilitate the reanalysis process by which the overt RC subject replaces the RC head that has been initially misanalysed as subject of the RC (Traxler et al., 2002, 2005). The interference-based account, on the other hand, predicts that ORs (as well as SRs) will be harder to comprehend when both RC heads and RC-internal nouns are animate and thus fully confusable compared to when they are distinct in animacy and thus induce less interference.

The comprehension accuracy data show that the participants were more accurate in comprehending the sentences containing an SR than those containing an OR, suggesting that ORs cause greater processing difficulty than SRs. The result replicates the OR penalty in final comprehension shown in Experiment 1E. More interestingly, the RCs with animate RC-internal nouns (i.e., A(A) condition) resulted in lower comprehension accuracy than those with inanimate RC-internal nouns (i.e., A(I) condition) both in the SRs and in the ORs, with the numerically lowest comprehension accuracy for the ORs with animate subjects. The results are consistent with the prediction of the interference-based account but not with the prediction of the reanalysis-based account. In the ORs with fully confusable nouns, the intervening RC subject

will cause interference when the RC head should be retrieved from memory to be integrated to the RC verb. But it is not possible for this kind of interference to occur within an SR because there is no intervening material between the RC verb and its arguments. The increased difficulty with the sentences containing an SR with fully confusable nouns instead might be due to the interference incurred at the main clause verb of the test sentences, at which point the main clause verb must integrate the RC head as its subject across the very similar noun phrase within the RC.

The online reading time results, however, appear to suggest a rather different story. There was evidence within the RC region that the RCs with fully confusable nouns (i.e., A(A) conditions) are read faster than those with nouns distinct in animacy (i.e., A(I) conditions), as indicated by the reliable main effect of animacy configuration and no interaction of animacy and RC type at the RC region. The result is consistent with the prediction of the reanalysis-based account but not with the interference-based account. The greater ease for the ORs with fully confusable nouns is readily explained by the facilitation of reanalysis by the animate RC subject. But the overall ease of reading the SRs with two animate nouns is not readily explainable by the reanalysis-based account because SRs do not involve reanalysis. The effect might be due to unidentified lexical properties of the RC verbs or plausibility differences across conditions. A possibility, for example, is that for some verbs used in the present experiment (e.g., *hate*, *love*, *fear*, *applaud*, etc.), animate entities could be a better Theme/object than inanimate entities.

There was no reliable evidence in reading time that RCs with fully confusable nouns cause greater processing difficulty than those with distinct nouns at the regions following the RC region, which would undermine the empirical adequacy of the interference-based account. However, there was suggestive evidence for greater processing difficulty with SRs with fully confusable nouns than those with distinct nouns at the MV region as can be seen in Figure 7. A mixed effects model fitted to the data from the SRs alone with animacy as the only fixed effect showed that this difference was reliable ($\text{Beta} = 33.96$, $\text{SE} = 13.54$, $t = 2.508$, $p < .05$)¹⁶. This might be due to some interference that arises when the sentential subject (i.e., the RC head) must be integrated with the main clause verb, as was suggested when discussing the lower comprehension accuracy for SRs with confusable nouns than for those with distinct nouns.

To summarize, the results of Experiment 3E provide evidence for the reanalysis-based account from the reading time data within the RC region. On the other hand, there was also

¹⁶ The model included random intercepts for participants and items as the random effects.

evidence for the interference-based account mainly in final comprehension accuracy and also in the reading speed at the regions following the RC region. How can these apparently conflicting pieces of evidence be resolved? The composite approach to the question as to the sources of OR penalty as formulated by Staub (2010) may be relevant here. According to Staub, there are at least two sources of difficulty that are responsible for the well-established OR penalty. One source is the violation of expectation that arises as soon as it becomes clear that a noun phrase is followed by a less expected OR rather than by a more expected SR construction, which would be when the subject of the OR is encountered. This prediction is also consistent with what the reanalysis-based account predicts, though from different theoretical perspectives. The other source lies in the memory limitations that are responsible for greater difficulty in integrating distant elements and for increased interference between similar elements at retrieval. Difficulty due to integration and interference would occur only at and/or after the processor encounters a lexical subcategorizer (e.g., the verb of an RC) that triggers the retrieval and integration of the preceding syntactic dependent (e.g., the head of an RC). The results of the present experiment seem to show a similar pattern of temporal dissociation between two different kinds of effects in regard to the OR penalty: the increased reading time for the ORs with inanimate RC subjects (i.e., OR-A(I) condition) compared with the ORs with animate subjects (i.e., OR-A(A) condition) and the greater comprehension difficulty for the ORs with confusable nouns (i.e., OR-A(A) condition) compared with the ORs with distinct nouns (i.e., OR-A(I) condition). In other words, the effect predicted by the reanalysis-based account appeared during online processing at the RC region, whereas the effect predicted by the interference-based account appeared mainly in the final comprehension data (as well as in the reading time data for the SRs *after* the RC region). Therefore, we might conclude that an animate RC subject indeed facilitates the reanalysis process involved with ORs even when the head noun is animate, resulting in faster processing of the ORs with two animate nouns. The interference between similar elements would be minimal during these processes because they are processed consecutively and thus the memory trace for each element should still be highly active and distinct (Traxler et al., 2005). However, when the already processed material must be searched and retrieved from memory after a substantial time lapse, either for interpreting a distant syntactic dependent or for finding an answer to a comprehension question, then confusable nouns would induce increased processing difficulty.

Finally, Experiment 3E provides further evidence for the role of WMC in sentence processing. First, WMC modulated the overall comprehension accuracy, which has been consistently observed in Experiment 1E and 2E. The result suggests that WMC is a reliable predictor of one's ability to explicitly find the correct answers for comprehension questions based on what he or she remembers from the results of online sentence processing. Second, as in Experiment 2E, there was evidence that WMC is associated with the ability to incorporate the information coming from different sources during sentence processing. As shown in Figure 6, the lower WMC participants experienced greater difficulty than the higher WMC participants in final comprehension in the OR-A(I) condition, where noun animacy information was in conflict with syntactic structure information. In the same light, as shown in Figure 8, the lower WMC participants showed generally greater sensitivity to the animacy manipulations in online reading compared to the higher WMC participants.

6.4. General discussion of L1 processing of English RCs

In this chapter, the results from three self-paced reading experiments with native speakers of English were reported, across which the role of syntactic complexity, noun animacy configuration, and individual differences in WMC were examined in the processing of RCs. In Experiment 1E, the participants showed evidence for the traditional SR-OR asymmetry in decreased comprehension accuracy and increased reading time for the ORs than for the SRs. Individual differences in WMC were associated with the overall accuracy in comprehension but not with syntactic complexity effects in either final comprehension or in online reading.

In Experiment 2E, the animacy of RC heads and that of RC-internal nouns were contrasted to test the role of noun animacy information in processing RCs. The results show that the traditional OR penalty disappears when the RC head is inanimate and the RC-internal noun is animate (e.g., *the movie that the director didn't praise*). The result is consistent with the reanalysis-based account that predicts that the reanalysis from the initially preferred SR analysis to the correct OR analysis is facilitated by the animate RC-internal noun because it is a good subject by being animate, and this kind of semantic information is appreciated by the comprehender at the stages of sentence processing during which reanalysis is carried out. Individual differences in WMC were associated with overall comprehension accuracy as in Experiment 1E. In addition, online reading speed was affected by WMC in such a way that

higher WMC participants are better at resolving the conflicts between information from different sources as in the ORs with animate heads and inanimate subjects.

In Experiment 3E, the head nouns of RCs were kept animate and the animacy of RC-internal nouns was manipulated in order to further test the role of noun animacy configuration in the process of reanalysis in OR processing. In the online reading time data at the RC region, there was evidence that inanimate RC-internal nouns induce greater processing difficulty than animate RC-internal nouns, with the asymmetry appearing to be more generalizable for the ORs across comprehenders with different WMC. The result is consistent with the reanalysis-based account that predicts a facilitation of reanalysis by animate RC subjects. But there was also evidence for the interference-based accounts in final comprehension, as reflected in the generally lower accuracy in comprehending the sentences containing the RCs with nouns fully confusable in terms of animacy. A hint of interference effect was also observed for the SRs at the main clause verb region in the reading time data. The results might suggest that multiple sources of difficulty conspire to bring about the OR penalty, which will be further discussed in Section 6.4.1. Individual differences in WMC modulated the overall accuracy in Experiment 3E, as was the case in Experiments 1E and 2E. But there was also a trend that WMC is associated with the experimentally manipulated syntactic complexity and noun animacy in final comprehension, unlike Experiments 1E and 2E. This trend was due to the fact that the lower WMC participants experienced greater difficulty in correctly comprehending the sentences containing an OR with an animate head and an inanimate subject, where the greatest difficulty is expected due to the conflicting information from a syntactic preference for an SR analysis and a semantic preference for an animate subject. The modulation of WMC was also evident in the online reading times at the RC region due to the fact that the lower WMC participants showed a more pronounced interaction between syntactic structure and noun animacy configuration, which is roughly consistent with the effects shown in the final comprehension data in Experiment 3E and in the online reading data in Experiment 2E.

6.4.1. Processing difficulty due to structure-based computation vs. lexical integration

In Section 4.1, several accounts for the processing asymmetry between SRs and ORs were reviewed, and following Staub (2010), a distinction was made between subsets of accounts on the basis of the kind of processes that each subset of accounts highlight as the main sources of OR penalty. The first set of accounts, which are represented by the reanalysis-based account for

the sake of discussion in the present dissertation, attribute the OR penalty to those processes that occur before encountering the verb within the RC (e.g., Gennari & MacDonald, 2008; Levy 2008; Traxler et al., 2002, 2005). These processes allow for the incremental construction of sentential representations exclusively based on the structural configurations of nouns (and other function words such as a complementizer) relevant to the interpretation of ORs, though with the exact mechanisms being different across specific accounts. Another set of accounts, which are represented by the interference-based account for the present purposes, prioritizes instead the processing complexity caused by memory limitations during the course of integrating syntactic dependents such as a lexical subcategorizer and its arguments (e.g., Gibson, 2000; Gordon et al., 2001, 2004; Van Dyke & Lewis, 2003). These two different kinds of processing difficulty highlighted by the two subsets of accounts may not be mutually exclusive because both structural computation and lexical integration would be inherent aspects of sentence interpretation. Consistent with this possibility, Staub (2010) presented evidence that both kinds of processing difficulty can be detected with different eye-movement measures. Specifically, the difficulty due to the non-canonical structure (e.g., violation of expectation) at the point of the RC subject in the OR was detected in increased regressive eye movements as compared to the RC object in the SR. On the other hand, the OR penalty arising at the RC verb in the OR took the form of increased reading time as compared to the same verb in the SR.

The results of the present experiments with native speakers of English also suggest that both structure-building and lexical integration processes could be relevant sources of the OR penalty and modulating effects of noun animacy, but they seem to further suggest that the former type of processes may be the *major* source of the online OR penalty. Evidence comes from the finding that across the three experiments, the OR penalty as reflected in reading time was modulated by the ways in which the reanalysis-based account would predict, but not in the ways in which the interference-based account would predict. Specifically, the traditional OR penalty as shown in Experiment 1E was eliminated or reduced when the subject of an OR was animate in Experiments 2E and 3E regardless of the confusability of the two relevant nouns, supporting the reanalysis-based account. On the other hand, the presence of two nouns that are distinctive in animacy in ORs with animate heads and inanimate subjects (i.e., OR-A(I) conditions) did not reduce the magnitude of OR penalty in Experiment 2E and actually *increased* that in Experiment 3E as compared to the ORs with two nouns that were both animate, inconsistent with the

interference-based account. However, a pattern of difficulty as would be predicted by the interference-based account was obtained in the comprehension accuracy data in Experiment 3E, where the sentences containing the RCs with confusable nouns were more difficult to correctly comprehend. There was also some evidence that confusable nouns can induce processing difficulty in online sentence processing, which was restricted to the main clause verb region of the SRs.

The apparently major role of structural processes in causing the OR penalty might be simply due to the fact that the relevant information is available earlier in the sentence than the trigger of the relevant lexical processes that would bring about interference. If it is the case, any effects that are attributable to increased interference between the two animate nouns would have been masked by the effects resulting from difficulty involved with earlier structure-building. Alternatively, the nouns relevant to the interpretation of simple ORs as represented by the materials of the present study may not cause so much interference- or integration-related difficulty as to incur the traditional OR penalty because there are just two of them and the distance between the head and the verb of an OR may not be so great with only one intervening discourse referent (Traxler et al, 2005; p. 206). Thus the interference-based effect may be localized to the later part of a sentence and final comprehension, when the semantic memory of the processed material should be searched. As time lapses after processing an RC, the memory traces for the nouns may gradually decay and there may be more information obtained from recently processed material in memory, both of which would increase interference.

Regardless of the exact mechanisms in which the different sources of processing difficulty interact, the present experiments contribute more evidence for the view that multiple sources of processing difficulty are responsible for the OR penalty. In addition, the results of the present experiments are free from the potential problem of comparing words in different positions in sentences as was the case in Staub (2010), because the crucial evidence for the difficulty arising from structural computation comes from the comparison of words at the same sentential position (i.e., an adverb preceding the RC verb) (cf. Gennari & MacDonald, 2008).

Now that we established how native speakers of English respond to the RCs with different animacy configurations, we are ready to test whether adult L2 learners show qualitatively the same behaviors as native speakers. If adult L2 learners process sentences incrementally, incorporating different types of information in fundamentally the same ways as

native speakers, they will show the same pattern of online processing difficulty that are primarily consistent with what the reanalysis-based account predicts. But if they prioritize the lexical integration between lexical subcategorizers and their dependents due to their reduced ability to appreciate the structural configuration of constituents as would be predicted by the shallow structure hypothesis (SSH) (e.g., Clahsen & Felser, 2006), we would expect that adult L2 learners will be more sensitive to the difficulty associated with lexical integration and thus they will be more vulnerable to interference effects than the native speakers were shown to be in the present experiments.

6.4.2. (Partial) dissociation between online sentence processing and WM span

Another major goal of the present experiments with native speakers of English was to further explore the role of individual WMC differences in online sentence processing and to establish the native speaker norm against which to compare the role of L2 WMC in L2 sentence processing. WMC was operationalized as the mean score of the two WM span tests to increase the reliability of the WM measures: the reading span and subtract-2 span tests.

The overall results of the present experiments seem to be consistent with a version of the separate-resource model of WMC as proposed in Traxler et al. (2005), according to which the cognitive processes involved in the early stages of sentence processing draw upon a dedicated pool of processing resources that are not tapped by traditional WM span measures, whereas some later processes in sentence processing responsible for evaluating non-structural information do draw upon the same processing resources as measured by WM span tests (cf. Caplan & Waters, 1999; Waters & Caplan, 2004). Across the three experiments with native English speakers, individual differences in WMC were consistently associated with overall comprehension accuracy in such a way that individuals with higher WMC were more accurate in final comprehension. Thus the ability to maintain the propositions read off from sentences and later search the semantic memory for the needed information seems to be reflected in one's WMC. On the other hand, syntactic complexity effects as reflected in OR penalty were rarely modulated by WMC either in comprehension accuracy or in online reading in the three experiments. It was only when the OR penalty was further moderated by noun animacy information that any effects due to WMC emerged, as observed in Experiments 2E and 3E. When the effect of WMC emerged in online reading speed, the pattern of the effect was consistent with the possibility that individuals with higher WMC are more efficient in incorporating multiple sources of information,

especially when the information from different sources are in conflict (cf. Pearlmutter & MacDonald, 1995). In Experiment 2E, for example, the main source of the WMC effect on reading times detected at the RC region seemed to lie in the lower WMC participants taking particularly long to read the SRs with inanimate heads and animate objects (e.g., *the movie that didn't please the director*), which displays non-canonical animacy – grammatical role mappings (i.e., an inanimate subject and an animate object).

The methods of the present experiments successfully captured the hypothesized distinction between the early structural processes responsible for the traditional OR penalty and the later reanalysis processes responsible for semantic evaluation of an interpreted sentential representation, as reflected in their dissociation/association with WMC as measured by the two WM span tests. If we take the dissociation between WMC and the early structural processes in sentence processing as evidence for the special ‘linguistic’ nature of those processes, then we might use the materials and design of the present experiments to test whether L2 sentence processing also involves such linguistic processes. The investigation of this issue could provide additional evidence on the nature of L2 sentence processing and would offer helpful insights in regard to the evaluation of the SSH that assumes qualitative differences in the structural computation implemented in L1 and L2 sentence processing. Now we turn to the experiments with L1-Korean L2-English learners.

Chapter 7

Experiments with L2 learners of English

In this chapter, three experiments with L1-Korean L2-English learners are presented. These experiments are labeled as Experiment 1K, 2K, and 3K. The materials and procedures for the experiments were the same as the corresponding experiments with native English speakers (i.e., Experiments 1E, 2E, and 3E), except that there were additional tasks such as the cloze test and the vocabulary knowledge check.

7.1. Experiment 1K

Experiment 1K tested the relationship between WMC and syntactic complexity effects in L2 sentence processing with 32 L1-Korean L2-English learners.

7.1.1. Materials and design

The materials and design were identical with Experiment 1E. The test sentences are repeated below as (68) – (71) for reference.

(68) The banker that irritated the lawyer played tennis every Saturday.

(69) The banker that the lawyer irritated played tennis every Saturday.

(70) The lawyer that irritated the banker played tennis every Saturday.

(71) The lawyer that the banker irritated played tennis every Saturday.

7.1.2. Predictions

It was predicted that the L2 learners would find ORs more difficult to comprehend than SRs as has been shown in previous studies (e.g., Hashimoto, 2007; Omaki & Ariji, 2005). If adult L2 learners rely more on the processing resources that are used for general cognitive functions due to some deficit in linguistic processes in the L2, then individual differences in online sentence processing might be modulated by individual differences in WMC. If, on the other hand, they use some kind of dedicated resources for L2 sentence processing as the native speakers seemed to do in Experiment 1E (or if syntactic processes in the L2 can become fairly automatic), then there should be a similar pattern of relationship between sentence processing and WMC as shown in L1 sentence processing, where WMC was associated with overall comprehension accuracy but not with syntactic complexity effects.

7.1.3. Results and discussion

Items that contained any crucial nouns and verbs unknown to a participant as identified by the vocabulary check list were excluded from analysis in the present and subsequent experiments. Mixed effects models were used for statistical analyses (e.g., Baayen et al., 2008; Jaeger, 2008). In order to control for the effect of general proficiency in English, the main effect of cloze test score and its interactions with the other predictor(s) were tested using the log-likelihood ratio test (e.g., Baayen et al., 2008). The models presented hereafter thus include as the fixed effects the main effects of the experimental factors in the respective experiments (RC type and cWM in Experiment 1K; RC type, animacy, and cWM in Experiments 2K and 3K) and their interactions by default, as well as the effects related to cloze test score (i.e., its main effect and interactions with the other effects) that significantly increased the fit of the model.

7.1.3.1. *Comprehension accuracy*

Mean accuracy was 81.8% (SD = 13.3) for the SR condition and 74% (SD = 16.0) for the OR condition. A logistic mixed effect model was fitted to assess this pattern and the effect of WMC. The estimated fixed effects from the selected model are presented in Table 17. The main effect of RC type was reliable, indicating that ORs were harder to comprehend overall than SRs. The main effect of cWM was also reliable, indicating that higher WMC was associated with more accurate comprehension. There was no interaction of the two effects, suggesting that the magnitude of the OR penalty as reflected in decreased comprehension accuracy did not systematically differ across L2 learners with different WMC. The main effect of cloze test score alone contributed to the model, indicating that the higher proficiency learners were more accurate in comprehension than the lower proficiency learners, regardless of the syntactic complexity of the test sentences or individual differences in WMC.

Table 17 Logistic mixed effects model for the comprehension accuracy in Experiment 1K

	Estimate	SE	z
(Intercept)	1.375	0.131	10.54***
RC type	-0.478	0.186	-2.57*
cWM	0.031	0.008	4.12***
RC type:cWM	0.010	0.015	0.66
Cloze	0.042	0.014	2.89**

The model included random intercepts for participants and items.

7.1.3.2. Reading time results

Figure 10 presents the mean reading times at each region by condition. The ORs were read more slowly than the SRs at the RC and MV regions. To assess this pattern, a mixed effect model was fitted at each of the three critical regions. The results are presented in Table 18.

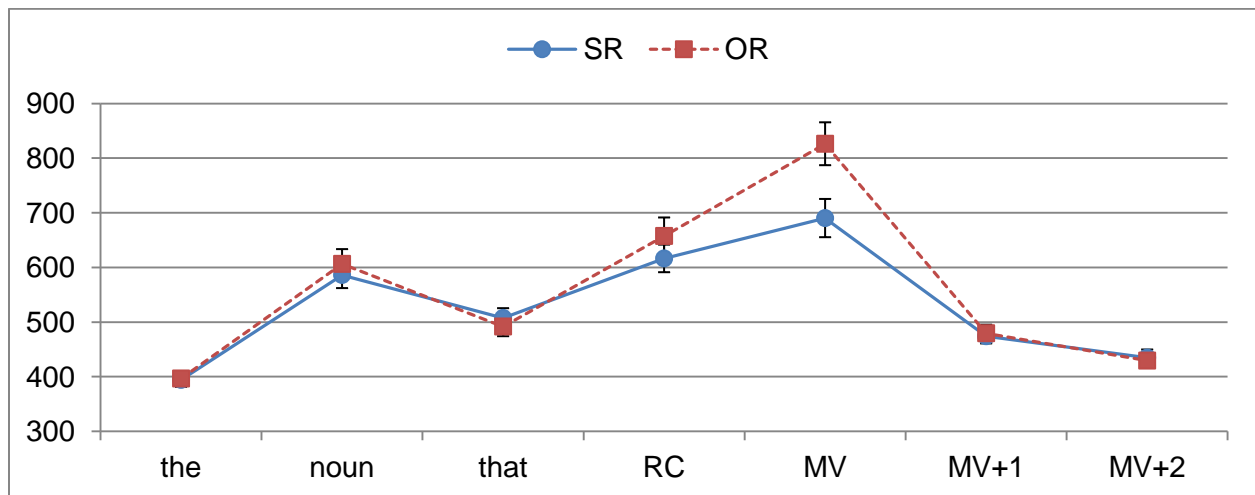
**Figure 10 Mean reading times by region and condition in Experiment 1K**

Table 18 Linear mixed effects model for the reading times at each critical region in Experiment 1K

Region	RC		MV		MV+1	
Fixed effects	Estimate (SE)	t	Estimate (SE)	t	Estimate (SE)	t
(Intercept)	641.47 (27.11)	23.66***	758.73 (34.94)	21.72***	479.75 (16.03)	29.93***
RC type	45.15 (16.22)	2.78**	137.24 (34.76)	3.95***	6.78 (13.10)	0.52
cWM	3.15 (2.11)	1.50	4.77 (2.62)	1.82 [†]	0.62 (1.00)	0.63
RC:cWM	1.55 (1.33)	1.17	-0.57 (2.82)	-0.20	0.26 (1.07)	0.24
Cloze	-12.39 (4.47)	-2.77**	n/a	n/a	-4.11 (2.10)	-1.96*
Cloze:RC	-2.24 (2.83)	-0.79	n/a	n/a	n/a	n/a
Cloze:cWM	-0.68 (0.35)	-1.92 [†]	n/a	n/a	-0.38 (0.17)	-2.31*
Cloze:RC:cWM	-0.53 (0.22)	-2.40*	n/a	n/a	n/a	n/a

Each model included random intercepts for participants and items.

n/a represents an effect that did not contribute to the model according to log-likelihood ratio tests.

* $p < .05$, ** $p < .01$, *** $p < .001$, [†] $p < .1$

At the RC region, the main effect of RC type was reliable, indicating that the ORs were read more slowly than the SRs. Neither the main effect of cWM nor the interaction of RC type and cWM reached significance. The main effect of cloze test score was reliable, with the negative coefficient indicating that the higher proficiency learners read faster than the lower proficiency learners. The interaction of cloze and cWM was marginally significant, reflecting the association of higher proficiency with faster reading on the one hand, and the trend that higher WMC is associated with slower reading (as reflected in the positive coefficient of the main effect of cWM, which was not significant at the RC region). But this interaction was further moderated by a reliable 3-way interaction of RC type, cWM, and cloze.

In order to explore the 3-way interaction of RC type, cWM and cloze, participants were divided into two groups at the median cloze test score (20 higher proficiency learners and 12 lower proficiency learners), and a separate mixed effects model was fitted to each proficiency

group with RC type and cWM as well as their interaction as the fixed effects. The estimated fixed effects are presented in Table 19. The results show relatively weaker effects of interest, probably due to the small number of participants in each group, but the trend of greater difficulty with the ORs was maintained. A notable difference between the two proficiency groups was that the interaction of RC type and cWM was marginally significant for the lower proficiency learners while it was not significant with the higher proficiency learners. The marginal interaction of the two factors for the lower proficiency learners was due to the trend that as WMC increases, the difference between the SR and the OR condition becomes larger¹⁷. Though this trend is suggestive at best, it might suggest that for the lower proficiency learners, a certain level of WMC must be available to incrementally construct a sentence structure and thus show immediate online behavioral signs of processing difficulty. As will be shown shortly, the learners showed an online OR penalty at the following MV region regardless of WMC. Thus, the absence of OR penalty for the lower WMC learners with the lower proficiency at the RC region should be seen as evidence for their relatively inefficient/slow structural computation rather than their inability to process incrementally. But if we take this as suggestive evidence for the overlap of the processing resources underlying lower proficiency learners' syntactic computation and those underlying their performance in the WM span tests, it may suggest the interesting possibility that as proficiency increases, the nature of the processing resources underlying syntactic computation becomes more native-like.

Table 19 Linear mixed effects models for the reading times at the RC region by proficiency group in Experiment 1K

	High proficiency (N = 20)			Low proficiency (N = 12)		
	Estimate	SE	t	Estimate	SE	t
(Intercept)	593.32	32.91	18.03***	705.39	51.94	12.58***
RC type	34.61	19.35	1.79 [†]	44.57	29.31	1.52
cWM	0.84	2.68	0.31	3.43	3.98	0.86
RC type:cWM	-0.70	1.58	0.44	4.09	2.39	1.72 [†]

Each model included random intercepts for participants and items.

¹⁷ Among 12 lower proficiency participants, 7 were high WMC and 5 were lower WMC participants. The mean difference in reading time between the SR and the OR condition was numerically larger for the higher WMC participants (M = 84.9ms, SE = 26.8) than that for the lower WMC participants (M = 0.3ms, SE = 57.4).

At the MV region, the main effect of RC type alone was reliable, indicating the ORs took longer to read than the SRs regardless of individual differences in WMC and proficiency. The main effect of cWM was marginal, indicating a trend that higher WMC participants read more slowly than the lower WMC learners. At the MV+1 region, the main effect of RC type was no longer significant. The only significant effects were the main effect of cloze and its interaction with cWM. These effects indicate that the higher proficiency participants generally read faster, while the higher WMC participants generally read more slowly. Note that the main effect of the cloze was reliable at the RC and MV+1 region, whereas the main effect of cWM did not reach significance at any region, suggesting that the interaction is mainly due to the strong effect of cloze test scores. Still it is not clear at first glance why higher WMC is associated with generally slower reading, if higher WMC represents more efficient processing. This apparently counterintuitive result is not unprecedented however. Some earlier studies on L1 sentence processing also reported slower reading by higher WMC readers under certain circumstances (e.g., Ferreira & Patson, 2007; MacDonald, Just, & Carpenter, 1992; Pearlmutter & MacDonald, 1995). In MacDonald et al. (1992), for example, high WMC participants read temporarily ambiguous sentences such as *The soldiers warned about the dangers before the midnight raid*¹⁸ more slowly than low WMC participants. The authors explained this result as suggesting that higher WMC participants are *better* at maintaining multiple syntactic representations, which in turn taxes processing resources and slows down reading. In the present experiment, the higher WMC participants might have performed more truthful analyses of the test sentences as reflected in higher comprehension accuracy, resulting in slower reading speed. The similar results of the present experiment and previous studies seem to suggest that the direction of the effect of WMC could be modulated by the specific demands of the task and characteristics of the stimuli.

To summarize the results of Experiment 1K, the participants with higher proficiency as measured by a cloze test were generally more accurate in final comprehension and faster in online reading. But the effects of interest for the present purposes showed reliable effects even after individual proficiency was statistically controlled for. First, participants showed a clear OR penalty both in final comprehension and in online reading, suggesting that the online processes responsible for the OR penalty (i.e., a process of reanalysis) are working properly in adult L2

¹⁸ The verb of the sentence *warned* can be initially interpreted as a past participle beginning a reduced relative clause, incurring a temporary syntactic ambiguity.

learners' sentence processing. Second, the effect of WMC was evident in final comprehension accuracy such that the learners with higher WMC were generally more accurate in comprehension. But there was little evidence that WMC modulates processing complexity associated with ORs either in final comprehension or in online reading. The only indication of potential association between WMC and online sentence processing in Experiment 1E was shown in the marginal interaction of RC type and cWM for the lower proficiency learners ($N = 12$). The trend for the interaction was due to the higher WMC learners ($N = 7$) showing the greater difference in reading time between SRs and ORs than the lower WMC learners ($N = 5$). If this trend is supported by more empirical evidence in future research, we might hypothesize that a certain level of WMC is required for lower proficiency learners to compute a given syntactic structure online and thus show some behavioral signs of difficulty involved with the structural computation.

7.2. Experiment 2K

Experiment 2K tested the interaction of syntactic complexity and noun animacy configuration in the course of RC processing, as well as the effect of WMC. The participants were the same as Experiment 1K.

7.2.1. Materials and design

The materials and design of Experiment 2K were the same as Experiment 2E. The test sentences are repeated below in (72)–(75) for reference. These sentences represent the SR-A(I), OR-A(I), SR-I(A), and OR-I(A) conditions, respectively.

(72) The director that didn't praise the movie received a prize at the film festival. *SR-A(I)*

(73) The director that the movie didn't please received a prize at the film festival. *OR-A(I)*

(74) The movie that didn't please the director received a prize at the film festival. *SR-I(A)*

(75) The movie that the director didn't praise received a prize at the film festival. *OR-I(A)*

The participants also provided post-hoc plausibility ratings. Table 20 presents a summary of the results from only those participants who were included for analysis in Experiment 2K. As for the propositions expressed by the RCs, the mean rating on the propositions with an animate subject and an inanimate object was significantly higher than that on those with an inanimate

subject and an animate object ($t(27) = 4.714, p < .001$), unlike native speakers. In other words, the L2 learners found the sentences like *The director didn't praise the movie* more plausible than the sentences like *The movie didn't please the director*. There was no reliable difference between the plausibility ratings between the two types of main clause propositions ($t(27) = 1.296, p > .1$), although there was still a numerical preference for the propositions with an animate subject. The result suggests that the perceived plausibility of the same sentences can be different between native speakers and L2 learners. Setting aside the possibility of different world knowledge systems at a cultural level, the suppressed plausibility ratings on the sentences with inanimate subjects might be due to overall comprehension difficulty arising from the non-canonical mappings of noun animacy and grammatical roles in those sentences (as well as the non-canonical thematic structure associated with the so-called psych verbs used for the majority of the RCs with inanimate subjects such as *please*) (c.f. de Swart, Lamers, & Lestrade, 2008)

The differences in the rated plausibility between the propositions with animate subjects and those with inanimate subjects are confounded with the effects of the other factors under study. Specifically, the OR-A(I) condition could be more difficult than the OR-I(A) condition not only due to animacy configuration but also due to plausibility difference between the propositions expressed by the RCs in the two conditions. Therefore, caution should be taken in interpreting the results of the present experiment. However, the potential difference in plausibility between the test conditions may not ultimately be detrimental to achieving the main goals of the present study, which is to explore the interaction of syntactic complexity, animacy configurations, and individual differences in WMC in the context of RC processing. Canonicity of the associations between noun animacy and certain grammatical roles may ultimately reduce to plausibility, and if so, any effects due to plausibility in the present experiment may be of the same nature as any effects that are driven by the animacy manipulation, which is, by hypothesis, orthogonal to the syntactic manipulation.

Table 20 Mean plausibility ratings on different propositions included in the test sentences in experiment 2K

Propositions	Relevant conditions	Mean rating (SD)
RC propositions with an animate subject and an animate object e.g., The director didn't praise the movie.	SR-A(I), OR-I(A)	5.82 (1.06)
RC propositions with an inanimate subject and an animate object e.g., The movie didn't please the director.	SR-I(A), OR-A(I)	4.86 (1.02)
Main clause propositions with an animate subject e.g., The director received a prize.	SR-A(I), OR-A(I)	5.60 (0.98)
Main clause propositions with an inanimate subject e.g., The movie received a prize.	SR-I(A), OR-I(A)	5.29 (1.16)

7.2.2. Predictions

If L2 learners incrementally construct sentential representations of RCs by immediately incorporating different types of information, they will show the same pattern of interaction between RC type and noun animacy during online sentence processing as did the native speakers in Experiment 2E (e.g., Jackson & Roberts, 2010; Omaki & Arij, 2005). More specifically, if L2 learners' incremental interpretation of the syntactic/thematic roles of noun phrases in a sentence fragment can be carried out independently of their lexical subcategorizer as opposed to what a strong version of the shallow structure hypothesis (SSH) would predict, then the online interaction of syntactic structure and noun animacy will appear before the RC verb is encountered in ORs, as was shown by the native speakers in Experiment 2E.

As for the role of WMC in L2 sentence processing, if L1 and L2 sentence processing share the processing mechanisms sub-served by the same kind of processing resources, and if L1 and L2 WM span tests measure at least some aspects of the shared processing mechanisms, L2 WMC will also modulate the ability to cope with conflicting information coming from different sources as shown by the native speakers in Experiment 2E (cf. Dussias & Piñar, 2010).

7.2.3. Results and discussion

7.2.3.1. *Comprehension accuracy*

Table 21 presents the mean accuracy for each condition. The general pattern seems to be similar to that from native speakers in that the L2 learners found ORs harder to comprehend than SRs when the heads of the RCs were animate, but it was not the case when the RC heads were inanimate. A notable pattern is that comprehension accuracy is much lower for the A(I) conditions than for the I(A) conditions.

Table 21 Mean comprehension accuracy in Experiment 2K (in percent)

	A(I)		I(A)	
	SR	OR	SR	OR
M	74.6	62.5	83.5	85.7
(SD)	(19.8)	(14.9)	(11.5)	(18.5)

A logistic mixed effects model assessed these effects as presented in Table 22. First, as the bottom row of the table shows, the interaction of cloze test score and RC type significantly contributed to the default model (i.e., the model with RC type, animacy, and cWM as the fixed effects). This interaction was due to the fact that the higher proficiency participants split at the median cloze score ($N = 20$) found the SRs easier to comprehend (Mean accuracy: 81.1%, SE: 2.3) than the ORs (Mean accuracy: 72.5%, SE: 3.0), but the lower proficiency participants ($N = 12$) did not show a noticeable difference between the two conditions (SR condition: $M = 75.6\%$, $SE = 4.5$; OR condition: $M = 76.8\%$, $SE = 3.1$). The result may suggest that a certain level of proficiency is required for L2 learners to show native-like sensitivity to structural manipulations of the sort tested in the present experiment, which may not be surprising.

Table 22 Logistic mixed effects model for the comprehension accuracy in Experiment 2K

	Estimate	SE	z
(Intercept)	1.541	0.205	7.52***
RC type	-0.200	0.190	-1.04
Animacy	1.132	0.190	5.95***
cWM	0.013	0.009	1.41
RC type:Animacy	0.695	0.378	1.84 [†]
RC type × cWM	0.007	0.015	0.48
Animacy × cWM	-0.014	0.015	-0.90
RC:Ani.:cWM	0.052	0.030	1.73 [†]
Cloze	0.022	0.018	1.20
Cloze × RC type	-0.081	0.030	-2.71**

The model included random intercepts for participants and items.

Now we turn to the main points of interest of the present experiment. The main effect of RC type was not reliable, but that of animacy was highly reliable. The interaction of RC type and animacy was marginally significant which was further moderated by an also marginally significant 3-way interaction of RC type, animacy, and cWM.

To explore this trend of 3-way interaction, participants were split into higher and lower WMC groups as in Experiment 1K, with the mean accuracy for each group provided in Figure 11. The figure suggests that the modulating effect of noun animacy on the OR penalty in final comprehension was more pronounced for the higher WMC participants than for the lower WMC participants, with the former group more closely resembling the native speakers' behavior as shown in Experiment 2E. Lastly, although the main effect of cWM did not reach significance, the numerical pattern of the effect as indicated by the positive coefficient was consistent with Experiment 1K, where higher WMC was reliably associated with higher comprehension accuracy.

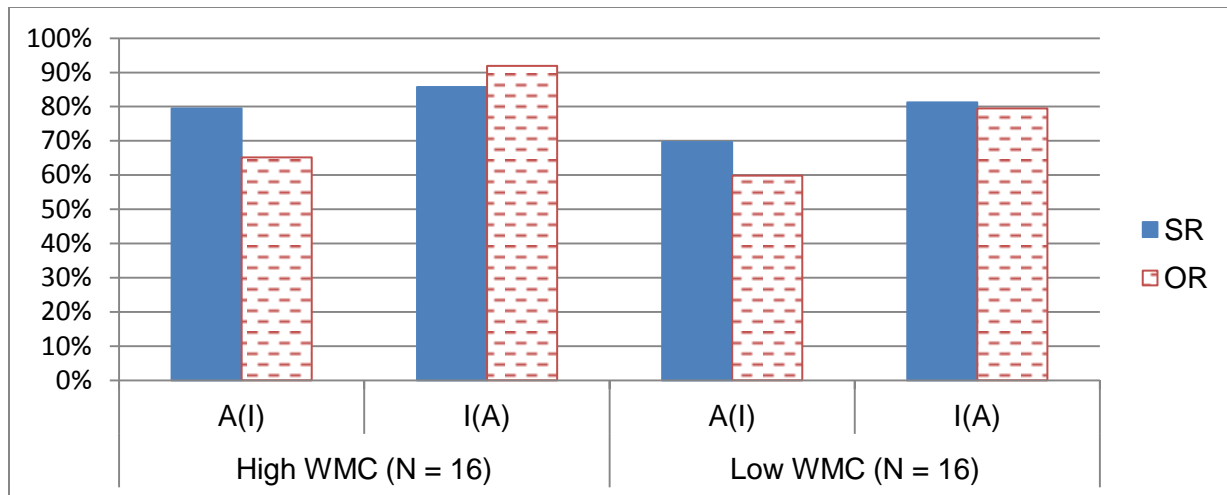


Figure 11 Mean comprehension accuracy by WMC group (in percent)

7.2.3.2. Reading time results

Figure 12 presents the mean reading times at each region by condition. The overall pattern is quite similar to the pattern shown by the native speakers in Experiment 2E. The OR-A(I) condition seems to be particularly difficult to read compared to all the other conditions. To assess this pattern, a linear mixed effects model was fitted for each critical region as presented in Table 23.

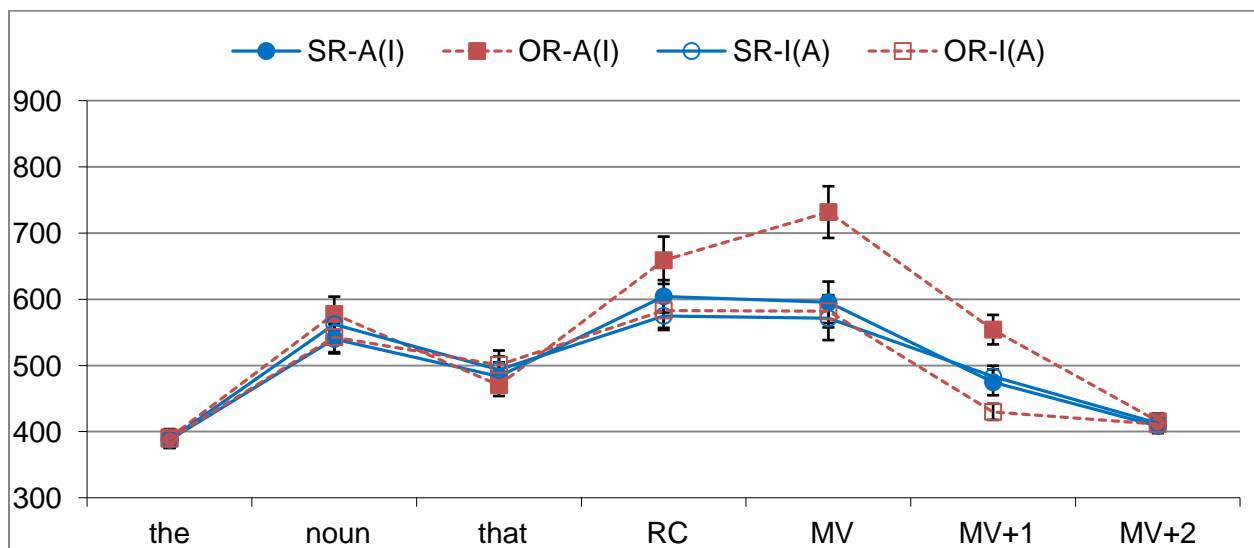


Figure 12 Mean reading times by region and condition in Experiment 2K

Table 23 Linear mixed effects model for the reading times at each critical region in Experiment 2K

	RC		MV		MV+1	
	Estimate (SE)	t	Estimate (SE)	t	Estimate (SE)	t
(Intercept)	611.28 (21.67)	28.21***	625.28 (29.42)	21.25***	488.50 (17.79)	27.45***
RC type	34.87 (12.56)	2.11*	75.92 (26.80)	2.83**	11.45 (16.51)	0.69
Animacy	-54.40 (17.02)	-3.20**	-83.73 (21.00)	-3.99***	-56.79 (13.48)	-4.21***
cWM	2.26 (1.72)	1.31	3.22 (1.90)	1.69 [†]	-0.55 (1.06)	-0.52
RC type:Animacy	-44.45 (35.22)	-1.26	-114.44 (42.04)	-2.72**	-129.37 (26.97)	-4.80***
RC type:cWM	0.34 (1.34)	0.26	0.46 (1.71)	0.27	1.87 (1.11)	1.69 [†]
Animacy:cWM	-0.53 (1.37)	-0.39	-3.47 (1.69)	-2.06*	-0.05 (1.09)	-0.04
RC:Ani.:cWM	-0.90 (2.83)	-0.32	1.59 (3.38)	0.47	-5.00 (2.18)	-2.29*
Cloze	-9.90 (3.50)	-2.83**	-6.46 (4.03)	-1.60	-1.67 (2.25)	-0.74
Cloze:RC type	-1.34 (2.48)	-0.54	n/a	n/a	n/a	n/a
Cloze:Animacy	-1.18 (2.58)	-0.46	n/a	n/a	-4.70 (2.23)	-2.10*
Cloze:cWM	-0.83 (0.28)	-3.00**	-0.59 (0.32)	-1.86 [†]	-0.37 (0.18)	-2.12*
Cloze:RC:cWM	-0.52 (0.20)	-2.67**	n/a	n/a	n/a	n/a
Cloze:Ani.: cWM	0.45 (0.20)	2.19*	n/a	n/a	n/a	n/a

The model for the RC region included random intercepts for participants and items, and by-participant random slopes for RC type, animacy, and their interaction. The models for the MV and MV+1 region included a by-items random slope for RC type as well as random intercepts for participants and items.

At the RC region, the main effect of cloze was reliable, indicating that higher proficiency is associated with faster reading. The interaction of cloze and cWM was significant, due to the fact that higher proficiency was associated with faster reading, whereas higher WMC was associated with slower reading, as was the case in Experiment 1K. The 3-way interaction of

cloze, RC type, and cWM, as well as the 3-way interaction of cloze, animacy, and cWM was reliable, suggesting that L2 learners' proficiency interacted with the experimental manipulations in complex ways. To explore the role of proficiency and make other effects more interpretable, the L2 learners were divided into higher and lower proficiency groups based on median cloze test score. Those with cloze scores lower than the median were categorized as lower proficiency learners (N = 12), with the others grouped as higher proficiency learners (N = 20). A separate mixed effects model was fitted to each group with RC type, animacy, cWM, and their interactions as fixed effects, which is presented in Table 24

Table 24 Linear mixed effects models for the reading times at the RC region by proficiency group in Experiment 2K

	High proficiency (N = 20)		Low proficiency (N = 12)	
	Estimate (SE)	t	Estimate (SE)	t
(Intercept)	571.99 (28.37)	20.16***	651.71 (36.55)	17.83***
RC type	39.59 (21.35)	1.85 [†]	8.72 (21.75)	0.40
Animacy	-64.37 (22.47)	-2.87**	-23.28 (21.55)	-1.08
cWM	-1.00 (2.31)	-0.43	4.27 (2.79)	1.53
RC type:Animacy	-85.53 (28.36)	-3.02**	20.61 (43.10)	0.48
RC type:cWM	-0.84 (1.75)	-0.48	1.51 (1.79)	0.84
Animacy:cWM	0.89 (1.83)	0.49	-2.32 (1.69)	-1.37
RC:Ani.:cWM	-4.61 (2.30)	-2.00*	3.62 (3.38)	1.07

The model for the high proficiency group included random intercepts for participants and items, and by-participants random slopes for RC type and animacy. The model for the low proficiency group included random intercepts for participants and items.

Notably, the lower proficiency group did not show any reliable effects due to the experimental manipulations, while the higher proficiency group showed several reliable fixed effects in the predicted directions. For the higher proficiency learners, the interaction of RC type and animacy was reliable, indicating that the magnitude of OR penalty differed depending on the

noun animacy configuration. This interaction, however, was further moderated by participants' WMC as indicated by a significant 3-way interaction of RC type, animacy, and cWM. In order to explore the pattern of this 3-way interaction, the higher proficiency learners were further divided into two groups based on their cWM scores¹⁹. The mean reading times in each condition are described in Figure 13 by WMC group. The figure suggests that the difference in processing difficulty between the OR-A(I) condition and all the other conditions was much larger for the high WMC participants (N = 9) than for the low WMC participants (N = 11). The result suggests that the higher WMC learners were more native-like in incorporating the different types of information efficiently within the RC region than were the lower WMC participants, producing a reading time pattern that is clearly more consistent with what the reanalysis-account would predict (i.e., the reduction or elimination of OR penalty with the ORs with inanimate heads and animate subjects). More native-like behaviors of the L2 learners with higher WMC were also shown in the comprehension accuracy data illustrated in Figure 11.

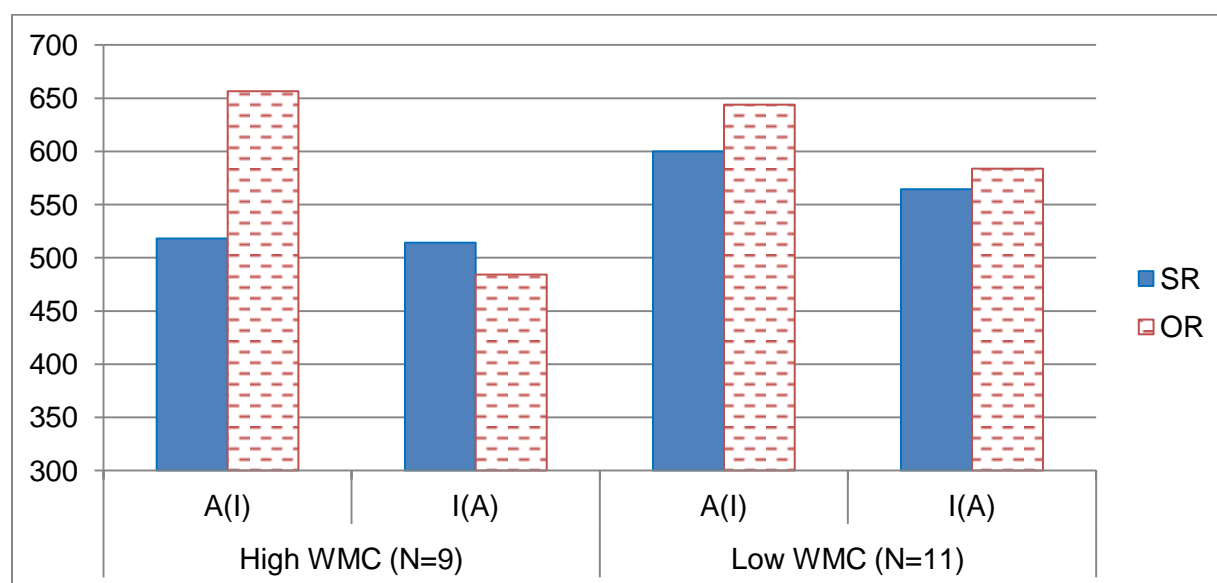


Figure 13 Mean reading times at the RC region by WMC group for the high proficiency L2 learners in Experiment 2K

At the MV region, any effects related to cloze did not reach significance. The main effects of RC type and animacy configuration were both reliable. Crucially, there was a significant interaction of RC type and animacy, indicating that the OR penalty was observable

¹⁹ The median cWM score based on all the 32 participants in Experiment 2K was used as the criterion of grouping.

only for the RCs with animate heads and inanimate RC-internal nouns as illustrated in Figure 12. This interaction was not moderated by proficiency, unlike at the RC region. It appears that the lower proficiency learners took longer to incorporate the relevant information during online sentence processing as compared to the higher proficiency learners, thus showing effects due to the experimental manipulations relatively later. The main effect of cWM was marginally significant, due to the trend that high WMC is associated with slower reading, as was the case in Experiment 1K. The interaction of animacy and cWM was reliable, indicating that participants with higher WMC read the MV region faster in the I(A) conditions than in the A(I) conditions to a greater extent than those with lower WMC. It might be related to some unidentified factors that are responsible for the greater difficulty in comprehending the sentences in the A(I) condition than those in I(A) condition, with higher WMC participants showing greater sensitivity to the differential difficulty. Though without a readily available explanation for this interaction, it seems suggestive that the effect of noun animacy configuration was associated with individual WMC differences, whereas the effect of RC type was not.

At the MV+1 region, cloze interacted with cWM as well as with animacy configuration. The former interaction was due to the two factors' opposite effects on reading time, which was also the case at the RC region. The interaction of cloze and animacy indicates that the higher proficiency learners showed a greater difference in reading time between the A(I) and I(A) conditions, which is analogous to the same pattern of interaction between animacy and cWM shown at the MV region. More importantly for our purposes, the interaction of RC type and animacy remained reliable, indicating that the asymmetrical difficulty involved with the OR-A(I) condition persists over the two words following the RC region. The interaction of RC type and cWM was marginally significant, but this interaction was further moderated by a significant 3-way interaction of RC type, animacy, and cWM, which was also reliable at the RC region for the higher proficiency learners. In order to further explore this 3-way interaction, participants were divided into higher and lower WMC groups and the mean reading times for each group are described in Figure 14. The overall pattern was quite similar to the pattern of reading times at the RC region from the higher proficiency learners as shown in Figure 13, which is consistent with the pattern predicted by the reanalysis-based account and thus closely resembles the reading time pattern of the native speakers in Experiment 2E.

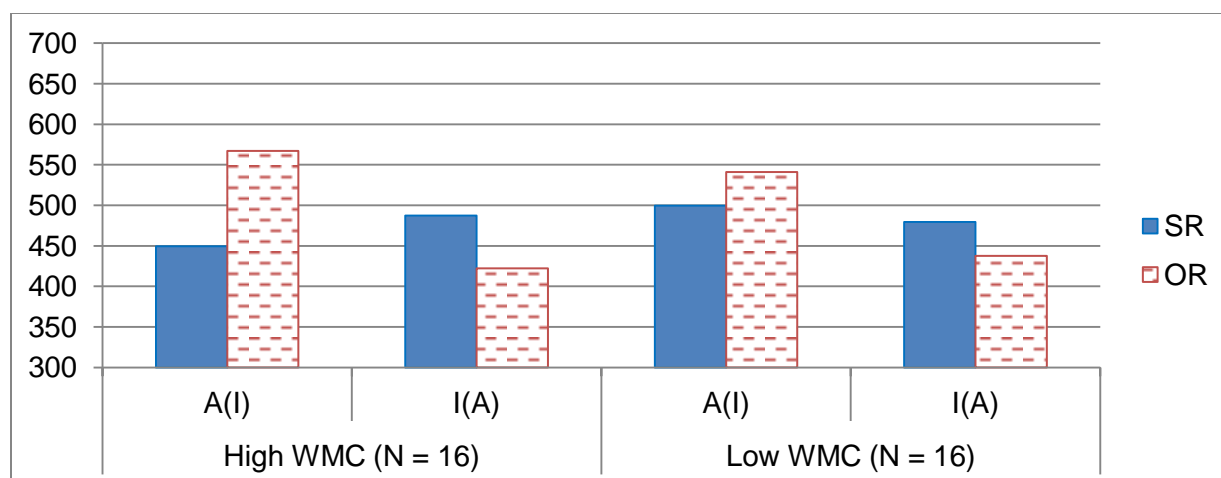


Figure 14 Mean reading times at the MV+1 region by WMC group in Experiment 2K

Lastly, the mean reading times at each word within the ORs with different animacy configurations were analyzed to test when the effect due to the relative difficulty of reanalysis in the two OR conditions began to appear for the L2 learners. Figure 15 suggests that the difficulty due to inanimate RC subjects in ORs is already evident at the adverb preceding the RC verb. The mixed effects models with animacy as the only fixed effect showed that this animate subject advantage in the OR was reliable both at the adverb and the RC verb (at the adverb, Beta = -97.45, SE = 30.85, $t = -3.16$, $p < .01$; at the RC verb, Beta = 194.65, SE = 68.74, $t = -2.83$, $p < .01$)²⁰. The result suggests that adult L2 learners, like native speakers, incrementally build sentential representations incorporating syntactic and semantic information without exclusively relying on lexical relations between lexical subcategorizers and their dependents.

²⁰ The model for the adverb region included random intercepts for participants and items as random effects. The model for the RC verb region included a by-participants random slope for animacy as well as random intercepts.

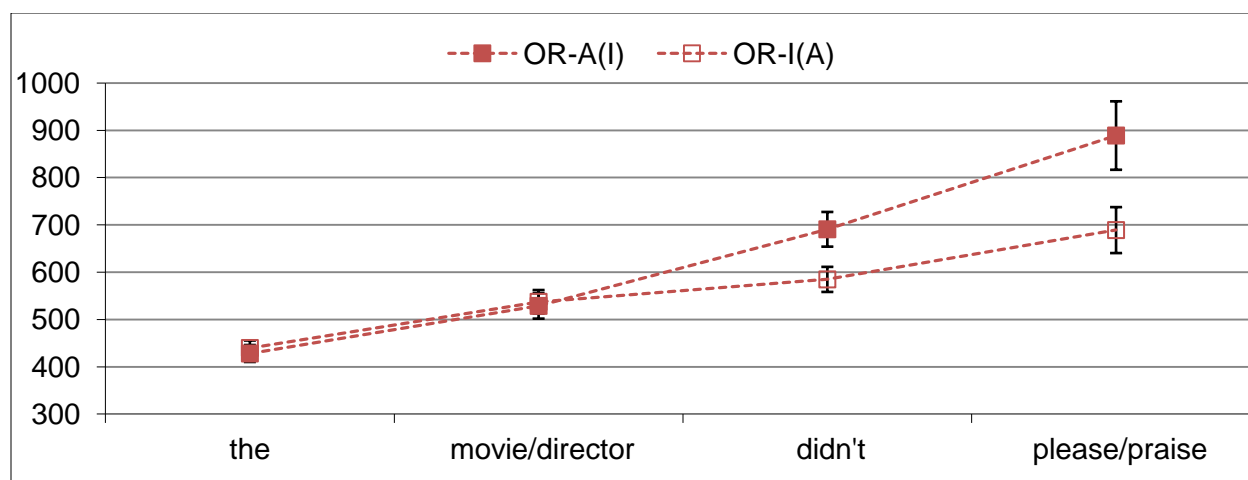


Figure 15 Mean reading times at each word within the ORs in Experiment 2K

In Experiment 2K, the effects of syntactic complexity, noun animacy configuration, and individual WMC differences on L2 sentence processing were investigated. To control for the effect of general proficiency in the L2, cloze test scores were also included in the statistical analysis. Not surprisingly, proficiency in the L2 as measured by the cloze test exerted some influence on sentence processing behaviors such that higher proficiency learners were more accurate in comprehension and generally faster in reading compared to lower proficiency learners. In some cases, only higher proficiency learners showed sensitivity to the experimental manipulations, as was shown in the reading time data at the RC region.

With the effects due to proficiency statistically controlled for, the results suggest that adult L2 learners processed sentences in the L2 in quite a similar way to native speakers, incrementally incorporating both structural and lexical-semantic information as reflected in the immediate online effect of the relative difficulty of reanalysis in the ORs. Evidence for the similarity between the sentence processing mechanisms that are used for L1 and L2 sentence processing also comes from the finding of a similar pattern of association/dissociation between sentence processing performance and individual WMC differences as shown in Experiment 2E and the present experiment.

Like the native speakers tested in Experiment 2E, the L2 learners in Experiment 2K found ORs more difficult to comprehend than SRs only when the RC head nouns were animate and the RC-internal nouns were inanimate. This was true both in online reading and in final comprehension, with higher proficiency or higher WMC learners generally showing greater

sensitivity to the experimental manipulations in the expected directions. The results are consistent with the predictions of the reanalysis-based account, according to which the asymmetrical difficulty with the ORs with animate heads and inanimate subjects is due to the relative difficulty of the reanalysis required for ORs.

Word-by-word comparisons within the ORs with different animacy configurations suggest that adult L2 learners' sensitivity to the relative difficulty of reanalysis in an OR begins to emerge before they process the RC verb that would trigger the lexical integration of the verb and the head noun of the RC. The results suggest that when processing RCs, adult L2 learners attempt to assign grammatical roles to the noun phrases based on 'structural' information such as word order and the presence of a relative pronoun before they encounter the verb of the RC. This finding would not be consistent with a strong version of the shallow structure hypothesis (SSH), which would predict that adult L2 learners primarily rely on lexical integration processes triggered by lexical subcategorizers due to their reduced ability to compute structural configurations between constituents.

WMC in L2 seems to modulate the ability to incorporate different types of information to build a coherent L2 sentential representation, as WMC in L1 does (see Experiment 2E). However, the direction of WMC effects was somewhat different from native speakers. For native speakers, higher WMC participants usually displayed a better ability to resolve conflicting information to reach a correct interpretation, resulting in relatively *smaller* differences between the conditions where a difference in processing difficulty was expected (e.g., between the SRs and ORs in the A(I) condition), compared to lower WMC participants. On the other hand, higher WMC in L2 seems to be associated with greater sensitivity to various types of information available from input during L2 sentence processing, as reflected in generally *larger* differences in reading time between the conditions where the degree of processing difficulty is expected to be different. In other words, the higher WMC learners showed more native-like behaviors in that the pattern of their behaviors fits better with native speakers' behaviors and what is predicted by the reanalysis-based account. This pattern of WMC effects on L2 sentence processing has also been shown in previous studies, in which only L2 learners with high WMC showed native-like sensitivity to semantic and structural information in filler-gap processing in L2 English and L2 Dutch (e.g., Dussias & Piñar, 2010; Havik et al., 2009).

7.3. Experiment 3K

Experiment 3K further tested the role of animacy of the RC-internal noun in determining the processing difficulty involved with ORs by keeping the RC head nouns animate across conditions. The role of WMC in L2 was also further tested. Thirty-one participants contributed data.

7.3.1. Materials and design

The materials and design of Experiment 3K were the same as Experiment 3E. The test sentences are repeated in (76) – (79) for reference. These sentences represent the SR-A(A), OR-A(A), SR-A(I), and OR-A(I) conditions, respectively.

(76) The teacher that applauded the actor was invited to the film festival.	<i>SR-A(A)</i>
(77) The teacher that the actor surprised was invited to the film festival.	<i>OR-A(A)</i>
(78) The teacher that applauded the movie was invited to the film festival.	<i>SR-A(I)</i>
(79) The teacher that the movie surprised was invited to the film festival.	<i>OR-A(I)</i>

The participants in Experiment 3K provided post-hoc plausibility ratings on the propositions derived from the RCs of the test sentences. The main clauses were the same across conditions, so plausibility was not checked for the propositions expressed by them. Table 25 presents the mean plausibility ratings for the propositions from each condition. A 2 (SR vs. OR condition) \times 2 (A(A) vs. A(I) condition) repeated measures ANOVA on the mean plausibility ratings by items aggregated across participants revealed that the propositions derived from the SRs were rated reliably more plausible ($M = 5.92$, $SD = 0.79$) than those derived from the ORs ($M = 5.61$, $SD = 0.92$) ($F(1, 27) = 6.440$, $p < .05$). The main effect of animacy configuration and its interaction with RC type were not significant (both $F_s < 2.034$, both $p_s > .1$). The lower perceived plausibility for the propositions expressed by the ORs might be due to the lexical difficulty associated with the lexical properties of the RC verbs, as was also a possibility in Experiment 2E. Because the lower perceived plausibility of the ORs is confounded with the greater syntactic complexity of the construction so that both characteristics predict greater processing difficulty, the interpretation of syntactic complexity effects should be qualified by the potential plausibility effect. But the lack of the main effect of animacy and the interaction of RC

type and animacy in the ANOVA makes the potential confound of syntactic complexity and lower plausibility irrelevant for the present purposes because the main interest of the present experiment lies in the interaction of syntactic complexity and noun animacy information. Specifically, any difference in the magnitude of SR-OR processing asymmetry between the different animacy configurations cannot be due to OR penalty amplified by lower plausibility because plausibility differences between SRs and ORs should equally affect the two animacy conditions.

Table 25 Mean plausibility ratings on the propositions derived from the RCs in Experiment 3K

Condition	Example of RC proposition	Mean plausibility rating (SD)
SR-A(A)	The teacher applauded the actor.	6.00 (0.56)
OR-A(A)	The actor surprised the teacher.	5.77 (0.59)
SR-A(I)	The teacher applauded the movie.	5.85 (0.98)
OR-A(I)	The movie surprised the teacher.	5.46 (1.16)

7.3.2. Predictions

In Experiment 2K, adult L2 learners showed a facilitation in OR processing when RC head nouns were inanimate and RC subjects were animate, suggesting that animate RC subjects facilitated the reanalysis process involved in the processing of ORs, consistent with the reanalysis-based account of Traxler et al. (2002, 2005). Based on the account, we could predict that if adult L2 learners are as efficient as native speakers in incorporating noun animacy information to carry out the reanalysis involved in the processing of ORs, the facilitation effect of animate RC subjects will appear even when the RC head nouns are animate, as shown by the native speakers in Experiment 3E. If adult L2 learners are not as efficient and flexible as native speakers in updating the current parse by incorporating newly coming information, then the increased plausibility of an initial SR (mis-)analysis due animate RC heads might cause increased processing difficulty when it must be revised to an OR analysis to the extent that the goodness of an animate entity as the subject cannot readily be appreciated to facilitate the reanalysis, resulting in the lack of animate subject advantage in the OR-A(A) condition as compared to the OR-A(I) condition.

Based on the shallow structure hypothesis (SSH), it could also be predicted that if the structural representations constructed by adult L2 learners are shallower and less detailed, and thus syntactic relations between words in a sentence are not as fully specified as in the structural representations built by native speakers, then the interference effect between semantically similar elements could be more pronounced in L2 sentence processing compared to L1 sentence processing because of decreased structural distinctiveness. This would result in the greater processing difficulty in the A(A) conditions than in the A(I) conditions.

7.3.3. Results and discussion

7.3.3.1. *Comprehension accuracy*

Table 26 presents the mean accuracy for each condition. The ORs seem to be harder to comprehend than SRs, especially in the A(I) condition. To assess this pattern, a logistic mixed effects model was fitted. The estimated fixed effects are presented in Table 27.

Table 26 Mean comprehension accuracy by condition in Experiment 3K (in percent)

	A(A)		A(I)	
	SR	OR	SR	OR
M	83.9	79.3	82.9	68.7
(SD)	(14.6)	(16.4)	(14.0)	(21.3)

Table 27 Logistic mixed effects model for the comprehension accuracy in Experiment 3K

	Estimate	SE	t
(Intercept)	1.593	0.184	8.64***
RC type	-0.643	0.199	-3.23**
Animacy	-0.467	0.200	-2.34*
cWM	0.016	0.011	1.43
RC type:Animacy	-0.232	0.398	-0.58
RC type:cWM	0.051	0.018	2.79**
Animacy:cWM	0.008	0.018	0.46
RC:Ani.:cWM	-0.062	0.037	-1.70 [†]
Cloze	0.068	0.025	2.70**
Cloze:RC type	-0.004	0.040	-0.10
Cloze:Animacy	0.092	0.040	2.28*
Cloze:cWM	0.001	0.002	0.84
Cloze:RC:Ani.	0.059	0.081	0.73
Cloze:RC:cWM	0.001	0.003	0.28
Cloze:Ani.:cWM	0.006	0.003	2.27*
Cloze:RC:Ani.:cWM	-0.015	0.006	-2.07*

The model included random intercepts for participants and items.

The main effects of RC type and animacy were reliable. A notable pattern is that the direction of the animacy effect was the opposite to the native speakers in Experiment 3E. Whereas the native speakers in the experiment found the A(A) condition *more difficult* to comprehend than the A(I) condition, the L2 learners found the A(A) condition *easier* to comprehend than the A(I) condition. The interaction of RC type and cWM was also reliable. The 3-way interaction of RC type, animacy and cWM was marginally reliable. But these effects should be interpreted in relation to the effects of cloze test scores. First, the main effect of cloze was reliable, indicating that higher proficiency was associated with higher accuracy in comprehension, which is not surprising. Cloze interacted with several other factors in a complex way, including a significant 4-way interaction of cloze, RC type, animacy, and cWM. To explore the effect of proficiency and to present the data in a more interpretable way, participants were divided into higher (N = 19) and lower proficiency group (N = 12) split at the median cloze score, and separate mixed effects models were fitted to each group's accuracy data. The results are presented in Table 28.

Table 28 Logistic mixed effects models for the comprehension accuracy by proficiency group in Experiment 3K

	High proficiency (N = 19)			Low proficiency (N = 12)		
	Estimate	SE	t	Estimate	SE	t
(Intercept)	1.705	0.208	8.20***	1.342	0.214	60.28***
RC type	-0.553	0.248	-2.23*	-0.616	0.284	-2.17*
Animacy	-0.164	0.248	-0.66	-0.726	0.289	-2.52*
cWM	0.027	0.017	1.62	0.021	0.014	1.51
RC type:Animacy	-0.024	0.496	-0.05	-10142	0.577	-1.98*
RC type:cWM	0.065	0.024	2.18**	0.035	0.022	1.61
Animacy:cWM	0.041	0.024	1.67 [†]	-0.013	0.022	-0.56
RC:Ani.:cWM	-0.106	0.049	-2.19*	0.027	0.044	0.61

Both models included random intercepts for participants and items.

As for the higher proficiency learners, there was a reliable main effect of RC type, indicating that the ORs were harder to comprehend than the SRs. Although the main effect of cWM did not reach significance despite a numerical trend of higher WMC – higher accuracy association, the interaction of RC type and cWM was reliable and the interaction of Animacy and cWM was marginally reliable. These interactions, however, were further subject to a significant 3-way interaction of RC type, animacy, and cWM. To explore this 3-way interaction, the higher proficiency learners were further divided into higher and lower WMC groups and their mean accuracy for each condition is described in Figure 16. The figure suggests that the main source of the 3-way interaction was the generally lower accuracy for ORs than for SRs in the lower WMC learners compared to the higher WMC learners, which was more pronounced in the A(A) condition. This modulation of the magnitude of OR penalty by WMC is not consistent with the results of Experiment 1K, where there was no interaction of RC type and WMC either in final comprehension or in online reading. Keeping the small number of participants in mind, the result could be taken as suggestive evidence for some difference between native speakers and adult L2 learners.

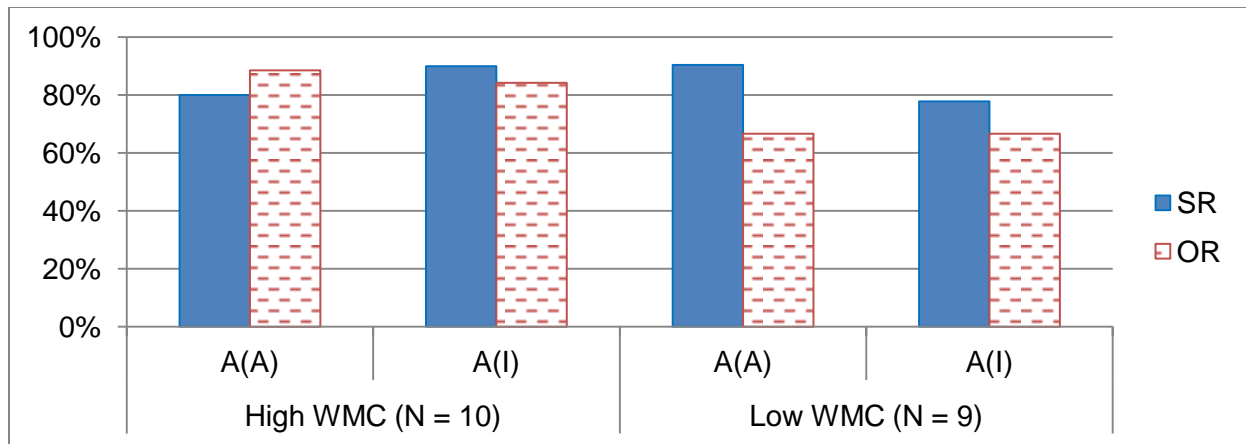


Figure 16 Mean comprehension accuracy in percent by WMC group for the high proficiency L2 learners in Experiment 3K (N = 19)

As for the lower proficiency learners' comprehension accuracy, the main effects of RC type and animacy were reliable as well as their interaction. But the main effect of cWM and the interactions involving it did not reach significance. The pattern of interaction of RC type and animacy is illustrated in Figure 17. The figure suggests that the OR-A(I) condition caused a particular difficulty in comprehension for the lower proficiency learners, which is consistent with the reanalysis-based account that predicts the most severe difficulty of comprehension for ORs with inanimate subjects.

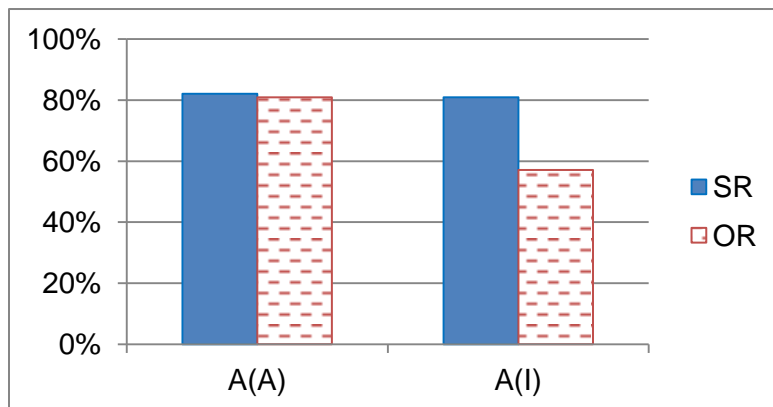


Figure 17 Mean comprehension accuracy in percent for low proficiency L2 learners in Experiment 3K (N = 12)

The overall pattern of results, as illustrated in Figures 16 and 17, suggests that the high WMC participants with higher proficiency did not show an overall comprehension advantage to

A(A) conditions, whereas the other participants did so. Recall that the native English speakers in Experiment 3E showed more accurate comprehension for the A(I) condition than for A(A) condition. The emerging picture is that L2 learners with high WMC and high proficiency seems to be on the way to native-like processing behaviors in that their accuracy pattern across conditions is something in the middle of the native speakers and the L2 learners with lower WMC or lower proficiency.

7.3.3.2. Reading time results

Figure 18 presents the mean reading times at each region by condition. The most salient pattern is the longer reading time for the ORs than for the SRs both in the A(A) and A(I) conditions at the MV region. At the MV+1 region, there was a hint of the trend that the A(A) conditions took longer than the A(I) conditions overall. These patterns were assessed by fitting a mixed effects model at each of the critical regions, and the fitted models are presented in Table 29.

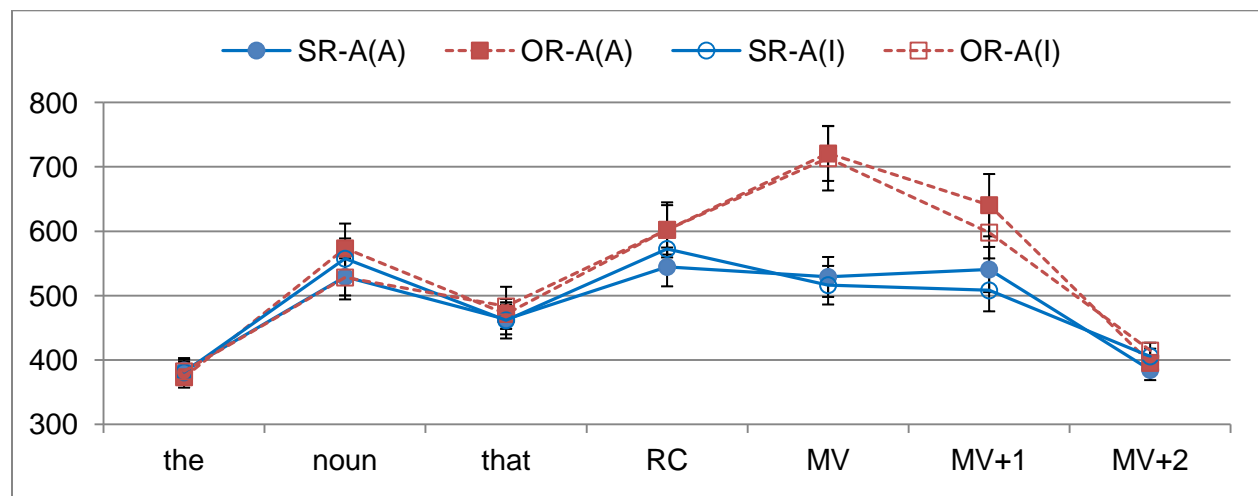


Figure 18 Mean reading times by region and condition in Experiment 3K

Table 29 Linear mixed effects model for the reading times at each critical region in Experiment 3K

Region	RC		MV		MV+1	
Fixed effects	Estimate (SE)	t	Estimate (SE)	t	Estimate (SE)	t
(Intercept)	581.10 (33.91)	17.14***	618.60 (31.20)	19.83***	571.91 (34.06)	16.79***
RC type	41.48 (15.02)	2.76*	193.49 (29.20)	6.63***	94.06 (34.74)	2.71**
Animacy	8.78 (15.07)	0.58	-10.13 (29.26)	-0.35	-37.85 (23.66)	-1.60
cWM	-5.08 (2.83)	-1.80 [†]	-3.00 (2.54)	-1.18	-4.67 (2.69)	-1.73 [†]
RC type:Animacy	-22.49 (30.12)	-0.75	6.55 (58.52)	0.11	-15.08 (47.29)	-0.32
RC type:cWM	-0.77 (1.33)	-0.58	-0.92 (2.56)	-0.36	-1.12 (3.04)	-0.37
Animacy:cWM	-1.21 (1.33)	-0.92	-1.08 (2.55)	-0.42	0.29 (2.07)	0.14
RC:Ani.:cWM	-3.43 (2.66)	-1.29	-0.17 (5.11)	-0.03	-2.76 (4.15)	-0.66

The models for the RC and MV region included random intercepts for participants and items. The model for the MV+1 region included a by-participant random slope for RC type as well as random intercepts for participants and items.

First, cloze test scores did not contribute to any of the models, suggesting that the pattern of reading times was comparable across the participants with different proficiency in English in the present experiment. As Table 29 shows, the only effect that was consistently significant across the 3 regions was the main effect of RC type, indicating that the ORs were read more slowly than the SRs. The main effect of cWM was marginally significant at the RC and the MV+1 region, indicating a trend that the higher WMC learners were generally faster in reading than the lower WMC learners. The trend of *faster* reading by higher WMC participants in the present experiment is in conflict with that of *slower* reading by higher WMC participants in Experiments 1/2K. Although the effect of WMC on general reading speed was not so robust as to be statistically significant, the opposite trends across the experiments seem to suggest that the direction of WMC effects on sentence reading may depend on specific characteristics of experimental stimuli and tasks as well as those of the target population. The other main effects and interactions were not reliable. The lack of any animacy effect is in contrast with the result

from the native speakers in Experiment 3E, where the A(A) conditions were read faster than the A(I) conditions in the OR condition and (for only the higher WMC participants) in the SR condition at the RC region. The result may suggest that L2 learners do not benefit from the animate subject of an OR in the course of reanalysis, unlike native speakers.

Finally, the reading times at each word within the OR conditions were compared between the A(A) and the A(I) condition to further test the role of animacy of RC subjects in processing ORs. As Figure 19 illustrates, the mixed effects models fitted to the reading times at each word with animacy as the only fixed effect did not show any reliable difference due to the animacy of the subject noun (e.g., *actor* vs. *movie*), unlike the native speakers in Experiment 3E, who showed significantly faster reading of the RC verb (e.g., *surprised*) when its subject was animate compared to when it was inanimate. This pattern of reading times within the OR is consistent with the lack of any animacy effect in the main analysis on the RC region.

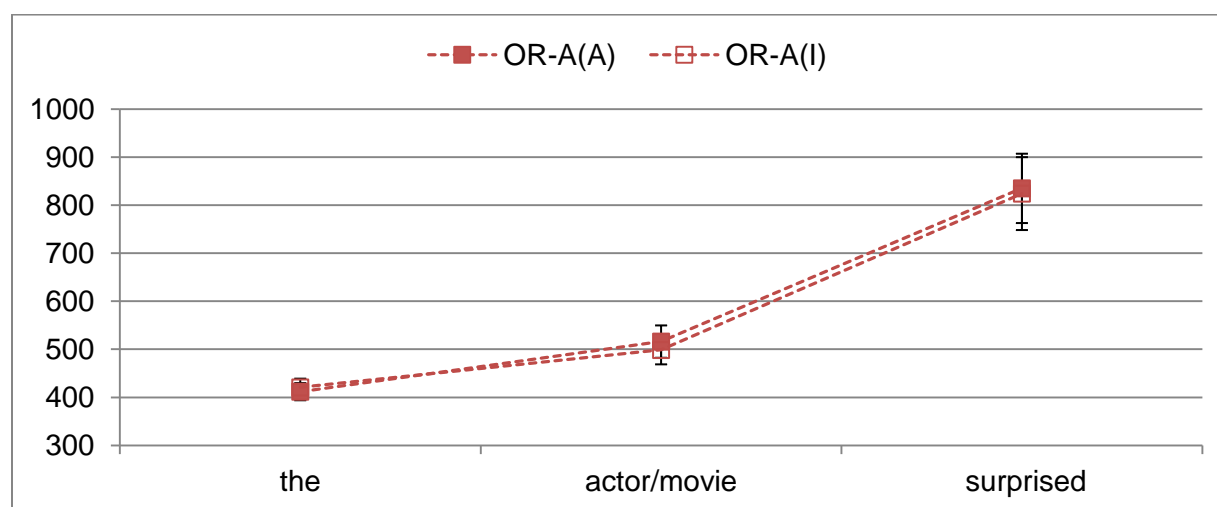


Figure 19 Mean reading times at each word within the ORs in Experiment 3K

To summarize, in Experiment 3K, participants' proficiency in English as measured by the cloze test modulated their performance in sentence comprehension as in Experiments 1K and 2K. But unlike those experiments, where the effect of proficiency was evident both in offline comprehension accuracy and online reading speed, the effect observed in the present experiment was restricted to final comprehension accuracy. The higher proficiency learners were overall more accurate in comprehending the test sentences. Furthermore, the modulating effect of WMC on the interaction of RC type and animacy in final comprehension was observed for the higher

proficiency learners but not for the lower proficiency learners. Given that there was a marginal trend that the interaction of syntactic complexity and noun animacy was further moderated by WMC in final comprehension for the native speakers in Experiment 3E, higher proficiency seems to be associated with more native-like behaviors, as was also the case in the Experiments 1K and 2K.

Now we turn to the effects of main interest for the present purposes. Overall, Experiment 3K showed some evidence for differences between L1 and L2 sentence processing both in terms of the ability to incorporate new information to update the current parse and in terms of the role of WMC in sentence processing. The results are not consistent with those of Experiments 1K and 2K, which showed patterns of L2 learners' behaviors very similar to those of native speakers. In the comprehension accuracy data, the L2 learners showed lower accuracy for RCs with animate head nouns and inanimate RC-internal nouns (e.g., *the teacher that the movie surprised*) compared to those with animate head nouns and RC-internal nouns (e.g., *the teacher that the actor surprised*), as indicated by the reliable main effect of animacy with a negative coefficient. In other words, the RCs with nouns distinct in terms of animacy were associated with greater comprehension difficulty than the RCs with confusable nouns. This direction of animacy effect is the opposite to that shown by the native speakers in Experiment 3E, where participants found RCs harder to comprehend when the two relevant nouns were both animate. These opposite patterns might be related to the difference between the L2 learners and the native speakers in perceived plausibility of the propositions expressed by the RCs. Recall that while the native speakers found the propositions equally plausible (see Table 13), the L2 learners rated those derived from the ORs to be less plausible than those derived from the SRs (see Table 25). Although animacy configuration did not exert a reliable effect on perceived plausibility for L2 learners, the numerical pattern of plausibility ratings across conditions matches that of comprehension accuracy quite well, with the A(I) conditions being lower than the A(A) conditions in plausibility rating and accuracy both in the SR and OR conditions. The result, however, is suggestive at best because it was not supported by statistical analyses and because the animacy effect on comprehension accuracy in the native speakers obtained independently of plausibility. Alternatively, if native speakers' lower accuracy for RCs with confusable nouns is due to interference between the two similar nouns in memory, as suggested in the discussion of Experiment 3E, the results of Experiment 3K might suggest that the interference-based difficulty

can be actually smaller in its strength for L2 learners than for native speakers in certain circumstances such as the present experiment. Then the result would not be consistent with the view of the SSH that L2 learners do not construct as detailed syntactic structures as native speakers during online sentence processing, because such a view would predict that semantic interference effects should be stronger for L2 learners, who would mark the nouns in a sentence less clearly for their syntactic status and thus render the semantically similar nouns less distinct in their syntactic properties.

A notable effect of WMC on final comprehension was observed in the reliable interaction of RC type and cWM shown for the higher proficiency learners, in which the higher WMC learners showed a smaller degree of OR penalty. As Figure 16 shows, this modulation of syntactic complexity effects by WMC was more pronounced for RCs with animate head nouns and RC-internal nouns. This interaction of RC type and WMC is not consistent with the null effect of WMC on syntactic complexity shown in Experiment 1K or with the findings of the experiments with native speakers. However, the interaction of syntactic complexity and WMC as shown in the present experiment is likely to be a spurious effect because it was based on a smaller number of participants and the materials were less strictly controlled in terms of plausibility as compared to Experiment 1K. Even if it is indeed a reliable effect, it would suggest that syntactic complexity effects in L2 sentence processing might be modulated by individual differences in L2 WMC *in final comprehension*, which is still consistent with the separate-resource view of WMC.

In online reading, the ORs with different animacy configurations were equally difficult for the L2 learners unlike for the native speakers, who showed a short-lived processing advantage for the ORs with animate subjects. The result cannot be due to L2 learners' general failure to reanalyze. When the initial analysis to be revised was not plausible, as was the case in the OR-I(A) condition in Experiment 2K, L2 learners could readily take advantage of the helpful semantic information to revise the initial analysis, suggesting that the structural processes underlying the reanalysis were working properly. Therefore, the result of Experiment 3K suggests that when an initially *plausible* analysis must be revised in ORs with animate head nouns, L2 learners may experience severe difficulty to the extent that they cannot benefit from helpful semantic information provided by animate RC subjects. The overall results seem to indicate reduced efficiency or flexibility of the sentence processing mechanisms available to

adult L2 learners, which could ultimately be attributable to reduced exposure to and experience with the target language, rather than a selective deficit in syntactic computation processes.

7.4. General discussion of L2 processing of English RCs

In the present study, three experiments investigated the effects of syntactic complexity, lexical-semantic information, and individual WMC differences in L2 sentence processing with adult L1-Korean L2-English learners. The results showed both similarities and differences between L1 and L2 sentence processing. Overall, L2 sentence processing by adult L2 learners seems to be qualitatively similar to L1 sentence processing in terms of (i) the nature of incremental processes involved in sentence interpretation and (ii) a potential distinction between a set of processing resources that are recruited for incremental assignments of syntactic and thematic structures to an incoming string of words and another set of processing resources that are recruited for later processes responsible for evaluating the overall semantic quality of the resulting interpretation and registering the final interpretation in semantic memory.

In Experiment 1K, the L2 learners showed clear sensitivity to the syntactic complexity as reflected in an OR penalty, not only in decreased comprehension accuracy but also in slower reading speed. The result establishes that L2 learners incrementally construct an interpreted representation of an unfolding sentence, showing immediate sensitivity to syntactic structure. Higher WMC as measured in L2 was associated with higher comprehension accuracy, suggesting that L2 WMC may be a valid predictor of the general reading comprehension ability in the L2 for adult L2 learners. By contrast, WMC was not associated with individual differences in the magnitude of syntactic complexity effects either in comprehension accuracy or in reading speed in the overall analysis. There was, however, a trend of an interaction between WMC and syntactic complexity effects for the lower proficiency learners in reading time, suggesting the possibility that the characteristics of L2 WMC may be modulated by one's current proficiency level in the target language, which will be discussed further in Section 7.4.2.

In Experiment 2K, the L2 learners also showed striking similarities with native speakers in terms of both the sensitivity to different types of information and the modulating effect of WMC on the ability to incorporate information from multiple sources during online sentence processing. For the adult L2 learners, ORs were no more difficult to comprehend than SRs when the head nouns were inanimate and the RC-internal nouns were animate, as was reflected in

increase reading time and decreased comprehension accuracy, whereas the effect of animacy configurations was much smaller for SRs. The results are consistent with the predictions of the reanalysis account. More importantly, the facilitation of reanalysis by animate subjects in ORs was observable before the RC verbs were encountered, as was the case with the native speakers in Experiment 2E. The result suggests that adult L2 learners initiate the reanalysis process when it is required for structural reasons, and that they do not exclusively rely on a lexical subcategorizer to assign appropriate syntactic/thematic roles to noun phrases. WMC seemed to modulate how efficiently an individual incorporated different sources of information to carry out this reanalysis process. Overall, participants with higher WMC showed greater sensitivity to the experimental manipulations, more closely resembling the native speakers in behavioral patterns. The results seem to suggest that WMC may be taken to be a valid indicator of the efficiency of L2 sentence processing.

Experiment 3K highlighted differences between L1 and L2 sentence processing, in contrast with Experiments 1K and 2K. The RCs with animate head nouns and inanimate RC-internal nouns induced generally lower comprehension accuracy than those with animate RC head and RC-internal nouns, with a numerically stronger effect for the ORs. This result is the opposite to the result from the native speakers in Experiment 3E, who showed generally lower comprehension accuracy for RCs with animate head and RC-internal nouns, again with a numerically stronger effect for the ORs. The result seems to suggest that L2 sentence processing can be more sensitive to structural/thematic configurations than L1 sentence processing under some circumstances, if we assume that the difficulty in comprehending ORs with animate head nouns and inanimate RC-internal nouns is due to the non-canonical thematic structure, and the difficulty in comprehending ORs with animate head and RC-internal nouns is mainly due to the interference between the two nouns in memory. This finding may not be consistent with the SSH, which would predict that adult L2 learners should show greater sensitivity to semantic interference because they would not fully specify syntactic properties of sentence constituents. The adult L2 learners in Experiment 3K showed a difference from the native speakers in Experiment 3E in online reading time patterns as well. Although the native speakers showed a facilitation of reanalysis from the animate subject of an OR as reflected in the faster reading of ORs with animate RC-internal nouns at the RC region, the L2 learners showed no evidence for

this animate subject advantage in OR processing. Possible reasons for this apparent insensitivity to helpful semantic information will be discussed further in the next section.

7.4.1. Structure-based incremental sentence processing in L2

One of the main goals of the present dissertation was to test the nature of incremental sentence processing in adult L2 learners. As discussed in Section 2.2, some researchers propose that adult L2 learners' interpretation of L2 sentences primarily relies on lexical mechanisms such as syntactic/thematic dependency between a lexical subcategorizer and its dependents, because their ability to compute syntactic relations among constituents is reduced due to later acquisition of the target language (e.g., Clahsen & Felser, 2006; Felser et al., 2003; Felser et al., in press; Felser & Roberts, 2007; Jackson, 2008; Marinis et al., 2005; Papadopoulou & Clahsen, 2003). This view is represented by the shallow structure hypothesis (SSH), according to which adult L2 learners are largely restricted to the shallow processing routes that interpret a given sentence heuristically without fully computing the syntactic structure of the sentence (cf. Christianson et al., 2001; Ferreira, 2003).

Alternative views to the SSH do not impose a categorical distinction between L1 and L2 sentence processing mechanisms (e.g., Dussias & Saggarra, 2007; Frenck-Mestre, 2002, 2005; Hopp, 2010; Jackson & Dussias, 2009). According to these views, syntactic processes underlying L2 sentence processing are not qualitatively different from those recruited for L1 sentence processing. Apparent differences in (morpho-)syntactic processes between L1 and L2 sentence processing are instead attributed to reduced experience with and exposure to the target language use and the resulting inefficiency in processing, and/or to compensatory transfer of L1 processing routines to L2 processing.

The three experiments conducted in the present dissertation with adult L1-Korean L2-English learners seem to be largely in harmony with the latter views that highlight the similarities between L1 and L2 sentence processing and attribute the differences between the two kinds of sentence processing to the reduced processing efficiency on the part of L2 learners. In Experiments 1K and 2K, the L2 learners showed processing patterns very similar to those of the native speakers in Experiments 1E and 2E. Specifically, they showed immediate sensitivity to syntactic structures as displayed in an online OR penalty in Experiment 1K, and this OR penalty was eliminated by helpful semantic information from the canonical mappings between noun animacy and grammatical roles (i.e., animate subjects and inanimate objects). Crucially, the

effect of noun animacy on the ease of reanalysis began immediately (i.e., similarly to native speakers) at the region where the reanalysis is required for structural reasons before encountering the verb inside an OR. Therefore, the reanalysis effect cannot be attributed to lexical integration processes triggered by a lexical subcategorizer.

In Experiment 3K, however, the facilitation of reanalysis by the animate RC subject of an OR was not observed with the L2 learners, although the effect persisted with the native speakers in Experiment 3E. It seems that the L2 learners in Experiment 3K could not take advantage of the fact that animate entities were good subjects, although they were clearly shown to use such information in Experiment 2K. The crucial difference between the two experiments was that RC heads were kept animate in Experiment 3K, whereas the animacy of RC heads was always contrasted with that of RC-internal nouns in Experiment 2K. Thus, an OR with an animate subject had an inanimate head noun in Experiment 2K but an animate head noun in Experiment 3K. The L2 learners in Experiment 2K may not have strongly committed to the initial SR analysis of an OR because inanimate RC head nouns are not a good subject in the first place, leading to an easier reanalysis when they encountered an animate RC subject. In Experiment 3K, in contrast, an initial SR analysis of an OR was always a plausible option because the head noun was always animate, resulting in increased processing difficulty due to the need to discard a plausible analysis (e.g., Roberts & Felser, 2011; Traxler & Pickering, 1996; Traxler et al., 1998; Williams, 2006; Williams et al., 2001). The inability of adult L2 learners to benefit from animate subjects in Experiment 3K should then be attributed to the increased difficulty of reanalysis. In other words, adult L2 learners may not be as efficient and flexible in discarding the old structure and carrying out a revision process, especially when the old structure to be discarded makes good sense (cf. Christianson et al., 2001; Pickering and Traxler, 1998).

7.4.2. Role of WMC in L2 sentence processing

Another main goal of the present study was to test whether the role of WMC in L2 sentence processing can be used as a diagnostic for the characteristics of the cognitive processes underlying L2 sentence processing, as compared to those underlying L1 sentence processing. This question capitalizes on the finding in the L1 sentence processing literature that at least some sub-processes involved in online sentence processing draw upon a dedicated set of processing resources that are not tapped by WM span tests, which has been taken by some researchers as evidence for the presence of ‘core’ linguistic processes (e.g., Caplan & Waters, 1999; Traxler et

al., 2005). The role of WMC in L2 sentence processing as revealed in the present study seems to point to the possibility that those core linguistic processes are implemented by the L2 sentence processing mechanism as well, though with some differences between L1 and L2 sentence processing in the specific pattern of the relationship between individual differences in WMC and sentence processing behaviors.

In Experiment 1K, individual WMC differences were reliably associated with overall comprehension accuracy but there was little evidence that they predicted individual differences in the magnitude of syntactic complexity effects. In final comprehension, the OR difficulty was not modulated by WMC at all. The same was true for online reading times with L2 learners with higher proficiency. Lower proficiency learners, by contrast, showed the possibility that the efficiency of online syntactic computation depends on one's WMC. But the specific pattern of the effect of WMC on syntactic complexity effects (i.e., the greater OR difficulty for higher WMC learners) differed from what would be predicted by the general assumption that comprehenders with lower WMC will experience *greater* difficulty with complex structures (e.g., King & Just, 1991; Traxler et al., 2005). This finding may indicate that a sufficient level of WMC is required for L2 learners to be able to incrementally build syntactic structures such as RCs. With a certain level of proficiency reached, WMC may no longer predict either the capability of immediate structure-building or the efficiency of online syntactic computation, as suggested by the results from the higher proficiency learners and from the native speakers in the present study. The overall results thus suggest that structural computation in L2 sentence processing by advanced learners may not draw upon the processing resources captured by WM span tests, as is the case with native speakers. An interesting possibility, however, is that L2 learners with relatively low proficiency in the L2 may resort to more general-purpose processing resources for online syntactic computation, which seems to deserve attention in future research.

In Experiments 2K and 3K, where noun animacy was also manipulated, individual differences in WMC were consistently associated with individual differences in the modulation of syntactic complexity effects by noun animacy configurations, in both online processing speed and final comprehension accuracy. The results suggest that semantic evaluation involved in reanalysis processes recruit those cognitive resources that are also drawn up on to perform WM span tests. One thing to note in the results of Experiment 3K is that there was some evidence in final comprehension accuracy that among the higher proficiency participants ($N = 19$), those

with higher WMC ($N = 10$) showed a smaller degree of OR penalty than those with lower WMC ($N = 9$). This might indicate the possibility that syntactic complexity effects are modulated by WMC in L2 sentence processing, in contrast with the results of Experiment 1K and the results of the experiments with native speakers. But the finding would still be consistent with the dedicated-resource view because final comprehension accuracy is hypothesized to reflect the result of primarily ‘post-interpretive’ processes that are tapped by WM span tests (Caplan & Waters, 1999).

Taken together, the overall results of the present experiments with adult L2 learners seem to be generally consistent with Traxler et al.’s (2005) formulation of the dedicated-resource view of WMC, suggesting that adult L2 learners can develop native-like processes for early structural computation that are sub-served by a dedicated pool of processing resources.

Despite the similarities at a general level in the role of WMC between L1 and L2 sentence processing, there were some notable differences in the specific directions of WMC effects. Higher WMC was usually associated with more native-like sensitivity to the experimental manipulations for the L2 learners, resulting in *increased* differences between conditions, whereas it was usually associated with a better ability to cope with difficult situations for the native speakers, resulting in *reduced* differences between conditions. Apparently, the higher WMC L2 learners showed a similar characteristic as the lower WMC native speakers. Recall that the native speakers’ WMC span scores were significantly higher than those of the L2 learners. This gradient effect of WMC between native speakers and L2 learners was also reported in other studies on L2 sentence processing (e.g., Dussias & Piñar, 2010; Havik et al., 2009; cf. Williams, 2006). The result thus seems to suggest that the role of L2 WMC reflects some aspects of the developmental trajectories that L2 learners go through. It is interesting here that at least for the L2 learners tested in the present study, cloze test scores and WM span scores did not correlate at a significant level. In addition, individual differences in sentence processing behaviors that were not captured by the cloze test were associated with individual differences in the WM span tests. This means that L2 WMC may predict some characteristics of L2 learners that are not readily explained by proficiency. Further research into the relationship between proficiency and WMC in L2 would provide valuable insights into the nature of L2 processing and individual differences therein.

Chapter 8

Conclusion and future research

The present dissertation asked whether and how much adult L2 learners' sentence processing in the L2 is similar to or different from normal L1 sentence processing by focusing on how structural and lexical-semantic information is incorporated in the course of online processing of English RCs. As an additional way to compare L1 and L2 sentence processing, the role of WMC was also tested.

The results from the three experiments with native speakers of English and the three parallel experiments with adult L1-Korean L2-English learners suggest that L2 learners process L2 sentences in quite similar ways to native speakers of the language. L2 learners incrementally interpreted sentences in the L2 like native speakers of the language, immediately incorporating both structural and lexical-semantic information available from input to construct the corresponding representations. There was little evidence that they primarily rely on semantic information retrieved from lexical items such as nouns and verbs for assigning syntactic relations to incoming words (cf. Jackson & Roberts, 2010). At the same time, however, there was evidence that when processing complexity increases due to, for example, the need to abandon a plausible interpretation, L2 learners' ability to continuously update the current interpretation degrades, indicating L2 learners' relatively limited processing resources or efficiency as compared to native speakers (e.g., Fesler and Roberts, 2011; Williams, 2006; Williams et al., 2001). The overall results are more consistent with the views that highlight the fundamental similarity between L1 and L2 processing routines such as the experience-based accounts of L2 sentence processing (e.g., Dussias & Sagarra, 2007; Frenck-Mestre, 2002), rather than with the views that focus on the qualitative differences in the syntactic representations and/or processing routines between the two types of sentence processing, such as the shallow structure hypothesis (SSH) (e.g., Clahsen & Fleser, 2006; Fesler et al., in press).

Additional evidence for the qualitative similarity between L1 and L2 sentence processing mechanisms comes from the characteristics of working memory (WM) resources that enable the mechanisms to work. Evidence suggests that there may be a similar kind of dissociation in L1 and L2 sentence processing between the processing resources recruited to assign syntactic/thematic structures to an incoming string of words, and those recruited to evaluate the semantic quality of the resulting interpretation and remember it for later recall, as reflected in

differential relationship with WMC. L2 sentence processing mechanisms thus could be characterized as a ‘language’ system rather than as a general purpose processing system appropriated for linguistic purposes. This finding is analogous to the view of some L2 acquisition researchers that L1 and L2 acquisition is guided by the same principles and thus results in the same kind of ‘linguistic’ knowledge system (e.g., Schwartz, 1992, 1998; Schwartz & Sprouse, 1996; Song & Schwartz, 2009).

In order to corroborate the findings of the present dissertation, different types of syntactic structures and lexical-semantic information need to be investigated with regard to their interaction in the course of L2 sentence processing. As has been discussed in Section 4.2, interpretation of RCs might not involve the construction of ‘hierarchical’ syntactic relations, in which case a central claim about the fundamental similarity between L1 and L2 sentence processing based on RC processing might be undermined to a certain extent. What is crucial would be to find appropriate test structures that unequivocally contain hierarchical configurations of sentence constituents but are not so complex as to overwhelm and break down normal L2 sentence processing routines.

Research needs to be extended to more peripheral or discourse-level semantic and pragmatic information because animacy of nouns is a very salient feature that plays a central role in determining the ways in which thematic roles are syntactically realized across natural languages (e.g., de Swart et al., 2008), which might be the main reason for the apparent similarity between L1 and L2 sentence processing shown in the present study. In Experiment 2 of the Traxler et al. (2002) study with native speaker of English, the processing of subject-extracted RCs (SRs) and object-extracted RCs (ORs) was compared, with the thematic fit of the RC verb with the RC head noun and the RC-internal noun manipulated so that only one of the nouns was plausible as the agent of the action described by the RC verb (e.g., *the policeman that arrested the thief* vs. *the thief that the policeman arrested*). This plausibility manipulation reduced the OR difficulty as compared to when both nouns were plausible agents of the RC verb (e.g., *the lawyer that irritated the banker* vs. *the banker that the lawyer irritated*), suggesting that the plausibility of the RC head noun as the object of the RC verb facilitated the reanalysis involved in the processing of an OR. Importantly, however, the plausibility manipulation did not eliminate the traditional OR difficulty, unlike animacy manipulation as implemented in Experiment 3 of Traxler et al. (2002) and in the present experiments (e.g., *the director that*

watched the movie vs. *the movie that the director watched*). This difference between verb–noun thematic fit and animacy manipulation might be due to the fact that the latter information becomes available *earlier* than the former. Specifically, the animacy of the RC head noun may affect the likelihood of the noun to be an object or a subject in the following RC as soon as the relative pronoun is encountered, whereas the thematic fit between the RC head noun and the RC verb in an OR can be evaluated only *after* the comprehender processes the RC verb. Since the RC head was always animate in the plausibility manipulation experiment of Traxler et al. (2002), the initial SR analysis of an OR should be plausible and thus may have been as hard to revise as in a thematically neutral OR until the semantic information of the RC verb becomes available. The effect of this kind of plausibility manipulation on L2 processing of RCs is worth testing to further explore the similarities and differences between L1 and L2 sentence processing. If adult L2 learners rely on semantic and pragmatic information to a greater extent than on structural information as compared to L1 speakers as proposed by the SSH, then the effect of the plausibility manipulation should be greater on L2 learners, resulting in greater reduction or elimination of the traditional OR difficulty. On the other hand, if adult L2 learners adopt native-like processing routines and initiate immediate reanalysis when encountering the RC subject in an OR as suggested by the results of the present study, then the difficulty involved with the reanalysis would persist until the thematic fit between the RC head and the RC verb can be evaluated. If it is the case, we will see a similar pattern of the effect of plausibility manipulation on L2 processing of RCs to the pattern as shown in Traxler et al. (2002).

The results of the present study also provide a hint of evidence that what L2 WMC reflects with regard to online L2 sentence processing may vary with one's general proficiency in the L2. The L2 learners with lower proficiency, but not those with higher proficiency, showed a trend that syntactic complexity effects are modulated by WMC in Experiment 1K, with only the higher WMC learners showing evidence for online sensitivity to syntactic complexity. As discussed earlier, this may suggest that only with a sufficient level of WMC can L2 learners incrementally build complex structures such as RCs. Alternatively, however, the finding might indicate that L2 sentence processing is initially sub-served by a pool of general-purpose processing resources, and as proficiency increases, it is gradually outsourced to more specialized processing resources that may be the same in nature as those underlying native language ability. Although the small number of participants in the present dissertation does not allow a robust

generalization, analogous ideas have been entertained in the relevant literature. For example, Ullman (2001) hypothesized that grammar rules are initially learned by adult L2 learners using ‘declarative’ memory which is posited to be specialized for representing lexical knowledge in L1, and that with extensive practice and exposure, some of the L2 rules are learned through procedural memory that is hypothesized to be responsible for grammatical computation in L1. In addition, multiple neurolinguistic studies showed that adult L2 learners’ neural responses to the morphosyntactic properties of the L2 became to resemble those of native speakers of the language as proficiency and length of training increase (e.g., Ojima, Nakata, & Kakigi, 2005; Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006; see Steinhauer, White, & Drury for a comprehensive review). In a similar vein, Segalowitz (2003) postulated the development in lexical and grammatical processing in L2 as a process of automatization, which involves the concentration of the brain region that is activated by the relevant cognitive processes as processing efficiency increases.

Further research with adult L2 learners with a greater range of proficiency and WMC is needed into the potential distinction in processing resources responsible for syntactic computation vs. semantic evaluation. Such research would provide a clearer picture of the development of the L2 sentence processing mechanism, which might be characterized by changes in the role of WMC. If it turns out that L2 sentence processing goes through a process of ‘specialization’ in underlying cognitive resources, then a categorical distinction between native-like full processing and non-native shallow processing as postulated in the SSH may not be a viable way to adequately describe L2 sentence processing. Instead, any adequate theory of L2 sentence processing would need to incorporate the developmental components as the experience-based view does. For that matter, the findings of the present dissertation leads to a conclusion that the role of WMC in L2 sentence processing could constitute a dimension of L2 sentence processing models along which a developmental path can be mapped.

The present study also suggests that the two WM span tests administered in the L2 in the present experiments (i.e., reading span and subtract-2 span tests) are quite reliable measures and yield potentially valid indicators of individual differences in the ability to immediately incorporate different kinds of information during the course of L2 sentence processing (e.g., Dussias & Piñar, 2010). However, the number of participants in the present study is rather small for individual differences research, which typically tests much more participants to obtain

reliable data (e.g., Kane et al., 2004; Water & Caplan, 2003, 2004). Thus, larger-scale studies on the psychometric characteristics of L2 WM span tests and their predictive forces with regard to L2 sentence processing are needed for a better understanding of the nature of processing resources underlying L2 sentence processing.

Lastly, one's WM span would much depend on how familiar the individual is with the stimuli to be processed in a specific WM test, as briefly discussed in Section 3.3. More familiar stimuli will be processed more efficiently, allowing a greater portion of available WM resources to be secured for storing to-be-recalled items in a WM span test. For this reason, any verbally-mediated WM span as measured by, for example, the reading span and minus-2 span tests would inevitably reflect how familiar/efficient an individual is with the language used for the tests. This raises the possibility of confound between L2 WMC and L2 proficiency/experience, undermining the status of L2 WMC as an independent predictor for individual differences in L2 sentence processing. Although the result of the present study suggests that L2 WMC may tap different aspects of L2 learners' individual differences from what L2 proficiency reflects, the small number of participants does not warrant a determinate conclusion. Here again, a larger-scale study will address this limitation of the present study. An alternative way to obviate the potential confound between L2 WMC and L2 proficiency would be to use language-independent, visuospatial WM span tests as individual-differences measures, which are taken to reflect the same construct as tapped by verbal WM span tests by some researchers (e.g., Engle, 2010; Kane et al., 2004). The role of language-independent WMC in L1 and L2 sentence processing seems to be a promising topic of future research as a way to further explore the similarities and differences between the two types of sentence processing and to explain individual differences in L2 processing and acquisition.

Appendix A Test sentences

Experimental sentences for Experiment 1 to 3 are presented. The subject-and object-extracted versions of each test sentence and the corresponding verification statements are presented together demarcated by “/”. The verification statements begin with “?”, which was not shown to the participants. At the end of each sentence is the correct answer to the preceding verification statements.

Experiment 1

The banker that irritated the lawyer/the lawyer irritated played tennis every Saturday. ? The banker irritated the lawyer./ The lawyer irritated the banker. True

The lawyer that irritated the banker/the banker irritated played tennis every Saturday. ? The lawyer irritated the banker./The banker irritated the lawyer. True

The child that chased the babysitter/the babysitter chased fell on the floor. ? The child chased the babysitter./The babysitter chased the child. True

The babysitter that chased the child/the child chased fell on the floor. ? The babysitter chased the child./The child chased the babysitter. True

The doctor that complimented the nurse/the nurse complimented asked for a date. ? The doctor complimented the nurse./The nurse complimented the doctor. True

The nurse that complimented the doctor/the doctor complimented asked for a date. ? The nurse complimented the doctor./The doctor complimented the nurse. True

The sailor that hated the captain/the captain hated signed a new contract. ? The sailor hated the captain./The captain hated the sailor. True

The captain that hated the sailor/the sailor hated signed a new contract. ? The captain hated the sailor./The sailor hated the captain. True

The businessman that married the secretary/the secretary married invited the singer to the party. ? "The businessman invited the singer. True

The secretary that married the businessman/the businessman married invited the singer to the party. ? The secretary invited the singer. True

The musician that ignored the painter/the painter ignored drove a red convertible. ? The musician ignored the painter./The painter ignored the musician. True

The painter that ignored the musician/the musician ignored drove a red convertible. ? The painter ignored the musician./The musician ignored the painter. True

The mechanic that divorced the waitress/the waitress divorced cheated on her often. ? The mechanic cheated on the waitress. True

The waitress that divorced the mechanic/the mechanic divorced cheated on him often. ? The waitress cheated on the mechanic. True

The writer that loved the photographer/the photographer loved worked for a national magazine. ?
The writer loved the photographer./The photographer loved the writer. True
The photographer that loved the writer/the writer loved worked for a national magazine. ? The
photographer loved the writer./The writer loved the photographer. True

The burglar that scared the policeman/the policeman scared appeared on the newspaper. ? The
burglar appeared on the newspaper. True
The policeman that scared the burglar/the burglar scared appeared on the newspaper. ? The
policeman appeared on the newspaper. True

The editor that angered the writer/the writer angered fired the entire staff. ? The editor fired the
entire staff. True
The writer that angered the editor/the editor angered wrote an article in protest. ? The writer
wrote an article in protest. True

The prisoner that attacked the guard/the guard attacked got seriously injured during the riot. ?
The prisoner got injured during the riot. True
The guard that attacked the prisoner/the prisoner attacked got seriously injured during the riot. ?
The guard got injured during the riot. True

The tourist that passed the fisherman/the fisherman passed disappeared from the sight. ? The
tourist disappeared from the sight. True
The fisherman that passed the tourist/the tourist passed disappeared from the sight. ? The
fisherman disappeared from the sight. True

The director that admired the dancer/the dancer admired gave her the leading role. ? The dancer
admired the director./The director admired the dancer. Not true
The dancer that admired the director/the director admired worked at a school. ? The director
admired the dancer./The dancer admired the director. Not true

The tenant that hated the landlord/the landlord hated called the lawyer to complain. ? The
landlord hated the tenant./The tenant hated the landlord. Not true
The landlord that hated the tenant/the tenant hated called the lawyer to complain. ? The tenant
hated the landlord./The landlord hated the tenant. Not true

The soldier that assisted the civilian/the civilian assisted died near the end of the war. ? The
civilian assisted the soldier./The soldier assisted the civilian. Not true
The civilian that assisted the soldier/the soldier assisted died near the end of the war. ? The
soldier assisted the civilian./The civilian assisted the soldier. Not true

The cowboy that killed the sheriff/the sheriff killed wore leather trousers. ? The sheriff killed the
cowboy./The cowboy killed the sheriff. Not true
The sheriff that killed the cowboy/the cowboy killed wore leather trousers. ? The cowboy killed
the sheriff./The sheriff killed the cowboy. Not true

The professor that criticized the student/the student criticized blushed and turned away. ? The student criticized the professor./The professor criticized the student. Not true
The student that criticized the professor/the professor criticized blushed and turned away. ? The professor criticized the student./The student criticized the professor. Not true

The carpenter that helped the electrician/the electrician helped retired after an accident. ? The electrician helped the carpenter./The carpenter helped the electrician. Not true
The electrician that helped the carpenter/the carpenter helped retired after an accident. ? The carpenter helped the electrician./The electrician helped the carpenter. Not true

The client that blamed the psychologist/the psychologist blamed felt bad and apologized. ? The psychologist apologized. Not true
The psychologist that blamed the client/the client blamed felt bad and apologized. ? The client apologized. Not true

The player that liked the coach/the coach liked bought a birthday present. ? The coach bought a birthday present. Not true
The coach that liked the player/the player liked bought a birthday present. ? The player bought a birthday present. Not true

The hunter that saw the policeman/the policeman saw ran off into the forest. ? The policeman ran off into the forest. Not true
The policeman that saw the hunter/the hunter saw fired a warning shot. ? The hunter fired a warning shot. Not true

The historian that criticized the freshman/the freshman criticized felt really bad afterward. ? The freshman felt bad afterward. Not true
The freshman that criticized the historian/the historian criticized felt really bad afterward. ? The historian felt bad afterward. Not true

The actor that visited the director/the director visited demanded a leading role. ? The director demanded a leading role. Not true
The director that visited the actor/the actor visited made a few changes in the script. ? The actor made a few changes in the script. Not true

The wolf that approached the deer/the deer approached attacked the helpless animal. ? The deer attacked the helpless animal. Not true
The deer that approached the wolf/the wolf approached ran away across the meadow. ? The wolf ran away across the meadow. Not true

Experiment 2

The driver that hardly witnessed the accident/the accident hardly frightened was investigated by the police. ? The police investigated the driver. True
The accident that hardly frightened the driver/the driver hardly witnessed was investigated by the police. ? The police investigated the accident. True

The singer that instantly composed the song/the song instantly delighted reminded Mary of her mother. ? The song reminded Mary of her mother. Not true

The song that instantly delighted the singer/the singer instantly composed reminded Mary of her mother. ? The singer reminded Mary of her mother. Not true

The policeman that didn't notice the robbery/the robbery didn't surprise was investigated by the detective. ? The policeman was investigated by the detective. True

The robbery that didn't surprise the policeman/the policeman didn't notice was investigated by the detective. ? The robbery was investigated by the detective. True

The man that severely feared the gun/the gun severely injured was known to be unreliable. ? The gun was known to be unreliable. Not true

The gun that severely injured the man/the man severely feared was known to be unreliable. ? The man was known to be unreliable. Not true

The farmer that recently repaired the tractor/the tractor recently wounded arrived at the store last night. ? The farmer arrived at the store last night. True

The tractor that recently wounded the farmer/the farmer recently repaired arrived at the store last night. ? The tractor arrived at the store last night. True

The child that always misplaced the knife/the knife always scared worried the teenage babysitter. ? The babysitter always misplaced the knife. Not true

The knife that always scared the child/the child always misplaced worried the teenage babysitter. ? The babysitter always scared the child. Not true

The director that didn't praise the movie/the movie didn't please received a prize at the film festival. ? The director received a prize. True

The movie that didn't please the director/the director didn't praise received a prize at the film festival. ? The movie received a prize. True

The teacher that never recommended the documentary/the documentary never surprised upset a few students. ? The students never recommended the documentary./The students never surprised the teacher. Not true

The documentary that never surprised the teacher/the teacher never recommended upset a few students. ? The students never surprised the teacher./The students never recommended the documentary. Not true

The buyer that greatly complimented the product/the product greatly excited was mentioned in the newsletter. ? The newsletter mentioned the buyer. True

The product that greatly excited the buyer/the buyer greatly praised was mentioned in the newsletter. ? The newsletter mentioned the product. True

The plumber that slightly damaged the wrench/the wrench slightly injured disappeared in the afternoon. ? The wrench disappeared in the afternoon. Not true

The wrench that slightly injured the plumber/the plumber slightly damaged disappeared in the afternoon. ? The plumber disappeared in the afternoon. Not true

The banker that didn't reject the loan/the loan didn't worry created a problem for the mayor. ?

The banker created a problem for the mayor. True

The loan that didn't worry the banker/the banker didn't reject created a problem for the mayor. ?

The loan created a problem for the mayor. True

The fireman that severely fought the fire/the fire severely wounded appeared in a magazine later. ? The fire appeared in a magazine. Not true

The fire that severely wounded the fireman/the fireman severely fought appeared in a magazine later. ? The fireman appeared in a magazine. Not true

The professor that immediately erased the notes/the notes immediately annoyed caused a big problem. ? The professor caused a big problem. True

The notes that immediately annoyed the professor/the professor immediately erased caused a big problem. ? The notes caused a big problem. True

The gardener that immediately arranged the flowers/the flowers immediately pleased impressed the lady a lot. ? The flowers impressed the lady a lot. Not true

The flowers that immediately pleased the gardener/the gardener immediately arranged impressed the lady a lot. ? The gardener impressed the lady a lot. Not true

The student that diligently attended the school/the school diligently educated was visited by the governor. ? The governor visited the student. True

The school that diligently educated the student/the student diligently attended was visited by the governor. ? The governor visited the school. True

The pilot that always feared the airplane/ the airplane always scared appeared in the newspaper. ?

The airplane appeared on the newspaper. Not true

The airplane that always scared the pilot/ the pilot always feared appeared in the newspaper. ?

The pilot appeared on the newspaper. Not true

The musician that never practiced the instrument/the instrument never impressed had been missing for a month. ? The musician had been missing for a month. True

The instrument that never impressed the musician/the musician never practiced had been missing for a month. ? The instrument had been missing for a month. True

The actor that always rehearsed the play/the play always excited was given first prize. ? The play was given first prize. Not true

The play that always excited the actor/the actor always rehearsed was given first prize. ? The actor was given first prize. Not true

The zebra that frequently drank the water/the water frequently satisfied was located in the back of the cage. ? The zebra was located in the back of the cage. True

The water that frequently satisfied the zebra/the zebra frequently drank was located in the back of the cage. ? The water was located in the back of the cage. True

The soldier that immediately erased the message/the message immediately alarmed had come from the headquarters. ? The message had come from the headquarters. Not true
The message that immediately alarmed the soldier/the soldier immediately erased had come from the headquarters. ? The soldier had come from the headquarters. Not true

The cop that strongly criticized the trial/the trial strongly offended was covered by the national media.
The trial that strongly offended the cop/the cop strongly criticized was covered by the national media.

N.B. Responses to verification statements for the above set were not analyzed due to an error in presentation.

The reporter that cautiously edited the article/the article cautiously criticized caused a big sensation. ? The reporter caused a big sensation. True
The article that cautiously criticized the reporter/the reporter cautiously edited caused a big sensation. ? The article caused a big sensation. True

The cook that really disliked the cheese/the cheese really surprised came from a foreign country. ?
The cheese came from a foreign country. Not true
The cheese that really surprised the cook/the cook really disliked came from a foreign country. ?
The cook came from a foreign country. Not true

The golfer that didn't complete the game/the game didn't satisfy was ignored by most sports writers. ? The game was ignored by most sports writers. Not true
The game that didn't satisfy the golfer/the golfer didn't complete was ignored by most sports writers. ? The golfer was ignored by most sports writers. Not true

The clerk that absolutely misplaced the jewelry/the jewelry absolutely fascinated bothered the old manager. ? The clerk bothered the old manager. True
The jewelry that absolutely fascinated the clerk/the clerk absolutely misplaced bothered the old manager. ? The jewelry bothered the old manager. True

The lady that strongly desired the dessert/the dessert strongly tempted was not very healthy. ?
The dessert was not very healthy. Not true
The dessert that strongly tempted the lady/the lady strongly desired was not very healthy. ? The lady was not very healthy. Not true

The fish that immediately attacked the bait/the bait immediately attracted impressed the fisherman. ? The fish attacked the fisherman./The bait attracted the fisherman. Not true
The bait that immediately attracted the fish/the fish immediately attacked impressed the fisherman. ? The bait attracted the fisherman./The fish attacked the fisherman. Not true

The scientist that constantly observed the volcano/the volcano constantly worried didn't interest the audience. ? The scientist constantly observed the volcano./The volcano constantly worried the scientist. True

The volcano that constantly worried the scientist/the scientist constantly observed didn't interest the audience. ? The volcano constantly worried the scientist./The scientist constantly observed the volcano. True

Experiment 3

The teacher that applauded the actor/the actor surprised was invited to the film festival. ? The actor was invited to the film festival. Not true

The teacher that applauded the movie/the movie surprised was invited to the film festival. ? The movie was invited to the film festival. Not true

The businessman that trusted the lawyer/the lawyer disappointed was surprised at the news. ? The businessman was surprised at the news. True

The businessman that trusted the contract/the contract disappointed was surprised at the news. ? The businessman was surprised at the news. True

The doctor that ignored the patient/the patient upset was notorious for his temper. ? The doctor was notorious for his temper. True

The doctor that ignored the regulation/the regulation upset was notorious for his temper. ? The doctor was notorious for his temper. True

The teenager that admired the waitress/the waitress excited was the student of a nearby middle school. ? The teenager admired the waitress/The waitress excited the teenager. True

The teenager that admired the bicycle/the bicycle excited was the student of a nearby middle school. ? The teenager admired the bicycle./The bicycle excited the teenager. True

The writer that praised the pianist/the pianist impressed was well-known among musicians. ? The pianist was well-known among musicians. Not true

The writer that praised the musical/the musical impressed was well-known among musicians. ? The musical was well-known among musicians. Not true

The bartender that feared the policeman/the policeman scared was going to quit the job. ? The bartender was going to quit the job. True

The bartender that feared the building/the building scared was going to quit the job. ? The bartender was going to quit the job. True

The reporter that hated the editor/the editor bothered didn't want to wait long. ? The reporter hated the editor./The editor bothered the reporter. True

The reporter that hated the meeting/the meeting bothered didn't want to wait long. ? The reporter hated the meeting./The meeting bothered the reporter. True

The director that evaluated the dancer/the dancer pleased was invited to the award ceremony. ?
The dancer was invited to the ceremony. Not true

The director that evaluated the opera/the opera pleased was invited to the award ceremony. ? The
opera was invited to the ceremony. Not true

The woman that feared the king/the king scared couldn't move at all for a while. ? The woman
couldn't move at all. True

The woman that feared the sea/the sea scared could not move at all for a while. ? The woman
couldn't move at all. True

The professor that criticized the student/the student worried was reluctant to help the research
team. ? The professor criticized the student./The student worried the professor. True

The professor that criticized the project/the project worried was reluctant to help the research
team. ? The professor criticized the project./The project worried the professor. True

The scientist that hated the assistant/the assistant bothered was crucial for completing the
project. ? The assistant was crucial for completing the project. Not true

The scientist that hated the climate/the climate bothered was crucial for completing the project. ?
The climate was crucial for completing the project. Not true

The senator that attacked the reporter/the reporter offended didn't want a press conference. ? The
senator attacked the reporter./The reporter offended the senator. True

The senator that attacked the proposal/the proposal offended didn't want a press conference. ?
The senator attacked the proposal./The proposal offended the senator. True

The driver that ignored the passenger/the passenger confused was the oldest in the company. ?
The driver was the oldest in the company. True

The driver that ignored the signal/the signal confused was the oldest in the company. ? The
driver was the oldest in the company. True

The detective that disbelieved the lady/the lady surprised didn't say a word. ? The detective
disbelieved the lady./The lady surprised the detective. True

The detective that disbelieved the evidence/the evidence surprised didn't say a word. ? The
detective disbelieved the evidence./The evidence surprised the detective. True

The student that disliked the teacher/the teacher annoyed was planning to move to another
school. ? The student disliked the teacher./The teacher annoyed the student. True

The student that disliked the festival/the festival annoyed was planning to move to another
school. ? The student disliked the festival./The festival annoyed the student. True

The engineer that distrusted the consultant/the consultant annoyed was assigned to a new team. ?
The consultant was assigned to a new team. Not true

The engineer that distrusted the machine/the machine annoyed was assigned to a new team. ?
The machine was assigned to a new team. Not true

The janitor that noticed the supervisor/the supervisor alarmed was walking down the stairs. ? The janitor was walking down the stairs./The janitor was walking down the stairs. True
The janitor that noticed the danger/the danger alarmed was walking down the stairs. ? The janitor was walking down the stairs. True

The lady that admired the captain/the captain fascinated was too old for a long trip. ? The captain was too old for a long trip. Not true
The lady that admired the yacht/the yacht fascinated was too old for a long trip. ? The yacht was too old for a long trip. Not true

The pianist that disliked the producer/the producer irritated was going to cancel her concert. ? The pianist disliked the producer./The producer irritated the pianist. True
The pianist that disliked the stage/the stage irritated was going to cancel her concert. ? The pianist disliked the stage./The stage irritated the pianist. True

The poet that praised the painter/the painter inspired was willing to write a letter to the curator. ? The poet was willing to write a letter. True
The poet that praised the painting/the painting inspired was willing to write a letter to the curator. ? The poet was willing to write a letter. True

The lady that loved the child/the child amused was planning to buy a piano. ? The lady loved the child./The child amused the lady. True
The lady that loved the music/the music amused was planning to buy a piano. ? The lady loved the music./The music amused the lady. True

The player that blamed the coach/the coach irritated was interviewed by the newspaper. ? The player blamed the coach./The coach irritated the player. True
The player that blamed the weather/the weather irritated was interviewed by the newspaper. ? The player blamed the weather./The weather irritated the player. True

The clerk that noticed the robber/the robber terrified was about to close the shop. ? The clerk noticed the robber./The robber terrified the clerk. True
The clerk that noticed the robbery/the robbery terrified was about to close the shop. ? The clerk noticed the robbery./The robbery terrified the clerk. True

The defendant that criticized the judge/the judge offended was covered by the newspaper. ? The judge was covered by the newspaper. Not true
The defendant that criticized the trial/the trial offended was covered by the newspaper. ? The trial was covered by the newspaper. Not true

The manager that distrusted the banker/the banker worried was reluctant to accept the offer. ? The manager distrusted the banker./The banker worried the manager. True
The manager that distrusted the bank/the bank worried was reluctant to accept the offer. ? The manager distrusted the bank./The bank worried the manager. True

The chef that recommended the waiter/the waiter pleased was famous for his recipes. ? The chef recommended the waiter./The waiter pleased the chef. True

The chef that recommended the cheese/the cheese pleased was famous for his recipes. ? The chef recommended the cheese./The cheese pleased the chef. True

The lawyer that misunderstood the client/the client confused was sued by the opponents. ? The lawyer misunderstood the client./The client confused the lawyer. True

The lawyer that misunderstood the document/the document confused was sued by the opponents. ? The lawyer misunderstood the document./The document confused the lawyer. True

The woman that witnessed the suspect/the suspect frightened was going to report to the police. ? The woman witnessed the suspect./The suspect frightened the woman. True

The woman that witnessed the accident/the accident frightened was going to report to the police. ? The woman witnessed the accident./The accident frightened the woman. True

Appendix B Cloze test

DIRECTIONS

1. Read the passage quickly to get the general meaning.
2. Write only one word in each blank next to the item number. Contractions (example: don't) and possessives (John's bicycle) are one word.
3. Check your answers.

NOTE: Spelling will not count against you as long as the scorer can read the word.

EXAMPLE: The boy walked up the street. He stepped on a piece of ice.

He fell (1) down but he didn't hurt himself.

MAN AND HIS PROGRESS

Man is the only living creature that can make and use tools. He is the most teachable of living beings, earning the name of Homo sapiens. His ever restless brain has used the Knowledge and the wisdom of his ancestors to improve his way of life. Since man is able to walk and run on his feet, his hands have always been free to carry and to use tools. Man's hands have served him well during his life on earth. His development, which can be divided into three major periods, is marked by several different ways of life.

Up to 10,000 years ago, (1) _____ human beings lived by hunting and (2) _____. They also picked berries and fruits, (3) _____ dug for various edible roots. Most (4) _____, the men were the hunters, and (5) _____ women acted as food gatherers. Since (6) _____ women were busy with the children, (7) _____ men handled the tools. In a (8) _____ hand, a dead branch became a (9) _____ to knock down fruit or to (10) _____ for tasty roots. Sometimes, an animal (11) _____ served as a club, and a (12) _____ piece of stone, fitting comfortably into (13) _____ hand, could be used to break (14) _____ or to throw at an animal. (15) _____ stone was chipped against another until (16) _____ had a sharp edge. The primitive (17) _____ who first thought of putting a (18) _____ stone at the end of a (19) _____ made a brilliant discovery: he (20) _____ joined two things to make a (21) _____ useful tool, the spear. Flint, found (22) _____ many rocks, became a common cutting (23) _____ in the Paleolithic period of man's (24) _____. Since no wood or bone tools (25) _____ survived, we know of this man (26) _____ his stone implements, with which he (27) _____ kill animals, cut up the meat, (28) _____ scrape the skins, as well as (29) _____ pictures on the walls of the (30) _____ where he lived during the winter.

(31) _____ the warmer seasons, man wandered on (32) _____ steppes of Europe without a fixed (33) _____, always foraging for food. Perhaps the (34) _____ carried nuts and berries in shells (35) _____ skins or even in light, woven (36) _____. Wherever they camped, the primitive people (37) _____ fires by striking flint for sparks (38) _____ using dried seeds, moss, and rotten (39) _____ for tinder. With fires that he kindled himself, man could (40) _____ wild animals away and could cook those that he killed, as well as provide warmth and light for himself.

References

- Alptekin, C., & Erçtin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading*, 33, 206–219.
- Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30, 191–238.
- Baayen, R.H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Baek, S. (2010). Where to reactivate? L2 processing of filler–gap dependency. In *Proceedings of the 34th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Proceedings Project.
- Bates, D., Maechler, M., & Bolker B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-41. <http://CRAN.R-project.org/package=lme4>.
- Betancort, M., Carreiras, M., & Sturt, P. (2009). The processing of subject and object relative clauses in Spanish: An eye-tracking study. *The Quarterly Journal of Experimental Psychology*, 62, 1915-1929.
- Bever, T., & McElree, B. (1988). Empty categories access their antecedents during comprehension. *Linguistic Inquiry* 19, 35–43.
- Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41–68). New York : Cambridge University Press.
- Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition*, 31, 175–198.
- 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.
- Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *The Modern Language Journal*, 64, 311–317.
- Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 22, 77–94.
- Carreiras, M., Duñabeitia, J. A., Vergara, M., de la Cruz-Pavía, I., & Laka, I. (2009). Subject relative clauses are not universally easier to process: Evidence from Basque. *Cognition*, 115, 79–92.
- Carroll, S. E. (1999). Putting ‘input’ in its proper place. *Second Language Research*, 15, 337–388.
- Chomsky, N. (1977). On WH movement. In P. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal Syntax* (pp. 71–132). New York: Academic Press.
- Christianson, K., Hollingworth, A. Halliwell, J. and Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology* 42. 368–407.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, 27, 3–42.
- Clifton, C., & Frazier, L. (1989). Comprehending sentences with long–distance dependencies. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 273–317). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Clifton, C., Traxler, M. J., Mohamed, M., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The Use of Thematic Role Information in Parsing: Syntactic Processing Autonomy Revisited. *Journal of Memory and Language*, 49, 317–334.

- Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O. & Engle, R.W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769–786.
- Cuetos, F. & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, 30, 73–105.
- Dallas, A., & Kaan, E. (2008). Second language processing of filler-gap dependencies by late learners. *Language and Linguistics Compass*, 2/3, 372–388.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in comprehending and producing words in context. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Daneman, M. & Merikle, P.M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, 3, 422–433.
- de Swart, P., Lamers, M., & Lestrade, S. (2008). Animacy, argument structure, and argument encoding. *Lingua*, 118, 131–140.
- Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition*, 25, 529–557.
- 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. & Piñar, P. (2009). Sentence parsing in L2 learners: Linguistic and experience-based factors. In W. Ritchie & T. Bathia (Eds.), *The New Handbook of Second Language Acquisition* (pp.). Amsterdam, The Netherlands: Elsevier.
- Dussias, P. E. & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis of *wh*-gaps by Chinese-English L2 speakers. *Second Language Research*, 26, 443–472.
- Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English L2 speakers. *Bilingualism: Language and Cognition*, 10, 101–116.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67, 547–619.
- Elman, J. L. (1991). Distributed Representations, Simple Recurrent Networks, and Grammatical Structure. *Machine Learning*, 7, 195–225
- Engle, R.W. (2010). Role of working-memory capacity in cognitive control. *Current Anthropology*, 51, S17–S26.
- Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, 54, 541–553
- Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (in press). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*.
- Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition: A Journal of Developmental Linguistics*, 11, 127–163.
- 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. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, 32, 299–331.
- 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.

- Fernández, E. M. (2002). Relative clause attachment in bilinguals and monolinguals. In R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 187–215). New York: Elsevier.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47, 164–203.
- Ferreira, F., & Clifton, C. Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348–368.
- Ferreira, F., & Patson, N. D. (2007). The ‘good-enough’ approach to language comprehension. *Language and Linguistics Compass*, 1, 71–83.
- Ford, M. (1983). A method of obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 203–218.
- Fodor, J. A. (1985). The précis of the modularity of mind. *The Behavioral and Brain Sciences*, 8, 1–42.
- Fodor, J. D. (2002, April). Psycholinguistics cannot escape prosody. In *Proceedings of the Speech Prosody 2002 Conference*. Available from <http://aune.lpl.univ-aix.fr/sp2002/papers.htm>
- Forster, I., Guerrero, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*, 41, 163–171.
- Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- Frazier, L. (1979). On comprehending sentences: Syntactic parsing strategies. Ph.D. dissertation, University of Connecticut. Indiana University Linguistics Club.
- Frazier, L. (1987a). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp.559-586). Hillsdale, NJ: Erlbaum.
- Frazier, L., (1987b). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, 5, 519–559.
- Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4, 93–126.
- Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6, 291–326.
- Frazier, L., & Rayner, K. (1982). The independence of syntactic parsing. *Journal of Memory and Language*, 25, 348–368.
- French-Mestre, C. (2002). An on-line look at sentence processing in the second language. In R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 217–236). New York: Elsevier.
- French-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *Quarterly Journal of Experimental Psychology*, 50A, 119–148.
- Gathercole, S. E. (2008). Working memory. In J. H. Byrne (Series Ed.) & H. L. Roediger III (Vol. Ed.), *Learning and memory: A comprehensive reference*. Vol. 2: Cognitive psychology of memory (pp. 33–51). Amsterdam: Elsevier.
- Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contribution of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58–93.
- Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, 58, 161–187.

- Gennari, S. P., & MacDonal, M. C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, 111, 1–23.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Gibson, E. (2000). The dependency locality theory: A distance–based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O’Neil (Eds.), *Image, language, brain* (pp. 95–125). Cambridge, MA: MIT Press.
- Gibson, E., Desmet, T., Grodner, D., Watson, D., & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics*, 16, 313–353.
- Gibson, E., & Pearlmutter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences*, 2, 262–268.
- Gibson, E., Schütze, C., & Salomon, A. (1996). The relationship between the frequency and the processing of linguistic structure. *Journal of Psycholinguistic Research*, 25, 59–92.
- Gibson, E., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in longdistance dependencies. *Syntax*, 7, 55–78.
- Gibson, E., & Wu, H.-H. I. (in press). Processing Chinese relative clauses in context. *Language and Cognitive Processes*.
- Gilboy, E., Sopena, J., Clifton, C., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English compound NPs. *Cognition*, 54, 131–167.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1411–1423.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, 51, 97–114.
- Gorrell, P. (1993). Evaluating the direct association hypothesis: A reply to Pickering and Barry (1991). *Language and Cognitive Processes*, 8, 129–146.
- Grodner, D. J., & Gibson, E. A. F. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, 29, 261–291.
- Gundel, J., Hedberg, H., & Zacharski, R. (1993). Referring expressions in discourse. *Language*, 69, 274–307.
- Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In *Proceedings of NAACL* (Vol. 2, pp. 159–166).
- Han, J. I. (1992). Syntactic movement analysis of Korean relativization. *Language Research*, 28(2), 335–357.
- Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, 14, 25–38.
- Hashimoto, K. (2007). Subject/object asymmetry in the comprehension of English relative clauses by Japanese learners of English. *University of Queensland Working Papers in Linguistics*. Retrieved October 1, 2009, from <http://espace.library.uq.edu.au/eserv/UQ:23700/HashimotoFINAL.pdf>.
- Havik, E., Roberts, L., van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing Subject–Object Ambiguities in the L2: A Self–Paced Reading Study With German L2 Learners of Dutch. *Language Learning*, 59, 73–112.
- Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language*, 75, 244–85.

- Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative-clause sentences. *Journal of Verbal Learning and Verbal Behavior*, 20, 417–430.
- Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22, 369–397.
- Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, 120, 901–931.
- Ishizuka, T., Nakatani, K., & Gibson, E. (2006). Processing Japanese relative clauses in context. Paper presented at *the 19th Annual CUNY Conference on Human sentence processing*, March 2006.
- Izumi, S. (2003). Processing difficulty in comprehension and production of relative clauses by learners of English as a second language. *Language Learning*, 53, 285–323.
- 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.
- Jackson, C., & Roberts, L. (2010). Animacy affects the processing of subject-object ambiguities in the second language: Evidence from self-paced reading with German second language learners of Dutch. *Applied Psycholinguistics*, 31, 671–691.
- 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.
- Juffs, A. (1998). Main verb vs. reduced relative clause ambiguity resolution in second language sentence processing. *Language Learning*, 48, 107–47.
- Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society Volume*, 102, 199–225.
- Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research*, 21, 121–151.
- Juffs, A. & Harrington, M. (1995). Parsing effects in L2 sentence processing: subject and object asymmetries in wh-extraction. *Studies in Second Language Acquisition*, 17, 483–516.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory capacity. *Psychological Review*, 99, 122–149.
- Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: new frontiers of evidence and arguments. *Psychological Review*, 103, 773–780.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processing in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228–238.
- Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, 130, 169–183.
- Kane, M. J., D. Z. Hambrick, S. W. Tuholski, O. Wilhelm, T. W. Payne, and R. W. Engle. (2004). The generality of working-memory capacity: a latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133, 189–217.

- Kang, Y. -S. (1986). *Korean syntax and universal grammar*. Unpublished Ph.D. Dissertation, Harvard University.
- Kanno, K. (2007). Factors affecting the processing of Japanese relative clauses by L2 learners. *Studies in Second Language Acquisition*, 29, 197–218.
- Keenan, E. L. & Comrie, B. (1997). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8, 63–99.
- Kim, J. H. (2008). *Working memory effects on bilingual sentence processing*. Doctoral dissertation, University of Illinois at Urbana Champaign.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30, 580–602.
- 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.
- Kurilecz, M. (1969). *Man and his world: A structured reader*. New York: Crowell.
- Kwon, N. (2008). *Processing of Syntactic and Anaphoric Gap-Filler Dependencies in Korean: Evidence from Self-Paced Reading Time, ERP and Eye-Tracking Experiments*. Doctoral dissertation. University of California, San Diego.
- Kwon, N., Polinsky, M., & Kluender, R. (2006). Subject preference in Korean. In D. Baumer, D. Montero, & M. Scanlon (Eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics* (pp. 1–14). Somerville, MA: Cascadilla Proceedings Project.
- Lim, J-H. (2011). *Second language processing in reading and translation*. Doctoral dissertation. University of Illinois at Urbana Champaign.
- Lin, C., & Bever, T. G. (2006). Subject preference in the processing of relative clauses in Chinese. In D. Baumer, D. Montero, & M. Scanlon (Eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics* (pp. 1–14). Somerville, MA: Cascadilla Proceedings Project.
- Lin, Y., & Garnsey, S. M. (2011). 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. 241–275).
- Lee, J.-H. (2010). A subject–object asymmetry in the comprehension of wh-questions by Korean learners of English. *Applied Linguistics*, 31, 136–155.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126–1177.
- Love, T., Maas, E., & Swinney, D. (2003). The influence of language exposure on lexical and syntactic language processing. *Experimental Psychology*, 50, 204–216.
- MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In B. MacWhinney (Ed.), *The emergence of language*. Mahwah, NJ: Erlbaum.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109, 35–54.
- MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, 24, 56–98.
- MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- MacWhinney, B., & Pleh, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition*, 29, 95–141.

- Mak, W., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, 47, 50–68.
- Mak, W., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language*, 54, 466–490.
- Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, 27, 53–78.
- Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 244, 522–523.
- McDonald, M. C. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55, 381–401.
- McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 4, 67–91.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63, 81–97.
- Miller, K. A. (2011). Rethinking nonnative processing constraints: Evidence from L2 French. In L. Plonsky & M. Schierloh (Eds.), *Selected Proceedings of the 2009 Second Language Research Forum: Diverse Contributions to SLA* (pp. 109–120). Somerville, MA: Cascadilla Proceedings Project.
- Mitchell, D. C., Cuetos, F., Corley, M., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, 24, 469–488.
- Miyamoto, E., & Nakamura, M. (2003). Subject/object asymmetries in the processing of relative clauses in Japanese. In G. Garding & M. Tsujimura (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics* (pp. 342–355). Somerville, MA: Cascadilla Press.
- Nicol, J. L., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18, 5–20.
- O’Grady, W. (1997). *Syntactic development: The acquisition of English*. Chicago, IL: University of Chicago Press.
- O’Grady, W., Lee, M., & Choo, M. (2003). A subject-object asymmetry in the acquisition of relative clauses in Korean as a second language. *Studies in Second Language Acquisition*, 25, 433–448.
- Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study on second language learning after childhood: effects of proficiency. *Journal of Cognitive Neuroscience*, 17, 1212–1228.
- Omaki, A. (2005). *Working memory and relative clause attachment in first and second language processing*. MA Thesis, Department of Second Language Studies, University of Hawaii.
- Omaki, A., & Arij, K. (2005). Testing and attesting the use of structural information in L2 sentence processing. In L. Dekydtspotter, R. A. Sprouse & A. Liljestr nd (Eds.), *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference* (pp. 205–218). Somerville, MA: Cascadilla Proceedings Project.
- Osterhout, L., McLaughlin, J., Pitkanen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning*, 56, 199–230.
- 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: a study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, 24, 501–528.
- Pearlmutter, N. J., & MacDonald, M. E. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, 34, 521–542.
- Pickering, M. J., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, 6, 229–59.
- 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, 940–961.
- Pickering, M. J., & Van Gompel, R.P.G. (2006). Syntactic parsing. In M.A. Gernsbacher & M.J. Traxler (Eds.), *Handbook of psycholinguistics*, 2nd ED. (pp. 455–503). San Diego: Academic Press.
- Pienemann, M. (1998). *Language processing and second language development*. Amsterdam: John Benjamins.
- R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org/>.
- Rah, A., & Adone, D. (2010). Relative clause versus main verb ambiguity in L2 learners at different proficiency levels. *Studies in Second Language Acquisition*, 32, 79–109.
- 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, 358–74.
- Real, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57, 1–23.
- Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, 32, 299–331.
- Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at gap positions in children's sentence processing. *Journal of Psycholinguistic Research*, 36, 175–188.
- Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57, 348–379.
- Schachter, J. & Yip, V. (1990). Grammaticality judgments: Why does anyone object to subject extraction? *Studies in Second Language Acquisition*, 12, 379–392.
- Schriefers, H., Friederici, A. D., & Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, 34, 499–520.
- Schwartz, B.D. (1992). Testing between UG-based and problem-solving models of L2A: Developmental sequence data. *Language Acquisition*, 2, 1–19.
- Schwartz, B. D. (1998). The second language instinct. *Lingua*, 106, 133–60
- Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, 12, 40–72.
- Segalowitz, N. (2003). Automaticity and second languages. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 382–388). Oxford: Blackwell.
- Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, 13, 272–281.

- Song, H. S., & Schwartz, B. D. (2009). Testing the fundamental difference hypothesis: L2 adult, L2 child, and L1 child comparisons in the acquisition of Korean wh-constructions with negative polarity items. *Studies in Second Language Acquisition*, 31, 323–361.
- Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research*, 25, 13–41.
- Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1, 227–245.
- Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, 116, 71–86.
- Swets, B., Desmet, T., Hambrick, D.Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology*, 136, 64–81.
- Traxler, M. J. (2007). Working memory contributions to relative clause attachment processing: A hierarchical linear modeling analysis. *Memory & Cognition*, 35, 1107–1121.
- Traxler, M. J., Morris, R., & Seely, R. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69–90.
- Traxler, M., & Pickering, M. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35, 542–562.
- Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, 53, 204–224.
- 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.
- Truscott, J., & Sharwood-Smith, M. (2004). Acquisition by processing: A modular perspective on language development. *Bilingualism: Language and Cognition*, 7, 1–20.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154.
- Ueno, M., & Garnsey, S. M. (2008). An ERP study of the processing of subject and object relative clauses in Japanese. *Language and Cognitive Processes*, 23, 647–688.
- Ullman, M.T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, 4, 105–122.
- Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 30, 1302–1321.
- Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 33, 330–407.
- Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A retrieval interference theory of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49, 285–316.
- Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55, 157–166.
- Van Patten, B. (2002). Processing instruction: an update. *Language Learning*, 52, 755–803.

- Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics*, 25, 315–339.
- Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 119–161). Cambridge, MA: MIT Press.
- Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, 85, 79–112.
- Waters, G. S., & Caplan, D. (1996a). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, 103, 761–772.
- Waters, G. S., & Caplan, D. (1996b). Processing resource capacity and the comprehension of garden path sentences. *Memory and Cognition*, 24, 342–55.
- Waters, G. S., & Caplan, D. (1996c). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, 49A, 51–74.
- Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, and Computers*, 35, 550–564.
- Waters, G. S., & Caplan, D. (2004). Verbal working memory and on-line syntactic processing: Evidence from self-paced listening. *The Quarterly Journal of Experimental Psychology*, 57A, 129–163.
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D., J., MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, 58, 250–271.
- Williams, J. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, 9, 71–88.
- Williams, J. N., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh-questions: parsing strategies and plausibility constraints. *Applied Psycholinguistics*, 22, 509–540.
- Yoon, J. (1995). The semantics of relative clauses in Korean. In N. Akatsuka (Ed.), *Japanese/Korean linguistics, IV* (pp. 413–428). Stanford, CA: Center for Study of Lang. & Information, Stanford University.