USE OF PLAUSIBILITY INFORMATION AND STRUCTURAL PREFERENCE IN RESOLVING SENTENCE AMBIGUITY IN KOREAN

BY

JINHEE CHOO

DISSERTATION

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

Urbana, Illinois

Doctoral Committee:

Associate Professor Kiel Christianson, Chair
Professor Susan Garnsey, Director of Research
Professor Silvina Montrul
Assistant Professor Annie Tremblay
ABSTRACT

This study investigated online processing of temporarily ambiguous relative clause (RC) constructions in Korean to determine how multiple sources of information, such as syntactic preference and semantic plausibility, were employed in reading comprehension by three groups of Korean speakers with different L1 backgrounds: native speakers of Korean (NSs), heritage speakers of Korean (HSs), and adult second-language (L2) learners of Korean (K2s). HSs in the current study were Korean-English early bilinguals whose dominant language was English while K2s were international students and immigrants living in Korea. The main research focus in this study was to explore how these three groups of Korean speakers, who exhibited different levels of Korean-language proficiency, would integrate information on semantic plausibility and syntactic preferences while resolving temporary ambiguities. Sentence processing data from NSs were used to test existing major theories of sentence processing and comparable data from HSs and K2s were considered to determine the extent to which each group performed native-like sentence processing while resolving temporary ambiguities.

The plausibility of a noun as a direct object of either the embedded RC verbs or the main verbs was manipulated for the two types of RC constructions in Korean: the SR (Subject + Relative clause) construction with two objects and the SOR (Subject + Object + Relative clause) construction with one object and a required structural reanalysis. In addition, individual working memory (WM) capacity was measured through a reading span test (RST) in Korean to determine whether individual variations in WM capacity could influence the use of information on semantic plausibility and syntactic preferences during the resolution of temporary ambiguities in Korean.
The processing load of information on syntactic preferences and semantic plausibility and the final interpretations of the two Korean RC constructions were measured by using a self-paced word-by-word moving-window reading paradigm. The dependent variables were reading times and accuracy of responses to the end-of-sentence comprehension questions. The sample included 40 NSs, 40 HSs, and 42 K2s. In addition, 26 NSs participated in a norming study for the plausibility of the experimental materials. Reading times at each word position in a sentence and responses to comprehension questions were analyzed for further statistical analyses.

The results indicate that both semantic plausibility and syntactic preferences for a simpler analysis had considerable influence on the overall sentence-parsing process in Korean. In general, the results for NSs provide support for both syntax-first serial models and parallel multiple constraint-based models. However, the generally mixed results for the main effects of the plausibility bias and the interaction effect of plausibility and structural preferences on both sentence processing and final interpretations provide support more for parallel multiple constraint-based models than for syntax-first serial models. In addition, these results provide a clear indication of Good-Enough (GE) or shallow parsing in the ambiguity resolution process as well as the final interpretation for both native and non-native speakers of Korean.

HSs showed slower reading times than NSs because of their lower Korean proficiency and provided mixed results for native-like processing, whereas K2s showed less native-like processing patterns than HSs. In terms of the use of plausibility information during the ambiguity resolution process, both HSs and K2s showed native-like sensitivity to plausibility information for both SR and SOR constructions. In general, HSs were more sensitive to plausibility information than K2s, but K2s showed a stronger structural preference for the SR construction than NSs and HSs.
WM had a main effect on the ambiguity resolution process during the online reading process as well as during the interpretation process with some group differences in the effect. There were significant differences between the three groups in terms of their reading proficiency and WM capacity in Korean. WM had stronger main effects on many of the measurements in the current study for HSs than for NSs and K2s.

The results, including those for the use of plausibility information and the syntactic preference in relation to individual WM capacity during sentence processing by these three groups of Korean speakers, provide valuable insights into bilingual sentence processing and present a new way of examine existing models of L1 sentence processing. In addition, these results allow for analyses of competing syntactic and plausibility factors influencing the resolution of temporary ambiguities in Korean, which should provide a better understanding the nature of language-processing mechanisms.
ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my thesis director, Dr. Susan Garnsey for her inspiration, guidance and support to complete the thesis. I would also like to give special thanks to my academic advisor, Dr. Kiel Christianson for being a tremendous help in completing my doctoral program. I am very grateful to the remaining members of my dissertation committee, Dr. Silvina Montrul and Dr. Annie Tremblay for their thoughtful and detailed comments.

Last, but certainly not least, I must acknowledge with tremendous and deep thanks to my family who are my biggest supporters through all my studies. I am forever indebted to them for giving me their love and heart.
# TABLE OF CONTENTS

DEFINITION OF TERMS ........................................................................................................... vii  

CHAPTER 1 INTRODUCTION ..................................................................................................... 1  

CHAPTER 2 SENTENCE PROCESSING THEORIES AND RESEARCH ........................................ 17  

CHAPTER 3 TEMPORARY AMBIGUITY RESOLUTION IN KOREAN ...................................... 93  

CHAPTER 4 METHOD ............................................................................................................. 121  

CHAPTER 5 RESULTS AND DISCUSSION ............................................................................. 140  

REFERENCES ...................................................................................................................... 193  

APPENDIX A KOREAN PROFICIENCY TEST ........................................................................ 206  

APPENDIX B ENGLISH PROFICIENCY TEST ..................................................................... 208  

APPENDIX C EXPERIMENTAL SENTENCES ..................................................................... 209
The following abbreviations are used to refer to various linguistic terms employed in the study:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>Accusative case marker/particle</td>
</tr>
<tr>
<td>DAT</td>
<td>Dative case marker</td>
</tr>
<tr>
<td>DEC</td>
<td>Declarative sentence-type suffix</td>
</tr>
<tr>
<td>DEL</td>
<td>Delimiter</td>
</tr>
<tr>
<td>IN</td>
<td>Indicative mood suffix</td>
</tr>
<tr>
<td>INF</td>
<td>Infinitive suffix</td>
</tr>
<tr>
<td>LOC</td>
<td>Locative case marker</td>
</tr>
<tr>
<td>NOM</td>
<td>Nominative case marker</td>
</tr>
<tr>
<td>PAS</td>
<td>Passive suffix</td>
</tr>
<tr>
<td>PAST</td>
<td>Past tense and perfect aspect suffix</td>
</tr>
<tr>
<td>PROG</td>
<td>Progressive suffix</td>
</tr>
<tr>
<td>REL</td>
<td>Relativizer suffix</td>
</tr>
<tr>
<td>SUBJ</td>
<td>Subject marker (nominative case marker + topic marker)</td>
</tr>
<tr>
<td>TOP</td>
<td>Topic marker</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1. L1 Sentence Processing

Sentence comprehension is an incremental process that sometimes leads to temporary ambiguities in how the words in a sentence are related. In the English sentence *The best man bought a new tuxedo is looking forward to the upcoming wedding*, there is some temporary syntactic ambiguity in the relationship between the verb *bought* and the preceding noun phrase (NP) *the best man*. The verb can be either the sentence’s main verb (MV), as in *The best man bought a new tuxedo*, or the verb in the reduced passive relative clause (RC), as it turns out to be in the example, whose non-reduced counterpart is *The best man who was bought a new tuxedo is looking forward to the upcoming wedding*. The sentences *The best man bought a new tuxedo* and *The best man bought a new tuxedo is looking forward to the upcoming wedding* begin identically, but the latter is followed by disambiguating information at the word *is*. Addressing the question of when and how individuals resolve such ambiguities can provide important insights into the properties of the mechanism underlying sentence comprehension.

Other languages also have similar temporary structural ambiguities. Although few studies have thus far focused on other languages, there is growing interest in such ambiguities in other languages (Lin & Garnsey, 2011; Lin & Bever, 2011; Papadopoulou & Clahsen, 2006; Demestre & Garcia-Albea, 2004; Kamide, Scheepers, & Altmann, 2003). Romance and Germanic languages have received the most attention, but typologically different languages such as Korean and Japanese have recently received increasing attention (Kim, 2005; Kim, 2004; Kamide, Altmann, & Haywood, 2003; Hirose, 2002).
An example of a temporary ambiguity in Korean, which is the language considered in the present study, is illustrated below in (1) and (2). These two sentences are identical until the fifth word (in bold type), which provides disambiguating information on the relationship between earlier words. In (1), the fifth word is an object-marked noun (cookies-ACC), but in (2) it is an obligatorily transitive verb (give). The sentence in (1) contains a Subject + RC (SR) construction with two direct objects (milk bottle and cookies) of two verbs (hold and buy, respectively). On the other hand, the sentence in (2) contains a Subject + Object + RC (SOR) construction with the only object milk bottle, which is initially analyzed as a direct object of the embedded RC verb hold, but should be interpreted instead as a direct object of the transitive MV give. At the fifth word in the SOR construction give, readers should reanalyze their initial incorrect analysis.

(1) **Subject + Relative Clause (SR) Construction**

엄마가 우유병을 들고있는 아기한테 과자를 사주었다.

*The mother bought cookies for the baby, who was holding his milk bottle.*

(2) **Subject + Object + Relative Clause (SOR) Construction**

엄마가 우유병을 들고있는 아기한테 떨어주려고 다가섰다.

*The mother approached to give a milk bottle to the baby she was holding.*

Both sentences contain RCs modifying the baby, but it is temporarily ambiguous whether *the milk bottle* is part of that RC or part of the main clause. The ambiguity arises because the beginnings of RCs are not marked in Korean. Thus, there is often ambiguity in multi-clause sentences with respect to which clause a particular word belongs to. This ambiguity is temporary because when the obligatorily transitive verb give appears in (2), *the bottle* has to be its object because no other object has yet appeared and objects always precede verbs in Korean.
There are two types of theories about the sentence processing involved in resolving ambiguities such as those in the sentences above. Syntax-first serial models (Ferreira & Clifton, 1986; Frazier, 1979, 1987) posit that only one syntactic structure is constructed during the initial parsing phase based on purely structural principles. In these models, various other types of information can influence later stages of sentence processing, but they do not influence the first stage of parsing.

In contrast, in parallel multiple constraint-based models (Garnsey, Pearlmutter, Myers, & Lotoky, 1997; MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, Tanenhaus, & Garnsey, 1994; Trueswell, Tanenhaus, & Kello, 1993), multiple syntactic structures are activated in parallel early in parsing based on all available information, including both syntactic and non-syntactic information. Information from different sources is combined, which can result in facilitation when the sources agree and delays when they conflict, i.e., competition (Garrett, 1990).

Previous studies have produced mixed results. Some have provided support for syntax-first serial models and others for parallel multiple constraint-based models (MacDonald & Seidenberg, 2006). This may be due to differences in experimental sentence types and structures, differences in how well the plausibility of the sentence materials is controlled, and differences in testing instruments and methods, among others (Jackendoff, 2007; McClelland, 1987; Trueswell, Tanenhaus, & Garnsey, 1994). As a result, the influence of non-syntactic information on initial sentence parsing remains a contentious issue in the sentence processing research.

Recently, some studies have questioned an assumption underlying both models. Both models expect the reader to accurately parse most sentences once the temporary ambiguity is resolved. However, some studies have provided evidence of incomplete parsing with lingering
interpretations from incorrect analyses of particularly difficult sentences (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Christianson, Luke, & Ferreira, 2010; Christianson, Williams, Zacks, & Ferreira, 2006; Ferreira, 2003; Ferreira, Bailey, & Ferraro, 2002; Sanford & Sturt, 2002; Swets, Desmet, Clifton, & Ferreira, 2008). Interestingly, a similar suggestion has been made about non-native language processing even for simpler sentences. Clahsen and Felser (2006a, b) proposed the “Shallow Structure Hypothesis” to account for parsing by adult learners of a second language (L2), suggesting that non-native speakers, even those with a fairly high level of proficiency, do not fully parse sentences in their L2, whereas native speakers’ incomplete parsing may be limited to sentences with more difficult temporary ambiguities.

Some studies have examined the role and effects of non-syntactic information on the process of resolving syntactic ambiguities and found that several non-syntactic factors sometimes play a role early in sentence parsing, including knowledge about verbs’ subcategorization possibilities and preferences (Mitchell, 1987; MacDonald et al., 1994; Trueswell, 1996; Trueswell et al., 1993; Garnsey et al., 1997), lexical frequency, and plausibility in resolving noun-noun vs. noun-verb ambiguities (Frazier & Rayner, 1987; MacDonald, 1993), and the combination of verb biases and thematic plausibility in the direct object vs. sentential complement ambiguities (Garnsey et al., 1997; Trueswell et al., 1993). Parallel multiple constraint-based models are especially compatible with the claim that sentence processing, including the resolution of temporary ambiguities, makes use of detailed frequency-based and combinatorial information on verb biases and the relative plausibility of particular word combinations, among others. (Patson, Darowski, Moon & Ferreira, 2009; Patson, Swensen, Moon & Ferreira, 2006).
Among various types of non-syntactic information, plausibility is the main focus of the present study. Among all possible types of information relevant to sentence processing, this study investigates 1) syntactic preferences for a simple structure and 2) whether the relative plausibility of nouns as direct objects of different verbs interacts with such preferences. The specific aspect of plausibility manipulated in this study is the thematic plausibility of a noun as a possible direct object of two different verbs in a sentence, the MV or the verb in an embedded RC verb (RCV). Thematic plausibility is one of the non-syntactic information constraints playing a leading role in the ambiguity resolution process not only in the final interpretation stage but also in the initial structure-building stage (Patson et al., 2009).

Another question with respect to the use of plausibility information during the process of resolving syntactic ambiguities is to what extent individual differences in working memory (WM) influence the processing of plausibility information. Given that sentence processing is complex and involves multiple cognitive components, a number of studies have examined the influence of cognitive factors such as WM on sentence processing. However, there is no clear consensus on what WM is and how it can be measured (Juffs, 2004). There is some evidence suggesting that differences in WM can influence the ability to make rapid use of plausibility information. For example, individuals with high WM spans have been shown to be more sensitive to plausibility information in general (King & Just, 1991; Pearlmutter & MacDonald, 1995). Some studies have found that individuals with high WM spans take longer to read ambiguous stretches of sentences (MacDonald, Just & Carpenter, 1992) because they are more likely to keep options open than those with low WM spans. However, other studies have found no significant effect of WM on sentence processing (Clifton, Traxler, Mohamed, Williams, Morris & Rayner, 2003).
In English, there is a prevailing tendency to take a post-verbal noun as the direct object of the immediately preceding verb, and this tendency is sometimes observed even when various types of linguistic information such as the type of verb, the type of argument, frequency, and collocation information provide evidence against that interpretation (Wiechmann, 2006). This trend has been labeled in various ways, but the best-known label is the Minimal Attachment (MA) principle in syntax-first serial models (Frazier, 1987). The MA principle is based purely on syntactic information and is hypothesized to be universal. In parallel multiple constraint-based models, the MA principle is just one of many, but it can be a very strong constraint.

There are constraints about the overall syntactic configuration such as the MA principle, and there are also ones based on particular words such as thematic requirements and frequency, among others. Verbs represent one of the most essential sentence components. By their nature, they strongly constrain what the sentence structure can be, as well as providing meaning and thematic relationships related to that meaning. Verbs occur in different locations across languages, including Subject-Verb-Object (e.g., English and French), Subject-Object-Verb (e.g., Korean and Japanese), Verb-Subject-Object (e.g., Tagalog and Classical Hebrew), and Verb-Object-Subject (e.g., Arabic). Given that verbs play a critical role in determining the syntactic structure of the sentence, investigating temporary structural ambiguities with languages having verbs in different locations should shed some light on the question of how the location of verbs influences ambiguity resolution process varies across languages.

In English, with its canonical SVO word order, verbs play a more important role in sentence processing than most other sentence components. Verbs are generally located right after a subject and thus constrain sentence structure relatively early in the parsing process. In addition, some types of information that are not purely syntactic, such as thematic relationships, are
closely related to verbs, so verbs are a major factor in deciding the sentence structure in the earlier stages of sentence processing in English.

In addition, the verb also plays a major role in the ambiguity resolution process involving reduced RCs with a non-canonical object-verb (OV) structure such as the example sentences give earlier: *(The best man (who was) bought a new tuxedo is looking forward to the upcoming wedding ceremony.)* The interpretation of reduced RC verbs in English can be influenced by various factors such as lexical frequency (Trueswell, 1996), morphological information (Rah & Adone, 2010), thematic relationship information (Trueswell et al., 1994), and syntactic biases, which reflect more structurally defined parsing principles, such as the MA principle (Grodner, Gibson & Tunstall, 2002; Frazier, 1979, 1987; Gorrell, 1995).

In English, people are more likely to have difficulty reading reduced RCs, which have a non-canonical OV order, particularly when the pre-verbal noun is less plausible as the object of the verb (Long & Prat, 2008; McRae, Spivey-Knowlton & Tanenhaus, 1998; Pearlmutter & MacDonald, 1995). This is referred to as “the plausibility effect,” that is, readers take longer on implausible words, but shorter at the later disambiguating words, presumably because the earlier apparent implausibility led to revision before the disambiguation was reached. Pickering and Traxler (1998, 2002, 2003) found similar plausibility effects in L1 sentence processing with different sentence structures. Depending on the degree of plausibility of the relationship between sentence components, different patterns of reading times can be observed even with the same syntactic structure.

One of this study’s main motives is to explore the universality of sentence processing mechanisms across different languages. Most sentence processing theories, including the two major models discussed earlier, are based mainly on data from L1 speakers of English. If
previous studies focusing on native speakers of English have produced mixed results about the resolution of ambiguities and the main research interest in sentence processing models is to identify and explain universal language rules and phenomena, then it should be meaningful to test sentence processing models with languages that are typologically different from English, such as Korean, which has SOV word order and many other linguistic features that are different from English but which has attracted little attention from researchers to date. Cross-linguistic comparisons can be useful for determining which sentence processing models provide a better explanation for universal phenomena across languages. In addition to providing a way to test the universality of parsing mechanisms across languages, investigating Korean on its own can be important in that it has several interesting features (Kim, 2005).

Korean poses theoretically challenging questions for major sentence parsing models, including both syntax-first serial models and parallel multiple constraint-based models (Clifton, 2000). Kim (2004) suggested that grammatical features of NPs such as case and number may play an especially important role in sentence processing in a head-final language such as Korean. They may be as important as verb information is in a head-initial language such as English. Because verbs do not appear until the ends of clauses in Korean, nouns may play a more influential role in deciding sentence structures than they do in English. In addition, Korean may require more WM resources than English because native speakers of Korean can make better and more frequent use of noun-related information, such as case markers and plausibility to project an initial sentence structure even before the verb appears (Kwon, 2008). Unlike in English, which employs early verb-based cues to project the syntactic structure of a sentence, incremental processing in Korean may place a heavier burden on WM to process and keep all the incoming linguistic input activated until the late-arriving verb at the end of the sentence. If non-syntactic
information such as the semantic plausibility of a noun as a direct object of the subsequent verb influences initial parsing or reanalysis procedures in Korean, which theoretical parsing model provides a better account of such effects?

1.2. Bilingual Sentence Processing

Previous psycholinguistic research on the resolution of ambiguities has typically focused on L1 sentence processing, and few parsing models make predictions about parsing patterns in L2. In general, bilingual processing is thought to be slower and less accurate than L1 processing. However, like L1 sentence processing, bilingual processing has been found to be incremental (Fender, 2001). One of the major research topics in bilingual processing has been whether L2 learners show native-like processing patterns (setting aside apparent quantitative differences such as the processing speed). Until recently, studies of bilingual processing focused more on syntactic processing than on semantic processing (Williams, 2006). Although previous studies have produced some mixed results for syntactic processing in L2, semantic processing in L2 is generally thought to be more native-like than syntactic processing.

The Shallow Structure Hypothesis (SSH), an influential theory of bilingual processing proposed by Clahsen and Felser (2006a, b), posits that even advanced L2 learners project a shallow syntactic structure based on simple information about simple syntactic structures and on non-syntactic information such as lexical and semantic information, so that adult L2 learners’ sentence processing is qualitatively different from that of native speakers. Previous studies of bilingual processing have tested this hypothesis and produced mixed results.

Like L1 sentence processing, bilingual processing can show strong plausibility effects. When advanced L2 learners showing native-like garden-path (GP) effects read sentences such as *After Bill drank the water proved to be poisoned*, they take longer to read the disambiguating
verb *proved*, which indicates that *the water* cannot be the direct object of *drank* even though it is highly plausible in that role (Juffs, 2004; Juffs & Harrington, 1996). In a self-paced L2 reading study with subject and object ambiguities in sentences such as *The inspector warned the boss/crimes would destroy very many lives* and *While the band played the song/beer pleased all the customers*, Felser and Roberts (2004, 2011) suggested that Greek L2 learners of English show stronger plausibility effects than native speakers of English do, and found that the plausibility effect in the form of instant recovery from the initial misanalysis of tested constructions (complement vs. adjunct clauses).

A growing trend in research on bilingual processing has been the inclusion of adult L2 learners with a high level of proficiency in an immersed environment (Rah & Adone, 2010). Another interesting population in research on bilingual processing is heritage speakers, who have thus far received little attention from researchers. Heritage speakers are those who have been exposed to both languages fairly early but tend to have “incomplete” L1 knowledge (Montrul, 2008). In the present study, the term “heritage speakers” refers to early Korean-English bilinguals whose first language was Korean but current language is primarily English. This study’s sample of bilinguals lived in the U.S., and their dominant language was English while their heritage language (HL) was Korean (based on their self-evaluation and proficiency tests). These heritage speakers had lived immersed in English since before puberty, and many were exposed to English and Korean simultaneously at home.

Heritage speakers of Korean share some common characteristics with children of L1 speakers of Korean because both have been exposed to Korean since birth and have limited proficiency in Korean. They are also like adult L2 learners of Korean in that they tend to have limited proficiency in Korean. Thus, the question of whether heritage speakers or adult L2
learners of Korean show syntactic preferences and plausibility effects in Korean similar to those found for native speakers remains unanswered. Given that they have been exposed to their HL since birth but may not have achieved native-like proficiency in grammar or language processing, how do heritage speakers handle multiple sources of information during their resolution of ambiguities? Will they rely more on syntactic information (like children of L1 speakers) (Clahsen & Felser, 2006b) or on non-syntactic information (like adult L2 learners)? In other words, it should be interesting to compare native speakers’ sentence processing with that of heritage speakers as well as adult L2 learners to investigate the effects of the age of language acquisition, which has typically been found to be confounded by proficiency in studies of adult L2 learners (Flege, 2007). Such a comparison should provide a better understanding of sentence processing.

This study is guided by the following three research questions:

(i) Is there a general structural preference for SRs over SORs in Korean?

(ii) When and how do Korean speakers at different levels of language proficiency make use of plausibility information to help resolve structural ambiguities? Does this plausibility influence initial structural decisions as well as final interpretations?

(iii) Does WM capacity influence readers’ ability to process information plausibility information during the resolution of structural ambiguities?

In this study, three groups of participants were considered: a baseline group composed of native speakers of Korean and two experimental groups of non-native speakers of Korean, namely heritage speakers and adult L2 learners of Korean (i.e., those with a different L1). Because the two non-native groups needed to be fairly proficient in Korean to be able to read Korean RCs in the experiment, the participants were screened using cloze tests and a survey of
their language history. Each experimental session consisted of surveys, cloze tests of Korean- and/or English-language proficiency, a reading span test (RST) in Korean to assess WM capacity, and the main experiment with a self-paced moving-window reading task.

Two types of temporarily ambiguous RC constructions in Korean were employed: the SR (Subject + RC) and SOR (Subject + Object + RC) constructions (Kim, 2004). The SR construction has previously been found to be preferred over the SOR construction (Kim, 2004) because it is structurally simpler, more common, and more consistent with the default word order in Korean. Plausibility was manipulated to be biased toward one or the other syntactic construction or to be neutral (SR-biased, Neutral, and SOR-biased). The plausibility of the noun as a direct object of the subsequent RCV or MV was controlled using two norming surveys.

A semantically biased condition can show either concurrence or conflict with general world knowledge and collocational frequency (e.g., entertain a dessert vs. serve a dessert). If we ignore plausibility relationships between nouns and verbs, as syntax-first serial models do, then the two nouns in an SR syntactic construction can be interpreted as the direct objects of the two verbs following them in order (i.e., the first noun with the embedded RC verb and the second noun with the main verb (MV)). On the other hand, the single noun in the SOR syntactic condition is likely to be interpreted initially as the direct object of the immediately following embedded RC verb and reanalyzed later as the direct object of the MV when the reader encounters a MV that requires a direct object. As a result, the SOR syntactic structure is likely to require structural reanalysis, and thus, it may be more computationally complex than the SR construction.
Table 1

*Design and Conditions: 2 Constructions x 3 Plausibility Conditions*

<table>
<thead>
<tr>
<th>Sentence Structure</th>
<th>SR-Biased</th>
<th>Neutral</th>
<th>SOR-Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SR Construction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(two objects)</td>
<td>SR-SR</td>
<td>SR-N</td>
<td>SR-SOR</td>
</tr>
<tr>
<td></td>
<td>(facilitating)</td>
<td></td>
<td>(conflicting)</td>
</tr>
<tr>
<td>N-obj₁ + RCV (Plaus)</td>
<td>N-obj₁ + RC/MV (Plaus)</td>
<td>N-obj₁ + RCV (Implaus)</td>
<td></td>
</tr>
<tr>
<td>N-obj₁ + MV (Implaus)</td>
<td>N-obj₂ + MV (Plaus)</td>
<td>N-obj₁ + MV (Plaus)</td>
<td></td>
</tr>
<tr>
<td>N-obj₂ + MV (Plaus)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SOR Construction</strong></td>
<td>SOR-SR</td>
<td>SOR-N</td>
<td>SOR-SOR</td>
</tr>
<tr>
<td>(one object with reanalysis)</td>
<td>(conflicting)</td>
<td></td>
<td>(facilitating)</td>
</tr>
<tr>
<td>N-obj₁ + RCV (Plaus)</td>
<td>N-obj₁ + RC/MV (Plaus)</td>
<td>N-obj₁ + RCV (Implaus)</td>
<td></td>
</tr>
<tr>
<td>N-obj₁ + MV (Implaus)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the condition labels used here, the letters before the dash describe the sentence structure and the letters after the dash describe which structure is consistent with contextual plausibility.

Each condition represents the factorial combination of two manipulations: Sentence structure and contextual plausibility. Six conditions results from crossing two sentence structures (an SR construction with two objects vs. an SOR construction with one object) by three plausibility conditions (SR-biased, SOR-biased, and Neutral). For instance, SR-SR means that the sentence has the SR syntactic construction (with two objects) and contextual plausibility supports that SR-structure, while SOR-SR means that the sentence has the SOR syntactic construction (with one object) but contextual plausibility supports the other SR construction.

(1) **Subject + Relative Clause (SR) Construction**

The mother bought cookies for the baby, who was holding his milk bottle.

If the reader initially processes the first object-marked noun in the sentence as the object of the immediately following first verb in the sentence, which is the embedded RCV at word 3, then the continuation of the sentence in the SR-SR and SR-N conditions provides support for that initial interpretation and there is thus no need for structural reanalysis when the second object-
marked noun appears at word 5. In these conditions, the meaning of the RCV provides support for the initial structural decision and thus for the rest of the sentence. However, in the SR-SOR condition, the first object-marked noun is not plausible as the object of the RCV that immediately follows it, so the SR structure is implausible. This should lead to some processing cost for the embedded RCV because of either the implausibility of the structurally preferred relationship between the noun and the verb or the switch to a structurally less preferred SOR interpretation if plausibility influences structural choices quickly enough, or both. If it turns out at word 5 that the sentence actually has the SR structure after all, then there should be an increase in the processing cost. However, if the readers do not make use of plausibility in their early structural decisions, then there should be no processing cost at word 5 for the SR-SOR condition because it eventually confirms the SR analysis that the reader has continued to pursue despite its implausibility.

(2) Subject + Object + Relative Clause (SOR) Construction

The mother approached to give a milk bottle to the baby she was holding.

In contrast, in the SOR construction, it turns out at the obligatory transitive MV (word 5) that the earlier object-marked noun has to be its object rather than the object of the RCV, so reanalysis is required. Syntax-first serial models predict that the SOR construction should always require reanalysis regardless of the contextual plausibility manipulation and thus should be approximately equally difficult at word 5. In contrast, parallel multiple constraint-based models predict that when plausibility provides support for the SOR syntactic construction in the SOR-SOR condition, the processing cost at word 5 should be lower than in the SR-SOR condition where the context does not support the SOR construction. In the SOR context, if the first object-
marked noun is not plausible as the object of the immediately following RCV, then there is an expectation that another verb will appear later in the sentence for which it is a plausible object, thereby providing support for the SOR structure rather than the SR structure. If the sentence then turns out at word 5 to be consistent with that expectation in the SOR-SOR condition, then either no analysis should be required or reanalysis should be easier than in the SOR-SR or SOR-N conditions. The processing cost of reanalysis required at word 5 for the SOR-SOR condition is lower than that for the SOR-SOR or SOR-N conditions. Table 2 shows the predictions of the two kinds of models’ about reading times at words 3 and 5 (and the words following them, since the effects are likely to spill over to subsequent word positions). The crucial differences in their predictions are indicated in bold type.

Table 2

*Expected Reading Time Patterns at Critical Positions*

<table>
<thead>
<tr>
<th>Critical Word Position</th>
<th>Serial Models</th>
<th>Parallel Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Word 3</td>
<td>Word 5</td>
</tr>
<tr>
<td><strong>SR construction</strong></td>
<td>SR bias &lt; SOR bias</td>
<td>(SR-SR=SR-N)=SR-SOR</td>
</tr>
<tr>
<td><strong>SOR construction</strong></td>
<td>SR bias &lt; SOR bias</td>
<td>SOR-SR=(SOR-N=SOR-SOR)</td>
</tr>
<tr>
<td></td>
<td>SR-SOR &lt; SOR-SOR</td>
<td>SR-SOR &lt; SOR-SOR</td>
</tr>
</tbody>
</table>

The rest of this paper is organized as follows: Chapter 2 provides a review of sentence processing theories and research in L1 as well as previous research on bilingual sentence processing with respect to the use of information on semantic plausibility and syntactic preferences during the process of resolving temporary ambiguities. Chapter 3 provides an overview of Korean syntax and discusses the structure of the Korean RCs considered in the study.
in the context of the research questions. Finally, Chapter 5 presents the results and their implications for research on L1 and L2 sentence processing.
Chapter 2 provides an overview of existing theories of L1 and L2 sentence processing and a review of previous studies of temporary ambiguity resolution in English focusing on the use of information on semantic plausibility and syntactic preferences, and the influence of working memory. In the next section, a review of previous studies relevant to this study will be presented, starting with an overview of two major models of L1 sentence processing: syntax-first serial models and parallel multiple constraint-based models.

2.1. Major L1 Sentence Processing Theories and Research

2.1.1. Syntax-first serial models vs. Parallel multiple constraint-based models

Sentence processing by native speakers is assumed to be a nearly automated process to incorporate the incoming linguistic input into the ongoing representation in order to develop sentence structures and, ultimately, interpretations. Because this is a very rapid and unconscious process, it is very difficult to analyze. As a result, one of the ways to investigate online sentence processing is to employ a temporarily ambiguous sentence structure, which can induce the reader to choose one interpretation among many options based on some available information.

Sentences that are temporarily ambiguous but grammatically correct (e.g., The horse raced past the barn fell) can be easily misunderstood by the reader and is referred to as GP sentences (Bever, 1970). Individuals are ultimately presumed to be able to resolve temporary structural ambiguities when they encounter additional disambiguating information later during their sentence parsing. When individuals read a sentence with temporary ambiguities, the pace of their reading often slows down as they try to understand the sentence. In this regard, by measuring reading times at each word, we can determine the places where the reader typically slows down in the sentence. This processing difficulty associated with temporary ambiguities is
referred to as GP effect, which can help the researchers to draw inferences about what
information sources are used for the choice of the structure.

Multiple sources of information are assumed to be available during parsing. There are
two well-known theoretical approaches that differ in terms of their views on when certain types
of information become available to the parser: syntax-first serial models and parallel multiple
constraint-based models.

Bever (1970) claimed that serial models can account for the preference for the simplest
syntactic structure, inducing the GP effect observed in many reading tasks. Several versions of
serial models have been proposed, including the Sausage-Machine (Frazier & Fodor, 1978), the
Garden Path Model (Frazier, 1979, 1987), the Phrase-Structure Approach (Gorrell, 1995), the
Construal Model (Frazier & Clifton, 1996), the Delay Model (Just & Carpenter, 1980), the
Information-paced Model (Inoue & Fodor, 1995), and the Head-driven Model (Pritchett, 1991).

The most influential version of the serial model is the GP Model (Frazier, 1987), which
involves two principles for the increased efficiency and economy of sentence processing:
minimal attachment and late closure. The minimal attachment (MA) principle is a way of
connecting a newly added phrase by using a minimum number of nodes in the phrase structure
tree needed to create a syntactically valid structure (Frazier, 1987). In English, this principle
often manifests itself as the NP attachment, where the NP is preferentially attached as the direct
object of the preceding verb. For example, according to the MA principle, the suspect is initially
parsed as the direct object of the verb knew in both (3a) and (3b), although it is in fact the subject
of a new clause in (3b).

(3a) I knew the suspect.

(3b) I knew the suspect had a gun.
On the other hand, the late closure (LC) principle connects a new phrase to the currently processed phrase or clause in the sentence and is typically employed for the adverbial or prepositional phrase (PP) attachment. For example, the sentence *I met a friend of mine who married a week ago* is parsed as *I met a friend of mine, and he/she married a week ago*, not as *I met a friend of mine a week ago, and he/she married*. The LC principle, along with other similar principles such as right association and locality, projects a structure with modifying and incoming new phrases being attached to the structurally closest possible node (Frazier & Fodor, 1978; Gibson & Pearlmutter, 1998). The MA principle is thought to be universally relevant to all languages (Garrett, 1990; Kim, 2004).

For the resolution of syntactic ambiguities, serial models presume distinct stages of syntactic parsing and interpretation. In the first stage of parsing, only syntactic information is prioritized for the initial structure-building process, and syntactic information includes the word class of each lexical item (e.g., a noun or a verb). According to syntactic principles such as the MA principle, with some temporary structural ambiguity, the reader selects the simplest structure in the first stage instead of pursuing multiple syntactic analyses or delaying the analysis of the input. Conducting one analysis at a time may require fewer memory resources, and thus, such an approach may be less costly than conducting multiple interpretations simultaneously (Frazier, 1979). In other words, preferences for simple structures in sentence processing may result from general cognitive limitations such as WM constraints. In the second stage, non-syntactic information (e.g., lexical, semantic, pragmatic, discourse, and general world knowledge information) is employed to monitor the validity of the initially projected structure. If there is some conflict between the first analysis conducted by the simplest structurally favored syntactic preference and non-syntactic information, then there is a GP effect that triggers reanalysis.
(Frazier, 1987; Frazier & Rayner, 1982). Accordingly, one of the disadvantages of serial models is the increased likelihood of a misanalysis or reanalysis in the second stage (Kim, 2004).

A number of studies of L1 reading have provided support for the serial view (Frazier, 1987; Frazier & Rayner, 1987; Y. Kim, 2001; Rayner, Carlson & Frazier, 1983). For instance, eye-tracking studies (Frazier & Rayner, 1982) have found that the reading time per character is longer for non-minimal structures than for minimal structures. Individuals tend to make more regressive eye movements when reading non-minimally attached sentences than minimally attached sentences, suggesting the reprocessing of the non-minimally attached sentences.

Ferreira and Clifton (1986) conducted an eye-tracking experiment to analyze the resolution of ambiguities in MVs and reduced RCs in English and provided support for the MA principle. They manipulated the plausibility (+/- animacy) of the first NP as an agent of the verb of the main clause and the types of embedded RC verbs (reduced or unreduced) to consider four conditions (animate/reduced, inanimate/reduced, animate/unreduced, and inanimate/unreduced).

(4) *The evidence/defendant examined by the lawyer turned out to be unreliable.* (reduced verb condition with an inanimate/animate agent)

They found that for the example (4), the participants preferred an interpretation of the main clause (*the defendant/evidence* as the subject of *examined*) before the disambiguating region by the lawyer regardless of the semantic plausibility condition.

They subsequently conducted an eye-tracking experiment and a self-paced reading experiment by employing reduced RCs and PP attachments to determine whether discourse information (MA-biased, non-MA-biased, and neutral) would guide the participants’ initial syntactic decisions.

(5a) *The editor played the tape/and agreed the story was big.* (MA: RC)
(5b) *The editor played the tape/agreed the story was big.* (Non-MA: RC)

(5c) *Sam loaded the boxes on the cart/before his coffee break.* (MA: PP)

(5d) *Sam loaded the boxes on the cart/onto the van.* (Non-MA: PP)

According to the results of both experiments, the participants consulted contextual information during their reanalysis. Based upon these results, they suggested that non-syntactic information is employed as extra help for a different analysis during the reanalysis.

Rayner, Carlson, and Frazier (1983) investigated the interaction between semantic/pragmatic information and syntactic information in online sentence processing by measuring eye movements and found no effect of real-world plausibility constraints on the initial processing of reduced RCs. They used semantically biased ambiguous sentences and tested MA and non-MA preferences for PPs.

(6a) *The spy saw the cop with binoculars but the cop didn’t see him.* (semantic plausibility for the minimal attachment principle)

(6b) *The spy saw the cop with a revolver but the cop didn’t see him.* (semantic plausibility against the minimal attachment principle)

In (6a), if the PP (*with binoculars*) is attached to the NP (*the cop*), not to the verb phrase (VP) (*the spy used binoculars to see the cop*), then an additional node is required. Therefore, the semantic plausibility in (6a) follows the MA principle. The participants in the study expressed a preference for the MA principle in the initial parsing stage even when semantic plausibility mitigated against the MA principle. This implies that plausibility did not influence the initial parsing process. However, plausibility influenced the final interpretation of the ambiguous sentence during the subsequent paraphrase task. They found that reading times were shorter in the facilitating condition in which semantic plausibility matched the syntactically preferred
structural analysis and suggested that there exist independent processors in the sentence comprehension system and temporal effects of multiple constraints on the sentence comprehension process.

In a self-paced reading study, Grodner, Gibson, and Tunstall (2002) found evidence of the effect of structural complexities and suggested that structural complexities play an important role in resolving structural ambiguities, such as the MV/reduced RCV ambiguity and the noun-noun/RC ambiguity.

(7a) The witness who the evidence (that was) examined by the lawyer implicated turned out to be unreliable. (MV/reduced RCV in a RC)

(7b) The witness said that the evidence (that was) examined by the lawyer was unreliable. (MV/reduced RCV in a sentence complement)

During in the ambiguity resolution process, the participants had difficulty in more structurally complex analyses even in conditions in which non-syntactic information such as lexical and contextual constraints facilitated the analyses. In their first experiment, sentences with embedded reduced RCs induced processing difficulties, which were enough to change the participants’ preferences in the MV/RR ambiguity. In other words, the participants preferred syntactically simpler analyses, although non-syntactic factors conflicted with such analyses. In this regard, they provided evidence for the effect of syntactic structural biases. In addition, they suggested that syntactic biases could take priority over lexical-level constraints even when such constraints are closely associated with the syntactic structural analysis (Frazier, 1979, 1987; Gorrell, 1995; Grodner et al., 2002).

Traxler (2002) examined self-paced reading and found that structural factors may take priority over semantic or pragmatic plausibility in children’s sentence processing. He compared
the processing of temporary ambiguities by L1 speakers of English between children and adults to determine whether they would respond to the semantic plausibility of the misanalysis and subcategory information in the same way during the resolution of ambiguities. For this, verb subcategorization information and plausibility were manipulated with sentences such as When Sue tripped(,) the girl/table fell over and the vase was broken. Like adults, young participants (8 to 12 years old) were able to use commas (an overt structural marker) and were sensitive to plausibility information. However, they tended to underuse subcategorization information by misanalysing sentences regardless of the verb subcategorization information of the initial verb. According to the results of a series of correlation analyses, however, subcategorization information had considerable influence on the level of difficulty for critical nouns and MVs. In addition, for young participants, plausibility effects were found before disambiguating information. Their reading times increased when they encountered an implausible direct object of the embedded RCV, whereas their reading times decreased for a plausible object even after syntactic disambiguating information triggered reanalysis. There was no difference in the reading time after syntactic disambiguation information regardless of the plausibility condition. Younger participants were considerably more likely to interpret the post-verbal nouns the girl and the table as the direct object of the verb regardless of plausibility. These results illustrate that young L1 speakers are more likely to use structural information than non-syntactic information in sentence processing.

Traxler (2005) examined the use of subcategorization information and plausibility in a self-paced reading study with the following sentences:

(8a) When Susan fell(,) the policeman stopped and picked her up. (intransitive verb condition)
(8b) *When Susan tripped, the table crashed to the ground.* ((in)transitive verb + implausible direct object condition)

(8c) *When Susan tripped, the policeman stopped and picked her up.* ((in)transitive verb + plausible direct object condition)

The participants spent more time reading the ambiguous NPs underlined above in the implausible condition regardless of the subcategorization type of verb. In addition, reading times were longer for MVs, which served as a syntactic disambiguating cue. The participants showed a strong tendency to make direct-object interpretations during the initial parsing regardless of verb subcategorization or plausibility information. Like subcategorization information, plausibility information played a role only in the recovery process. Neither subcategorization information nor plausibility information could prevent the participants from a syntactic misanalysis, but it determined the level of processing difficulty caused by recovering from the initial syntactic misanalysis. These results provide support for serial models in that only syntactic information is considered in the initial structure building.

On the other hand, parallel multiple constraint-based models differ from syntax-first serial models in that they are typically interactive and parallel with multiple sources of information in all processing steps as soon as they become available. In parallel multiple constraint-based models, multiple sources of information are used simultaneously to arrive at a particular interpretation from the initial structural decision (MacDonald et al., 1994). Immediate references to all available information (context, frequency, discourse, and syntactic information) are supposed to be made during language comprehension (Romeo, 2006).

In the case of resolving structural ambiguities, each information constraint such as structural preferences, lexical information, probabilistic patterns and discourse information
competes with all other constraints simultaneously as sentence structures are built through an instantaneous combination of syntactic and non-syntactic information. The decision is made based on the relative strength of each constraint. For instance, syntactic preferences such as the MA preference for the simplest structure can be considered just one of many constraints available and competing with other non-syntactic preferences during sentence processing. However, when other non-syntactic information constraints are not informative enough, this syntactic preference may appear more clearly in the form of a conventional GP effect (Kim, 2004). According to this approach, ambiguities can be resolved in one large step. As a result, however, the WM or long-term memory system can be overloaded with increased cognitive processing required for holding all the activated information and structures.

Although specific details of parsing procedures vary according to the specific model, there are several different versions of parallel multiple constraint-based models, including the Parallel Processing Hypothesis (Fodor, Bever & Garrett, 1974), the Unrestricted Parallel Model (Marslen-Wilson & Welsh, 1975), the Interaction/Connectionist Model (McClelland, 1987), the Referential Momentary Parallel model (Altmann & Steedman, 1988), the Interactive Activation Model (MacDonald et al., 1994), the Constraint-based Model (McRae et al., 1998; Trueswell et al., 1994), the Competition Model Hypothesis (MacWhinney, 1989), and the Lexically-driven Model (Bresnan & Kaplan, 1982; Pritchett, 1988).

Gorell (1989) found evidence supporting the parallel processing view in lexical decision tasks. To investigate the role of the syntactic context in sentence parsing, he employed three types of verbs/syntactic contexts for lexical decision targets with two ambiguity conditions (NPs vs. sentential complements (SC) with balanced ambiguities and MVs vs. past participles with
less balanced ambiguities). The three types of contexts included a structurally simple context, an ambiguous context, and a structurally complex context for each ambiguity condition.

(9a) NP/SC Ambiguity

\textit{It's obvious that Holmes saved the son of the banker/right away.} (simple context)

\textit{It's obvious that Holmes suspected the son of the banker/(right away/was guilty).} (ambiguous context)

\textit{It's obvious that Holmes realized the son of the banker/was guilty.} (complex context)

(9b) MV/Past participle Ambiguity

\textit{The company was loaned money at low rates/to ensure high volume.} (simple context)

\textit{The company loaned money at low rates/(to ensure high volume/decided to begin expanding).} (ambiguous context)

\textit{The company they loaned money at low rates/decided to begin expanding.} (complex context)

For each ambiguity condition, the simple context was presumed to provide preferred structural reading to resolve ambiguities based on the MA principle. However, the participants took longer to read for the verb target in the simple context than in the other ambiguous and complex contexts. Gorrell found that the participants conducted proper structural analyses of ambiguous input strings even before they encountered the verb target.

Syntax-first serial models are simple and economical, but the GP theory has become controversial with experimental data favoring parallel multiple constraint-based models based on the proper management of experimental materials (Kim, 2004).

For instance, Taraban and McClelland (1988) and McClelland (1987) questioned Rayner et al.’s (1983) findings, claiming that their findings may be due to the inappropriate manipulation
of semantic constraints with respect to syntactic ones, and suggested that Rayner et al.’s study employed materials that were semantically biased toward minimal completion (the advantage of the VP completion over the NP completion), not toward the syntactic structural preference.

Taraban et al. (1988) conducted a rating study and found that the final nouns in (10a) were more likely to be expected with respect to the verb than those in (10b).

(10a) *The spy saw the cop with binoculars.* (*binocular* as an instrument)

(10b) *The spy saw the cop with a revolver.* (*revolver* as a possession of the cop)

This semantic bias had considerable influence on the final interpretation of sentences, but McClelland (1987) wondered whether it could influence the initial parsing decision early enough. McClelland argued that GP effects in English could be easily exaggerated or diminished if non-syntactic information biases in experimental materials are not carefully controlled for. In this case, employing materials with non-minimal attachment preferences or with frequency and plausibility controlled for may be a better control method.

Trueswell et al. (1994) also questioned the plausibility manipulation (animacy) of Ferreira and Clifton’s (1986) items.

(11a) *The defendant examined by the lawyer turned out to be unreliable.*

(11b) *The evidence examined by the lawyer turned out to be unreliable.*

They argued that half the inanimate items were not plausible agents. Thus, they replicated Ferreira and Clifton’s (1986) experiment by using updated rating information (agents vs. patients/themes) for inanimate nouns and including morphologically unambiguous verbs. In addition, they employed a different display to present all the scoring positions placed on one line.
Ambiguous Verbs (animate vs. inanimate noun)

The prisoner/gold (that was) transported by the guards was closely watched. (1.1, 5.5)

Unambiguous Verbs (inanimate noun only)

The money (that was) taken by the student was finally returned. (1.2, 5.6)

(cf. mean agent and patient/theme ratings for inanimate nouns)

They analyzed eye movements by using improved items and found that thematic relationships based on animacy and lexical frequency information were immediately used in resolving MV/reduced RCs ambiguities, which was predicted by parallel multiple constraint-based models. In addition, it could prevent GP effects if the previous context provided support for the non-minimally attached analysis.

Through properly balanced control of semantic and syntactic constraints with direct object vs. sentential complement (DO/SC) ambiguities, Garnsey et al. (1997) conducted eye-tracking and self-paced reading analyses and found that both frequency-based verb biases and the thematic plausibility of particular word combinations contributed to the online processing of temporarily ambiguous sentences. In the analyses, they considered three types of verb biases: the DO bias, the SC bias, and the equi-bias (EQ bias). Some of the sentences they used are as follows:

(13a) The talented photographer accepted (that) the money/fire could not be … (DO bias condition)

(13b) The sales clerk acknowledged (that) the error/shirt should have been … (EQ bias condition)

(13c) The ticket agent admitted (that) the mistake/airplane had been … (SC bias condition)
Garnsey et al. found some interaction between the verb bias and plausibility information in the comprehension of temporary ambiguous sentences. The effect of the verb bias did not entail significant interactions with plausibility information in the DO bias and SC bias conditions. They found that the information on the verb bias led to initial structure building as shown by overriding the DO interpretation of NPs in the SC bias condition. On the other hand, they found a plausibility effect in the disambiguating region only in the EQ bias condition. In other words, the plausibility bias had measurable effects only when verbs were not biased. However, they suggested a need for further research for generalizing the interaction patterns of the effects of verb biases and plausibility. In addition, their manipulation of the plausibility effect and findings are not clear in that the strong effect of the verb bias may reflect the fact that the ambiguous region was short and verbs preceded ambiguous NPs in their sentences.

In this section, two leading models of the ambiguity resolution process have thus far been presented with some conflicting research findings. Some studies have provided support for serial models, whereas others provide evidence that non-syntactic information can contribute to initial sentence processing (Gibson & Pearlmutter, 1998). Thus, although a number of studies have examined in the ambiguity resolution process, it remains unclear which model better explains human language processing. In addition, individual differences in the cognitive resource capacity may influence initial parsing decisions.

In the last three decades, these two classes of models have increasingly become complementary because they have been modified to account for new experimental evidence (Lewis, 2000). For example, Jackendoff (2007) argued that syntax-first serial models are obliged to either adopt “lexically conditioned alternatives that imply either some degree of parallelism” (p. 12) or allow for an incredibly speedy recovery from an erroneous first analysis to account for
all the data. On the other hand, parallel multiple constraint-based models are forced to devise principled criteria for ranking alternatives. One of the main reasons for this can be found in research methods for controlling for the degree of the bias associated with experimental materials or conditions. Because sentence processing occurs extremely fast and is thus hard to detect, we may not detect the whole process or may end up with results from less-balanced experimental materials. As a result, there is a need for well-balanced experimental materials for determining which model better explains sentence processing.

2.1.2. Good-Enough/Shallow Parsing in L1

If we set aside the timing issue associated with the reanalysis or revision process of resolving ambiguities, both models assume that individuals engaged in incremental parsing eventually end up with a complete and correct interpretation based on appropriate structure building with or without a revision or reanalysis when they encounter disambiguating cues during their resolution of ambiguities. Adult native speakers are able to use syntactic as well as non-syntactic information appropriately for either communication or task completion purposes (Indefrey, 2006).

There were no empirical studies testing this assumption until several L1 studies have recently measured the interpretation of experimental sentences instead of focusing on in the online ambiguity resolution process and the use of different types of information. These studies have found that even native speakers are often satisfied with an incomplete (and possibly incorrect) structural analysis (i.e., shallow parsing) under certain circumstances such as time pressure, noise, complicated syntactic structures, task demands, a low level of WM capacity, emotional stress, and unsound physical conditions. This implies that individuals sometimes engage in shallow parsing even in their native language (Christianson et al., 2001; Christianson
et al., 2010; Christianson et al., 2006; Ferreira, 2003; Ferreira et al., 2002; Sanford & Sturt, 2002; Swets et al., 2008; Sturt, 2007). However, both of the sentence-processing models considered in the present study focus on the synchronous processing itself but pay little attention to the final interpretation of the reader during or after in the ambiguity resolution process (Fernandez, 2006; Ferreira, 2003). Some native speakers keep some remaining interpretation from their initial analysis, although it may turn out to be wrong or incompatible. In contrast to both of these models, good-enough (GE) sentence processing models (backed by evidence of shallow parsing among L1 speakers) highlight the importance of integrating non-syntactic information into the parsing process as well as the syntactic analysis.

For example, Christianson et al. (2001) found evidence of a partial reanalysis of GP sentences in L1 English, which may be due to some lingering initial misanalysis.

(14) While Anna bathed the baby spit up on the bed.

(14a) Did Anna bathe the baby?

(14b) Did the baby spit up on the bed?

When the participants were asked forced-choice (yes/no) questions such as (14a) and (14b), they were significantly more likely to answer “yes” to both questions. They answered “yes” incorrectly approximately 60% of the time for (14a), whereas they correctly answered “yes” approximately 90% of the time for (14b). This indicates that the participants were likely to interpret the baby as the direct object of the word bathed, which is not grammatically correct, while simultaneously assigning it as the subject of the subsequent MV spit up. However, the forced-choice question used in the study has been questioned because it might have influenced the participants to retain their initial misanalysis (van Gompel, Pickering, Pearson & Jacob, 2006).
In addition, van Gompel et al. (2006) investigated the lingering misinterpretation of misanalyzed transitive structures by using structural priming. Using sentences such as those shown below (15a and 15b as prime sentences and 15c as target fragments), they found that the participants were more likely to provide a transitive structure interpretation with temporary ambiguities in (15a) than in (15b) with a disambiguating cue.

(15a) While the man was visiting the children who were surprisingly pleasant and funny played outside.

(15b) While the man was visiting, the children who were surprisingly pleasant and funny played outside.

(15c) When the doctor was visiting...

Although they did not provide a final interpretation of the sentence directly, they claimed that the initially misanalyzed transitive structure continued to have considerable influence on interpretation. The findings of these two studies suggest that individuals do not easily overwrite the initial structure they project, particularly when it is plausible.

Some researchers have proposed that individuals make use of shallow parsing to reduce the amount of effort, conserve cognitive resources, or save time so that they may process complicated sentences faster and more efficiently. In many cases, without resorting to all options, the reader can achieve efficiency and save time by having some fined list of closely related structural options and choosing one of them. Such processing strategies are called heuristics, which are experience-based and efficient. For language processing, Ferreria (2003) suggested several commonly applied heuristics such as NVN (Noun-Verb-Noun), plausibility, and animacy. The NVN heuristics is heuristics based on the order of words and take an NVN sequence as Subject (agent) - Verb (action) - Object (patient/theme) (Ferreira, 2003; Townsend
It is similar to the MA principle in that both take the second object noun as the direct object of the preceding verb in principle when processing the NVN sequence in English. This sequence is particularly common in English and was one of the main heuristics used in sentence processing in Ferreria’s (2003) correlation analysis. The next strongest heuristics is thematic processing strategies based on plausibility and animacy. These two strategies are closely related and belong to the same semantic domain.

Ferreira (2003) conducted three experiments with unambiguous active/passive and subject-cleft/object-cleft sentences to test two different but parallel parsing routes: heuristic and algorithmic parsing. Heuristic routes employ simplified strategies to understand sentences, whereas algorithmic parsing routes build detailed syntactic structures. The participants listened to sentences and stated aloud the thematic roles in the sentences, such as the agent and the patient. The accuracy of the identification of thematic information and decision times were analyzed. Ferreira found that passives with implausible thematic roles (e.g., a mouse chasing a cat) were misinterpreted more often and that subject-clefs, which are fairly atypical, were processed as easily and accurately as active sentences. Based on these findings, Ferreira suggested the preference for the NVN strategy over detailed syntactic structure building and proposed the existence of simple and shallow heuristic processing as well as algorithm-based syntactic processing in L1 language comprehension. This shallow processing, named “good-enough” processing by Ferreira (2001, 2003), may frequently override detailed syntactic processing. However, Ferreira noted a need for further research to elucidate the relative strength of various potential heuristics and the coordination of heuristic and algorithmic parsing streams.

their study, 72 native speakers (adults) of English read sentences and answered comprehension questions about the thematic roles (agent vs. patient) of plausible and implausible passive/active sentences (the same ones in Ferreira (2003). Listening comprehension was followed by a production task involving picture description. In general, their comprehension findings provide support for Ferreira (2003).

However, in terms of production data, they found an interesting result: Plausible passives and implausible actives were more likely to prime passive structures in oral descriptions. This structural priming effect on subsequent production and its interaction with plausibility were not related to the type of question used to probe comprehension. They suggested that this was clear evidence for GE language processing (Ferreira et al., 2002), which hypothesizes that underspecified final representations result from language processing under certain circumstances. Based on their findings, semantic plausibility may have a stronger effect on the final interpretation and structural priming when there is conflicting information from two separate processing routes, i.e., morphosyntactic and semantic processing routes (Christianson et al., 2010).

However, as noted by Christianson et al. (2010), their findings provide no information on the time course and processing load of conflicting information sources for initial processing because the main focus of their study was on the final interpretation and structural priming for subsequent production. In this regard, Ferreira (2003) and Christianson et al. (2010) provided similar findings despite using different questioning methods.

Thus, GE parsing heuristics may represent a hybrid of the two major models of sentence processing: the MA principle and the plausibility constraint (which focuses on the interpretation process). In this regard, determining whether both types of information or constraints are used
from the initial parsing stage as well as the final interpretation expected by a syntactically precise structural analysis, depending on the relative strength of each constraint should provide important insights into sentence processing.

In an ERP study of semantic anomalies involving conflicting thematic interpretations between syntactic parsing and plausibility heuristics, such as *The fox that hunted the poacher*, Vissers, Chwilla, and Kolk (2007) proposed parallel but independent syntactic and semantic parsing mechanisms and suggested that plausibility heuristics can project the basic structure of a sentence only by its thematic interpretation without resorting to syntactic information, which is consistent with the findings of Christianson et al. (2010).

Townsend and Bever (2001) proposed two parsing routes based on their late assignment of syntax theory (LAST): pseudo-syntax first and full syntactic analysis later. Pseudo-syntax refers to building an initial meaning form hypothesis based upon semantic and statistical information such as possible thematic roles. For instance, local phrases are first built using function-word templates or canonical sentence patterns such as the NVN pattern. Then detailed syntactic information such as grammatical derivations is employed to build a complete syntactic structure that generates another representation of the meaning. Here in the second stage, the preliminary analysis generated by the pseudo-syntax limits the number of structural options for the real syntax. Then the output of this grammatical derivation is compared with the linguistic input. If these two do not match, then reanalysis (including another pseudo-syntax analysis and the subsequent grammatical derivation) is followed for successful comprehension.

Sanford and Sturt (2002) offered evidence of underspecified processing and representations in L1 language comprehension. They found that college students employ simpler strategies to make reasonable speculations instead of performing full syntactic parsing. Hence,
they suggested that language processing and the final interpretation during comprehension are not always complete or incremental as many researchers generally assume.

In the same vein, Clahsen and Felser (2006a) proposed the so-called “shallow structure hypothesis” (SSH), a model of bilingual processing. The SSH posits that even advanced L2 learners project a shallow syntactic structure based on simple syntactic structure information and non-syntactic information such as lexical and semantic information. In their comparative study of sentence processing by L1 speakers (children and adults) and L2 learners (adults), they hypothesized that adult L1 speakers theoretically have two available parsing processes for language comprehension. The first one is a semantically mediated shallow processing route derived from fairly simple surface structure information; morphological rules; statistical patterns; and lexical, semantic, and pragmatic information. The second one is a full parsing route based upon a fully specified syntactic analysis for more accurate and complete processing. Native speakers employed both these two parsing routes properly for their purposes. On the other hand, language processing by adult L2 learners was fundamentally different from that by native speakers in that L2 learners’ language processing lacked a fully specified syntactic analysis, and this lack of detailed syntactic information was likely to be supplemented by their heavy use of lexical-semantic cues during sentence processing (Clahsen & Felser, 2006a, b). According to the SSH, L2 learners should make increased use of non-syntactic information to understand the L2 linguistic input because their syntactic knowledge and ability to deploy syntactic knowledge incrementally in real-time processing are not fully established.

This dependence on non-syntactic information during sentence parsing is consistent with the GE parsing principle in L1 sentence processing. In other words, shallow and partial L1 (Ferreira, 2003, 2007) and L2 (Clashen & Felser, 2006a, b) sentence processing place great
emphasis on the effective use of non-syntactic information for the resolution of syntactic ambiguities. An important difference is that according to the SSH, full syntactic parsing is not available to L2 learners, whereas it is for native speakers, and thus, there is a qualitative difference in sentence processing between the two groups. However, the GE framework expects these two groups to be quantitatively different, but qualitatively more or less similar processing behaviors based on the input, the task at hand, and individual characteristics of speakers (e.g., WM and L2 proficiency) (for a full discussion, see Lim, 2010).

Why do individuals often tolerate or disregard incomplete processing and thus incorrect interpretations? If the reader intentionally does not attend to the exact meaning of each sentence or does not want to spend a lot of time and effort for their daily communication, then the issue is more a selective control problem associated with human psychology and behavior. However, given that at least some part of GE or shallow parsing may result from cognitive resource constraints during online linguistic processing without any conscious decision or effort, there remains the question of which factor is responsible for individual variations in L1 sentence processing.

In the following section, I provide a review of previous research on the role of plausibility information in the resolution of ambiguities in L1 sentence processing.

2.1.3. Use of information on semantic plausibility in L1 sentence processing

Among the multiple sources of information available in sentence processing, syntactic structural information is considered the most influential one, particularly in the initial stage of sentence processing. On the other hand, there are diverse types of non-syntactic information that can contribute to the ambiguity resolution process, including animacy, case markers, the word order, thematic information, the verb bias, frequency information, and referential information.
Some studies have found that non-syntactic information such as plausibility may play an important role in the early stage of sentence processing, building an initial surface structure in conjunction with syntactic information (Patson, Darowski, Moon & Ferreira, 2009; Patson, Swensen, Moon & Ferreira, 2006; Vissers et al., 2007).

Studies of sentence processing by L1 speakers of English have found that individuals are likely to have more difficulty recovering from a plausible interpretation than from an implausible one during the reanalysis. For instance, Traxler and Pickering (1996) provided evidence of the plausibility effect in L1 sentence processing by using the following sentences:

(16a) We like the book that the author wrote ___ unceasingly and with great dedication about (real gap) while waiting for a contract. (write a book – a plausible direct object interpretation)

(16b) We like the city that the author wrote ___ unceasingly and with great dedication about (real gap) while waiting for a contract. (write a city – an implausible direct object interpretation)

After verb wrote, there is a gap in both sentences. Based on the filler-driven strategy, Frazier and Clifton (1989) found that native speakers of English showed longer reading times at the gap in the second sentence, which indicates that they were likely to calculate not only the structural filler (taking the previous NP, the book or the city, as the direct object of the verb) but also the plausibility of the filler because. However, at the relevant gap at the region about while, reading times for the revision of the initial analysis were longer for the first sentence than for the second sentence. This suggests that the more plausible the initial analysis, the harder it is to reanalyze, which can be interpreted as a cross-over pattern of plausibility effects, that is, longer reading times at the verb because of the implausibility of the direct-object interpretation but
shorter reading times at the real gap because of the previous expectation of another real direct object. Pickering and Traxler (1998, 2003) found similar plausibility effects.

Van Gompel and Pickering’s (1998) eye-tracking study with subject/object ambiguities demonstrated that verb subcategorization information as well as plausibility information could facilitate the reanalysis.

(17a) After the dog struggled the vet and his new assistant took off the muzzle.

(17b) After the dog scratched the vet and his new assistant took off the muzzle.

They found that the participants had more difficulty processing the ambiguous phrase the vet and his new assistant in the first sentence (clear intransitive verb subcategorization information) than in the second sentence (obvious transitive subcategorization information). In addition, the participants had more difficulty reanalyzing the plausible misanalysis when they encountered the disambiguating word. However, these findings shed no light on the role of plausibility information in initial parsing.

Matzen and Garnsey (2006) conducted an event-related potential (ERP) study to investigate the effects of noun plausibility in English by using (18a) and (18b) and found that native speakers of English used noun plausibility information as a disambiguating cue for temporarily ambiguous verbs.

(18a) The young boy admitted the stone had broken the window. (implausible direct object condition)

(18b) The young boy admitted the lie had been his idea. (plausible direct object condition)

In the implausible DO condition, large N400s (a sign of semantic anomaly detection) were found at the subsequent nouns. On the other hand, in the plausible DO condition, P600s (a
sign of syntactic reanalysis or integration) were found at the disambiguating words (subordinate clause verbs). Because there was no P600 effect at the disambiguating word in the implausible DO condition, Matzen and Garnsey claimed that plausibility information was used early enough to construct a proper sentence structure, which prevented possible GP effects in the later stages of reading.

As discussed in the previous section, Christianson et al. (2010) presented further evidence of a plausibility effect on structural priming as well as the final interpretation of sentence comprehension for adult L1 speakers of English. In the next section, a review of previous research on WM in L1 sentence processing with respect to plausibility information is provided.

**2.1.4. Working Memory in L1 sentence processing**

Cognitive resources such as working memory (WM) have been assumed to influence the process of resolving temporary structural ambiguities because the capacity of such resources is expected to determine how much information can be stored and used during this process. For instance, some researchers have indicated that individuals with a high level of WM capacity are good at considering multiple structural alternatives and can make better use of this information during the ambiguity resolution process, whereas those with a low level of WM capacity may activate only one of the structural alternatives and engage in less efficient processing (MacDonald, Just & Carpenter, 1992). On the other hand, other researchers have suggested that individuals with a high level of WM capacity tend to choose a single preferred structural option in the case of structural ambiguities early enough for more efficient processing, whereas those with a low level of WM capacity tend to consume their WM resources faster by considering multiple structural alternatives (Friederici, Steinhauer & Mecklinger, 1998). This view is consistent with syntax-first serial models in that readers are likely to project the simplest
structure based on any available linguistic input regardless of differences in their WM capacity. In addition, there is no clear consensus on the definition of or measurement methods for WM. WM is a complicated concept, and thus, it is difficult to come up with a short definition. Further, diverse aspects of WM systems have been discussed and emphasized separately by a number of researchers. WM is one of the central components of cognitive resources, and in general, WM is thought to be short-term and capacity-limited and can be subdivided into multiple components to keep information active for sentence processing. An individual’s WM capacity is his or her cognitive capability to keep information activated and available for ordinary activities despite competition for limited attention (Sanders, 2004). Multiple factors such as an individual’s age, experience, and language proficiency are likely to influence the individual’s WM capacity. WM temporarily holds the meaning and mental representation of linguistic expressions in language tasks. Like other complex cognitive activities, language comprehension involves the coordination of the human memory system for storing and processing linguistic information. Accordingly, various models of language comprehension offer differing views on the role of WM in sentence processing.

In fact, major sentence-processing theories are influenced by the effect of WM capacity on sentence processing. Syntax-first serial models (e.g., GP models) (Frazier, 1979) propose one simple structural analysis to guide initial sentence processing (e.g., the MA principle) because of WM constraints. Only syntactic information is applied to initial structure building for minimizing the WM burden. Non-syntactic information such as plausibility is hypothesized to be applied in the second stage of parsing for confirming the initial syntactic analysis.

On the other hand, parallel multiple constraint-based models posit that readers can hold multiple structural alternatives in their WM and propose that individual differences in WM
capacity can explain individual variations in parsing performance. For example, individuals with a high level of WM capacity are likely to hold more applicable constraints such as plausibility information in addition to syntactic information for processing the linguistic input in the early parsing stage. Hence, those with a high level of WM capacity tend to perform more efficient and accurate sentence processing than those with a low level of WM capacity (e.g., Just & Carpenter, 1992; MacDonald et al., 1992; Pearlmutter & MacDonald, 1995).

In contrast, the GE approach presumes that language comprehension mechanism does not always lead to complete and perfect full syntactic structures but provides “good-enough” representations for sentences to conserve cognitive resources. However, the GE approach does not directly specify the role of WM capacity in sentence processing. Replicating Christianson et al.’s (2001) study by considering a sample of young and older adults, Christianson, Williams, Zacks, and Ferreira (2006) found evidence of GE processing for both age groups. Although both groups showed similar GP effects for comprehension questions, older adults tended to respond with more transitive interpretations with optionally transitive verbs such as hunt, chew, and read, and those with a low level of WM capacity were less likely to succeed in reactivating syntactic structures for reflexive reading. Christianson et al. suggested that older adults’ increased reliance on inferences may be driven by their need to compensate for the age-related decrease in their WM capacity.

One of the most well-known and widely-used WM models has been proposed by Baddeley and colleagues (Baddeley, 2003; Baddeley & Hitch, 1974). This multiple-component model consists of four segments: the phonological/articulatory loop, the visual-spatial sketchpad, the central executive, and the episodic buffer. The episodic buffer has recently been added to Baddeley’s (2000, 2003) model to address short-term operations through communication with
episodic long-term memory. The visuo-spatial sketchpad facilitates visual and spatial information retention and processing. Both the phonological loop and the central executive have been widely investigated in the psycholinguistic field for language processing and acquisition. The central executive distributes information processing and cognitive resources to perform a given task and is known to play an important role in language comprehension (Ardila, 2003; Marton & Schwartz, 2003).

If Baddeley’s model defines WM from a structural point of view, then Cowan’s (1988, 1995) Embedded-Process model describes WM in terms of its function. Cowan (1988, 1995, 1999) suggested that WM is merely some part of long-term memory and that it actively processes incoming information based on time and capacity limitations. His model proposes separate components for passive information storage and processing by differentiating memory activation from attention and highlights the control of attention with respect to WM. Both activation and attentional focus are necessary for linguistic information processing.

However, specific attributes of WM with respect to sentence processing remain unclear (Just, Carpenter & Woolley, 1982; MacDonald & Christiansen, 2002; Waters & Caplan, 1996). For instance, WM models disagree on when and how WM influences parsing, that is, whether its influence starts from initial parsing decisions or occurs in “post-interpretive processes” (Patson et al., 2006). In terms of cognitive resources that can refer to both the information-processing structure and capacity, there are two approaches to sentence comprehension: single-resource (SR) theory (Just & Carpenter, 1992; King & Just, 1991; MacDonald et al., 1992) and separate language interpretation resource (SLIR) theory (Caplan & Waters, 1999). A key question regarding WM and sentence comprehension is whether there is a special type of cognitive resource exclusively for language processing.
According to the SR theory, cognitive systems tend to be domain-general processes, and thus, cognitive resources such as WM are employed for various cognitive functions, including language processing. There is one general and shared WM resource pool for sentence comprehension involving both linguistic processing and general task-related executive processing (Just & Carpenter, 1992; King & Just, 1991; MacDonald et al., 1992). Because cognitive resources are limited, if any sub-component of the WM system is overloaded with a given task, then it may drain the shared cognitive resources that can otherwise allocated to other subcomponents. For instance, when language users perform a linguistic task and the task itself demands a high level of WM capacity, sentence processing takes longer or generates more errors.

In addition, individual differences in WM capacity may account for individual variations in the processing strategy (Hummel, 2002). Some studies have provided support for the model of shared resource pools (Just & Carpenter, 1992; King & Just, 1991) by presenting evidence based on the diverse performance of readers with different WM spans through the use of Daneman and Carpenter’s (1980) reading span test (RST). According to the SR theory, individuals with a low level of WM capacity tend to have fewer cognitive resources for sentence processing and thus show less effective non-linguistic processing.

On the other hand, the SLIR theory hypothesizes a separate and independent verbal WM resource system for online linguistic processing (Caplan & Waters, 1999; Waters & Caplan, 2001). It distinguishes between interpretation and post-interpretation processes in sentence processing. The process of interpreting a linguistic input is instant and specialized and is managed by a separate verbal WM resource system instead of drawing on generalized cognitive resources. As a result, interpretive syntactic processing is not likely to be influenced by
individual differences in WM capacity (Waters & Caplan, 1996; Waters & Caplan, 2001). According to this model, individuals with both high and low WM spans process linguistic inputs in a similar way, that is, their processing differences stem from the demand of the post-interpretation process such as answering comprehension questions for meanings or retaining observed content for recall.

Caplan and Waters (1999) and Rochon, Waters, and Caplan (2000) employed data from aphasic patients and provided support for this model. They found that aphasic patients had no problem with interpretive processing (including phonological and syntactic processing) for understanding the linguistic input, whereas they had difficulty mapping the meaning onto propositions in long-term memory, that is, post-interpretive processing. In addition, in their replication of MacDonald et al. (1992) and other similar studies with different types of GP sentences, Waters and Caplan (1996) found no difference between WM groups in terms of reading times or accuracy.

A different WM perspective proposed by Engle (2002) emphasizes the processing side of WM, that is, it emphasizes the control of attention to a given task/goal and the avoidance of proactive interference in the performance of a task. According to this perspective, individuals with a high level of WM capacity tend to concentrate better on competing tasks with less interference, whereas those with a low level of WM capacity are more likely to be easily distracted.

MacDonald and Christiansen (2002) and Long and Prat (2008) provided an experience-based analysis of WM effects in language processing and suggested that the individual’s experience and skill play important roles in language processing. MacDonald and Christiansen (2002) suggested that there is no theoretical rationale for WM and claimed that WM test scores
simply reflect individuals’ language experience and that capacity differences are dependent on the amount of language experience. Hence, individuals with a low level of WM capacity are likely to have more processing difficulty with less frequent words or phrases.

Roberts and Gibson (2002) suggested that WM is only some part of the memory system and is not directly associated with measures such as reading comprehension. Previous studies investigating L1 sentence processing by using WM capacity for sentence comprehension tasks have produced mixed results for the WM effect on syntactic processing. A large portion of this discrepancy is often attributed to diverse methods and experimental materials employed in studies.

Some studies have found no WM effect even for complicated syntactic structures (Long & Prat, 2008; MacDonald & Christiansen, 2002). In general, L1 syntactic processing is not closely related to WM differences in terms of speed and accuracy (Waters & Caplan, 2001; Juffs, in press). However, there is some evidence of the WM effect on a number of propositions. Individuals with a low level of WM capacity tend to have more difficulty recalling the propositional content of a sentence than those with a high level of WM capacity (Juffs, in press).

Clifton, Traxler, Mohamed, Williams, Morris, and Rayner (2003) measured eye movements by using the following sentences and found no WM effect on syntactic disambiguation with reduced RC ambiguities:

(20a) *The evidence (that was) examined by the lawyer turned out to be incorrect.*

(inanimate subject + past participle condition)

(20b) *The defendant (that was) examined by the lawyer turned out to be incorrect.*

(animate subject + past tense condition)
The participants showed the same processing patterns regardless of their WM capacity. In other words, they experienced the disruption of processing when they encountered syntactically disambiguating materials in both sentences regardless of their WM capacity. In addition, in contrast to Trueswell et al. (1994), Clifton et al. (2003) provided support for syntax-first serial models.

On the other hand, other researchers have shown meaningful effects of WM capacity on sentence processing, although their findings show some inconsistency across various research methods. For instance, Just and Carpenter (1992) found a positive correlation between the RST of verbal WM and language comprehension ability and proficiency in reading.

King and Just (1991), the first to examine the effect of the WM span, found differences in reading times and accuracy between high- and low-span readers for object-relative sentences such as *The senator that the reporter attacked admitted the error*. Low-span readers had more difficulty processing object-relative sentences because of their insufficient processing capacity. King and Just (1991) showed that low-span native speakers took longer for complex regions than their high-span counterparts.

Replicating Ferreira and Clifton (1986) by using MV and RR ambiguity constructions (see (4), Just and Carpenter (1992) argued that individual differences in WM capacity could account for the reader’s difficulty resolving ambiguities in the disambiguating region. For instance, low-span readers are not likely to make use of animacy information in the reduced RC (RRC), whereas high-span readers take less time reading the inanimate subject NP than reading the animate subject NP.

(21) *The experienced soldiers warned about the dangers before the midnight raid (MV).*

*/conducted the midnight raid (RRC).*
Just and Carpenter determined that reading times at the end of the sentence were longer for high-span readers for MVs than for low-span readers and found no WM effect on syntactic information processing. Regardless of animacy information in the first noun, both groups of readers spent more time on reduced RCs (i.e., a less preferred structure). However, high-span readers showed some evidence of active interactions between syntax and semantics, whereas low-span readers were insensitive to the effect of semantics on initial syntactic parsing decisions.

On the other hand, in a self-paced reading study using GP sentences in English with temporary main vs. reduced RCV ambiguities, MacDonald et al. (1992) found somewhat surprising WM differences in reading times and accuracy for the comprehension questions. Regardless of the preference the type of verb, high-span readers spent more time reading ambiguous sentences. MacDonald et al. suggested that this difference might be due to the processing cost of holding all structural alternatives up to the disambiguating cue.

Previous studies have provided some evidence of the effect of structure complexity on sentence processing with respect to WM (Grodner, Gibson & Tunstall, 2002). Recently, neuroimaging studies using positron emission tomography (PET) or functional magnetic resonance imaging (fMRI) have provided evidence of the effect of structural complex on neural activation (Fiebach, Vos & Friederici, 2004). In addition, WM has been suggested to play an important role in this effect (Fiebach et al., 2004).

Plausibility information is thought to be influenced by cognitive resource constraints to a certain degree. Previous L1 research has suggested that readers with a high level of WM capacity may be more sensitive to semantic or plausibility cues than those with a low level of WM capacity (King & Just, 1991; Pearlmutter & MacDonald, 1995). Pearlmutter and MacDonald (1995) found WM differences between these two groups of readers in their use of cues with main
vs. reduced RC interpretations. They investigated reading times for MV/reduced RC ambiguities with subsequent disambiguating PPs such as *The soup cooked in the pot but was not ready to eat.* Consistent with the findings of MacDonald et al. (1992), high-span readers spent more time reading PPs. In contrast, low-span readers did not use probabilistic frequency information or contextual constraints during the online reading process. Pearlmutter and MacDonald employed tasks for rating plausibility to determine whether the participants would use plausibility information for their interpretations and found that both high- and low-span readers made effective use of plausibility cues. However, according to a series of regression analyses, individual variations in the participants’ sensitivity to plausibility information on potential alternatives led to differences in reading times for online processing between the two groups of readers. Low-span readers were less likely to be sensitive to the plausibility of alternative syntactic analyses in online processing than their high-span counterparts.

Long and Prat (2008) conducted a series of self-paced reading experiments with English-speaking college students to investigate the relationship between syntactic processing and differences in the level of WM capacity. For this, they employed sentences with MV/reduced RCV ambiguities (e.g., *The first question asked/written on the exam was particularly difficult*). All these sentences were expected to be structurally interpreted as reduced RC constructions. The plausibility of the first noun as the object (thematic patient) of the embedded RC verb was manipulated, in addition to the type of verb bias (biased toward the past tense and biased toward the past participle). Individual WM was measured via Danemann and Carpenter’s (1980) RST one day before the main experiment to allow for a counter-balanced arrangement for span scores for two sets of material. There were three versions of the main experiment. In the first trial of experiment 1, to test the effect of WM on the process of resolving syntactic ambiguities, they
employed two groups of participants: one group for reading times and the other group for rating. They considered a sample of 42 college students for reading times and a sample of 180 college students for rating and sentence completion tasks. In terms of reading times, the participants had more difficulty processing sentences biased toward the MV than those biased toward the reduced RCV regardless of their WM spans. However, there was some ambiguity effect for span groups. High-span readers showed longer reading times earlier in the sentence, whereas mid/low-span readers showed longer reading times even after the MV following the PP.

Unlike the first experiment (in which comprehension questions were included only for filler sentences), in the second experiment, all test items were accompanied by paraphrased true/false comprehension questions for checking the participants’ interpretations (e.g., The question was asked on the exam). A grammaticality judgment task was used in the third experiment to test whether the participants could effectively resolve ambiguities. The second and third experiments provided similar results for reading times. High-span readers had difficulty processing sentences biased toward the MV, and all readers showed a verb bias effect at the end of sentences. In sum, there were differences in processing patterns for ambiguous sentences between high- and low-span readers.

Long and Prat (2008) conducted subsequent regression analyses of the effects of different sources of information (e.g., verb frequency, plausibility, and contextual predictability) on reading times and suggested that the effect of plausibility on reading times varies according to WM spans. Both high- and low-span readers were sensitive to structural ambiguities. According to the results of rating and sentence completion tests in the first experiment, low-span readers had the linguistic knowledge to understand the interaction between plausibility and structural ambiguities. However, only high-span readers used plausibility and predictability information
during their online processing, whereas low-span readers made use of only morphological information for sentences biased toward past participles. In contrast, in the second experiment, when low-span readers had repeated exposure to the same test materials, they became sensitive to plausibility information. Based on these findings, Long and Prat (2008) suggested that the influence of WM capacity on sentence parsing depends on the individual’s experience and experimental stimuli.

Patson, Swensen, Moon, and Ferreira (2006) found that WM could influence the reanalysis process but not initial parsing decisions in L1 sentence processing. They examined the effects of WM on the reanalysis process and the final interpretation of GP sentences by considering a paraphrased question paradigm. They employed 24 GP sentences from Christianson et al. (2001) and two verb conditions (optionally transitive verbs vs. reflexive absolute transitive verbs) under two paraphrasing conditions (one paraphrasing the sentence the participant read immediately after the grammaticality judgment task vs. another counting numbers for 30 seconds before paraphrasing). Patson et al. (2006) made use of a paraphrasing paradigm that is considered to be less intrusive than the forced yes/no choice.

(22a) While the man hunted the deer that was brown and graceful ran into the woods.

(optionally transitive verbs)

(22b) While Anna bathed the baby that was cute and cuddly spit up on the bed.

(reflexive verbs)

Paraphrase examples for the reflexive verb sentences above:

Full: The baby spit up on the bed while Anna took a bath.

Partial: Anna bathed the baby and it spit up on the bed.

Failed: Anna bathed the baby.
The participants were divided into two groups based on their WM span. Patson et al.’s (2006) assumption about the effects of WM on sentence processing was based on Hasher and Zachs’s (1988) inhibition theory of WM, which states that high-span individuals can effectively avoid unnecessary information processing in WM to focus their cognitive resources on processing important information. They hypothesized that high-span readers would be more efficient in preventing misinterpretation at the beginning of sentences than low-span readers and thus that high-span readers would end up with a more robust reanalysis for the final interpretation.

According to the results for the proportion of paraphrases to the type of verb in the facilitating and delay conditions, WM had considerable influence on the reanalysis process. Consistent with the findings of Christianson et al. (2001), the participants generally showed the tendency to partially reanalyze GP sentences across the conditions. This implies that initial misinterpretations persisted for some participants across all the conditions. However, high-span readers, particularly those in the delay condition, were much more likely to perform full reanalysis than low-span readers. This means that both groups exhibited GP effects to a similar degree but high-span readers recovered more efficiently from their initial misanalysis than low-span readers. Patson et al. (2006) claimed that these findings provide support for the SLIR (e.g., Waters & Caplan, 1996) because WM influenced the reanalysis process but not initial online syntactic decisions. Although Patson et al. (2006) provided some evidence of the effect of WM on the reanalysis process, it remains unclear whether WM influences the initial online processing because their research questions focused on question-type effects in the reanalysis process. Thus, to address the question of whether WM influences online processing, one needs to apply online processing measures during parsing with respect to WM.
Replicating Patson et al.’s (2006) study, Patson, Darowski, Moon, and Ferreira (2009) provided similar results for the partial reanalysis in L1 (English) sentence processing. Thus, both Patson et al. (2006) and Patson et al. (2009) provided empirical evidence of GP processing.

In sum, in terms of L1 sentence processing, L1 speakers with a high level of WM capacity can read faster and more accurately than those with a low level of WM capacity and are more sensitive to non-syntactic information such as plausibility (King & Just, 1991; Pearlmutter & MacDonald, 1995). However, they sometimes slow down when they encounter temporary ambiguities because they hold more information such as all possible syntactic structures and interpretive alternatives (MacDonald et al., 1992). In addition, there is a correlation between WM and plausibility preferences in L1 sentence processing. Low-span readers are less likely than their high-span counterparts to make use of plausibility information for parsing.

Some aspects of WM, including its components, roles in sentence processing, measurement techniques, remain the subject of much debate (Baddeley, 2000; MacDonald & Christiansen, 2002). In addition, there are diverse tests for the WM span (e.g., the counting span, the operation span, and the reading span) and various scoring techniques for each test. These tests are thought to estimate the reader’s performance in certain types of cognitive activities such as reading and language comprehension. It is difficult to select appropriate WM tests and scoring methods for specific purposes, and this may explain why previous WM research has produced mixed results for language acquisition and development in both L1 and L2 (Conway, Kane, Bunting, Hambrick, Wilhelm & Engle, 2005; Rodriguez, 2008).

Among various WM span tests, the RST is one of the most widely used tests for measuring verbal WM capacity in sentence processing (MacDonald & Christiansen, 2002). With a dual-task mode, the RST analyzes the storage and processing of test items through word recall.
and sentence judgments. The RST is considered an assessment tool designed to measure the function of the central executive during language comprehension. In the classic RST (Daneman & Carpenter, 1980), participants are asked to remember the final word of each sentence while reading sentences aloud. The number of irrelevant words they recall is recorded to measure their WM capacity. The RST assumes a trade-off relationship between its storage and processing components in L1 sentence processing (Alptekin & Ercetin, 2010).

RST performance is known to be closely related to the ambiguity resolution process (Friederici, Steinhauser & Mecklinger, 1998; Just & Carpenter, 1992; King & Just, 1991; Vos, 2001). Despite some criticisms about the correlation between the RST and the criterion task (Hummel, 2002), the test has been widely used in language research (Waters & Caplan, 1996). In terms of the relationship between WM and sentence processing, previous studies have generally suggested that existing WM tests such as the RST can measure some (but not all) aspects of language comprehension.

In the present study, the RST was used to measure the participants’ WM capacity during their sentence processing in Korean. All three groups took the RST in Korean. Heritage speakers, whose first language was Korean but dominant language was English, took the RST only in Korean because previous studies have demonstrated that there is a consistent relationship across languages between L1 and L2 speakers in terms of their WM capacity (Alptekin & Ercetin, 2010; Juffs, 2005; Miyake & Friedman, 1998; Osaka & Osaka, 1992; Osaka et al., 1993; Van den Norrt, Bosch & Hugdahl, 2006) and that the RST in the L2 is more closely related to L2 reading comprehension than the RST in the L1 (McDonald, 2006; Miyake & Friedman, 1998).

In the RST in Korean, the participants read sentences aloud and remembered irrelevant sentence-final words for later recall and were asked to judge the acceptability of sentences.
Although the RST measure of WM capacity does not fully represent all the cognitive resource capacity involved in language processing, the processing portion of RST scores (the acceptability test) may be closely correlated with the bilingual processing capability in Korean. This study’s operational definition of WM is as follows: some part of general cognitive resources used to process and store online linguistic information during the process of resolving temporary ambiguities. In addition, this study’s main assumption regarding WM is that reading comprehension is a cognitive task processed with limited cognitive resources and thus that individual variations of WM capacity may or may not make a difference in the use of plausibility information and syntactic preferences during the ambiguity resolution process for native speakers of Korean, heritage speakers of Korean, and L2 learners of Korean. In this study, we measured the level of WM capacity through the RST—developed by Conway et al. (2005) and adapted into a Korean version by Kim (2008)—in Korean. Hence, the RST score was as an indicator of the participant’s cognitive capacity with respect to his or her reading skill in Korean.

Because the RST is a dual task involving both the storage and processing of information, there are two sources of RST data: storage and processing data. Because of the ceiling effect of processing scores and the strong correlation between processing and storage scores, previous L1/L2 research has generally employed storage scores as a measure of WM capacity. Hence, RST storage scores were used in the statistical analysis to examine WM effects in the current study. RST processing scores were not considered in the analysis, but were used as a pre-screening tool because data from those participants whose RST processing scores were below a certain level—according to Conway et al. (2005), 85% on average—had to be removed from further analysis. Accordingly, the RST storage score for those participants whose RST processing score (i.e., the score for the logical decision task) was less than 80% were excluded.
This study examined the potential effects of WM on the use of different types of information during the process of resolving temporary ambiguities to account for considerable within-group/between-group differences in processing L1 Korean between the three groups and to determine how WM capacity and proficiency interact with the use of syntactic and plausibility information in online processing. Further, the individual differences in the processing of plausibility information in Korean were investigated. Previous studies of WM effects in L1 and L2 sentence processing have generally suggested that L2 learners are more likely to show lower RST scores than L1 speakers. However, some studies have found that high-span L2 learners and low-span native speakers tend to show similar processing patterns (Alptekin & Ercetin, 2010; Felser & Roberts, 2007; Juffs, 2005; Omaki, 2005; Williams, 2006).

Korean is a head-final language, which is more likely to require the reader to store and process data in WM than a head-initial language. Thus, the processing of a head-final language may entail stronger WM effects. For processing a head-final language, which has the MV at the end of the sentence, the reader may need to temporarily store the linguistic input to a greater extent than the reader of a head-initial language. Kwon (2008) suggested that sentence processing in Korean is incremental, but delayed within the clausal boundary. In the case of English, which has the SVO word order, the verb appears relatively earlier in the sentence processing, and post-verbal arguments are supposed to be attached to the verb according to the verb’s subcategorization information. Accordingly, verbs play an important role in initial structure building. On the other hand, some noun-associated cues such as case markers and plausibility are presumed to facilitate the initial process structure building in Korean, which has the SOV word order. For instance, case markers in Korean play an active role in predicting the structure of the incoming linguistic input (NPs) before the MV (Kim, 1999; Kwon, 2008). In the
present study, the plausibility relationship between the noun and the verb could be out of the clausal boundary because the experimental sentences consisted of embedded RCs and main clauses, and nouns could be an object of either the embedded RCV or the MV. As a result, the participants with a higher level of WM capacity were expected to process such sentence ambiguities more clearly and efficiently.

2.1.5. Processing vs. Interpretation

In this study, paraphrased comprehension questions were asked for all test materials to examine the participants’ final interpretations, and answering patterns and reading times for those questions were recorded and analyzed. The participants were asked to read Korean sentences with structural ambiguities with MV/reduced RCV-biased structures. After each sentence, they answered each paraphrased comprehension question.

For the end-of-sentence (EOS) comprehension questions, some studies have expressed concerns about the effects of such questions (Waters & Caplan, 1996). Readers with high WM span or high-verbal tendency may notice the nature of the experiment through comprehension questions targeting the ambiguity. According to Swets et al. (2008), the use of different types of questions influences sentence processing by shifting the reader’s ambiguity resolution strategies. That is, more specific and detailed questions are likely to force the reader to pay more attention to the meaning of the sentence as a whole in RC attachment preference experiments.

However, other studies of the ambiguity resolution process have employed different types of comprehension questions and found that the type of question has little effect on this process (Christianson et al., 2010; Christianson et al., 2001; Patson et al., 2006, 2009). This suggests that forced yes/no questions may not influence the reader’s comprehension.
In the present study, a forced yes/no question format under the paraphrase paradigm was employed to examine the participants’ final interpretations. Under the paraphrase paradigm, leading question sentences consist of two paraphrased sentences from each experimental sentence with its ambiguity resolved toward appropriate syntactic analyses. Paraphrased questions target the role of the first NP as the object of the embedded RCV or the MV. This paraphrase verification task was first used in Kim’s (2008) dissertation. Under the paraphrase paradigm for EOS questions, readers are instructed to determine whether the meaning of the EOS question sentence is identical to that of the sentence they just read regardless of whether the meaning makes sense or not.

There are two points to consider in the use of the paraphrase paradigm. First, comprehension questions are intended to ensure that the reader comprehends the sentence as much as they can and to let the researcher ask detailed information on different syntactic forms from the target sentences. At the same time, these questions should avoid priming readers with some expectation so that they could focus on certain parts of the sentence for their answers. Thus, forced yes/no questions with paraphrases of target sentences are expected to influence the way that readers parse the sentences as less detrimental as they could and keep them focused on the meaning of the sentences, although it is not possible to exclude the possibility that this type of question may implicitly explore the reader’s interpretations.

2.2. Bilingual Sentence Processing Research

2.2.1. Research on L2 sentence processing

Processing incoming inputs online is very important for L1 and L2 sentence processing and comprehension. However, most of the previous psycholinguistic studies of sentence processing have employed L1 data, and thus, there exist few parsing models for predicting the pattern of L2 parsing (cf. van Patten, 1996; Bates & MacWhinney, 1997). Even in the field of
second language acquisition (SLA), where language learners’ linguistic knowledge of lexis, morphology, or syntax has been systematically investigated, few studies have examined L2 sentence processing (Rodriguez, 2008). Instead, L2 researchers have typically focused on the ultimate attainment of L2 grammatical knowledge. In addition, these researchers have emphasized L2 learners’ linguistic competence by considering the role of the input, the effect of the L1 on L2 development, and the ultimate attainment of L2 proficiency, among others. A number of studies of bilingual processing have employed off-line grammaticality judgment tests. When measured by off-line measurements such as grammaticality judgment tests or language proficiency tests, bilinguals often show native-like or near-native-like proficiency in terms of accuracy scores (Papadopoulou & Clahsen, 2003). However, online parsing processes for L2 learners remain unclear (Klein, 1999).

There is growing interest in examining bilingual processing through online experimental methods such as cross-modal priming, eye-tracking techniques and self-paced reading/listening tasks. Researchers have focused not only on bilingual processing but also on sources of individual variations in readers’ cognitive capacity, such as WM (Clahsen & Felsers, 2006a, b). Accordingly, since early 2000, a small but increasing number of studies has examined the online process of bilingual parsing. In a multi-lingual age, investigating online bilingual processing by using diverse language populations should provide a better understanding of the nature of language processing for both L1 and L2 speakers and may provide empirical support for some theories of L2 acquisition and development.

As in research on L1 sentence processing, the ambiguity resolution process is one of the most commonly investigated topics in research on bilingual processing. Having L2 learners choose a certain structural alternative over another in the presence of structural ambiguities can
shed some light on their online parsing, including what information they use and when it is used (Clahsen & Felser, 2006a). Previous studies of bilingual sentence processing have considered the following four major linguistic categories influencing the ambiguity resolution process: RC/PP attachment preferences (Dussias, 2001, 2003; Frenck-Mestre, 2002; Frenck-Mestre & Pynte, 1997; Ha, 2005); filler-gap dependency principles by movement (Felser & Roberts, 2007; Gorrell, 1995; Williams, 2006; Williams, Mobious & Kim, 2001); subject vs. object ambiguity (Felser & Roberts, 2004; Frenck-Mestre & Pynte, 1997; Juffs & Harrington, 1996); and MV vs. reduced RCV ambiguity (Juffs, 1998). Factors such as L1 effects, incomplete grammatical knowledge, limited vocabulary, and constrained cognitive resources are generally expected to influence bilingual sentence processing (Clahsen & Felser, 2006a).

The RC attachment preference is one of the most frequently examined syntactic topics in research on bilingual sentence processing. However, previous studies have produced somewhat mixed results. In terms of sentence parsing by L1 speakers and L2 learners, some have found similarities, whereas others, differences (Felser, Marinis & Clahsen, 2003; Papadopoulou & Clahsen, 2003). This suggests that, regardless of the specific L1 attachment preference, even advanced L2 learners with native-like grammatical knowledge of the L2 (as measured by off-line tasks) are not likely to show native-like attachment preferences based on online measures, which require more cognitive resources to process. However, such studies have typically drawn their conclusions based on null results, which cannot be used to fully determine whether L2 learners show processing patterns that differ from those for L1 speakers. Given such mixed results, cross-linguistic variations in L1 parsing with this construction, and criticisms about various methodological issues associated with RC attachment preferences, this construction may not be
suitable for testing hypotheses about bilingual sentence processing online (Fernandez, 2006; Rodriguez, 2008).

Numerous issues associated with bilingual sentence processing remain controversial. One of the most interesting and intriguing questions in this field is whether advanced L2 learners and native speakers show similar behaviors during sentence parsing (Havik, Roberts, van Hout, Schreuder & Haverkort, 2009), but previous studies have produced mixed results. Some have provided evidence of native-like syntactic processing by L2 learners regardless of the L1 effect (e.g., Frenck-Mestre & Pynte, 1997; Havik et al., 2009; Indefrey, 2006; Juffs, 1998; Juffs & Harrington, 1995, 1996; Williams et al. 2001; Williams, 2006). Despite the overall latency difference, L1 and L2 sentence processing mechanisms are generally considered to be fundamentally the same (Dekydtspotter, Schwartz & Sprouse, 2006).

Juffs and Harrington (1996) considered reading times to examine the difference between L2 competence and parsing performance and found that Chinese-speaking ESL (English as a second language) learners demonstrated the same GP effect as their native counterparts.

(23) *After Bill drank the water proved to be poisoned.*

Juffs and Harrington (1996) were interested in the ambiguity resolution process when readers arrived at the disambiguating word *proved* and their accuracy on the grammaticality judgment task. The two groups were similar in terms of their accuracy and reading time patterns in general, but with a large difference in the reading time between the groups. Juffs and Harrington maintained that L2 learners showed poor parsing performance despite their linguistic competence. However, it is unclear how they controlled for the proficiency of the participants. In addition, they did not describe their statistical analyses in detail.
Juffs (1998a) examined the use of lexical information for the resolution of MV vs. reduced RCV ambiguities by ESL learners with different L1 backgrounds.

(24) *The bad boys criticized almost every day were playing in the park.*

The participants showed native-like parsing patterns that were consistent with the GP phenomenon and were sensitive to information in the verb argument structure during their ambiguity resolution.


(25) *Before Mary ate the pizza arrived from the local restaurant.*

According to their RST results, there was no significant relationship between the effect size of the GP phenomenon and the WM span. In addition, there was no group difference in accuracy scores for the grammaticality judgment task, and the pattern of reading times was similar to that in Juffs and Harrington (1996). L2 learners could revise their initial misanalysis to provide the same interpretation as native speakers. Thus, the L2 learners in Juffs (1996, 1998b, 2004) showed native-like parsing patterns. Juffs claimed that the L2 learners in his studies used syntactic information during the initial parsing process to resolve ambiguities and employed plausibility information during the reanalysis process. However, these studies do not pose a serious challenge to Clahsen and Felser’s SSH because these studies had a number of methodological problems.

On the other hand, a number of studies have provide evidence of non-native-like processing by L2 learners even when the learners show native-like linguistic competence in offline tasks (Dekydtspotter et al., 2006, Felser & Roberts, 2007; Marinis, Roberts, Felser &
Clahsen, 2005; Papadopoulou & Clahsen, 2003; Rah, 2008), although some (e.g., Marinis et al., 2005; Papadopoulou & Clahsen, 2003) reported null results.

In a comparison study considering reduced RC ambiguities with sentences such as The horse raced past the barn fell, Rah (2008) found that German L2 learners of English had native-like linguistic knowledge of English in a grammaticality judgment task but showed significant differences in processing in self-paced reading. In her study, the ambiguity originated from verb morphology, and materials presented good and bad post-ambiguity cues in ambiguous conditions.

(26a) The brown sparrow saw an insect on a high branch. (unambiguous example)

(26b) The brown sparrow seen by the hungry cat pecked at an insect. (ambiguous example with a good cue)

(26c) The brown sparrow noticed on an upper branch pecked at an insect. (ambiguous example with a bad cue)

(26d) The brown sparrow noticed almost every day pecked at an insect. (ambiguous distractor)

(26e) The brown sparrow noticed an insect on a high branch. (unambiguous distractor)

L2 learners spent more time reading the sentences and spent the largest amount of time on ambiguous sentences with a bad cue. This provides some support for Clahsen and Felser’s (2006a, b) shallow syntactic processing by L2 learners, who are assumed to make extensive use of non-syntactic information such as verb morphology and post-ambiguity cues (Rah, 2008). Rah (2008) suggested that the difference in sentence processing between L1 speakers and L2 learners may be due to their diverging processing strategies, not to a lack of linguistic competence.
Hopp (2006) claimed that mixed results from previous studies of L2 sentence processing may be due to the use of diverse types of experimental tasks, although he found that L2 learners can show native-like sentence processing. Studies employing reading for the meaning question format (e.g. Felser & Roberts, 2007; Marinis et al., 2005; Papadopoulou & Clahsen, 2003) are more likely to provide evidence of native-like sentence-processing behaviors among L2 learners (Havik et al., 2009; Hopp, 2006) than those studies employing tasks drawing explicit attention to the experimental treatment (e.g. Havik, Roberts, van Hout, Schreuder & Haverkort, 2009; Williams, 2006).

2.2.2. General characteristics of L2 sentence processing

In general, L2 learners tend to be slow in processing linguistic inputs, and in many cases, their responses are not as accurate as those of L1 speakers (Fernandez, 2002). This may be because they are less likely to engage in the automatization of lower-level processes (Frenck-Mestre, 2002). However, Hopp (2007) suggested that differences in the processing speed do not necessarily mean qualitative differences in L1 and L2 sentence processing.

In cognitive neuroscience, L2 sentence processing has typically been investigated in terms of the neural representation, that is, whether native speakers and non-native speakers activate the same area of the brain for language processing. Event-related potential (ERP) and neuroimaging studies of language processing have produced somewhat mixed results. Some studies have found that L2 sentence processing is generally different from L1 sentence processing because L2 sentence processing involves the wider activation of brain areas (Hopp, 2007). For instance, L2 sentence processing employs less efficient parsing routes for integrating syntactic information (Hahne, 2001). In terms of semantic processing, native-like semantic processing can come in the early stages of L2 development, whereas native-like syntactic processing is likely to appear with increased proficiency (Hahne, 2001). Other ERP studies have
demonstrated that L2 learners may show native-like electrophysiological brain signatures that depend mainly on their language proficiency (Hahne, Mueller & Clahsen, 2006; Hahne, 2001).

Previous L2 research has suggested that the level of L2 proficiency and the age of acquisition (AOA) play important roles in L2 sentence processing. Neuroimaging studies have found that L2 proficiency plays a more important role in neurocognitive processing mechanisms, and recent ERP studies dissociating the AOA from L2 proficiency have provided evidence of the native-like processing of morphosyntax by adult L2 learners (Steinhauer, White & Drury, 2009). L2 learners with a high level of L2 proficiency are likely to show brain activation patterns that are similar to those of native speakers, whereas those with a low level of L2 proficiency are likely to show diverse and qualitatively different activation patterns (e.g., N400) (Weber-Fox & Neville, 1996; Hahne et al., 2006).

Some studies have suggested that even adult L2 learners can attain native-like proficiency (Birdsong, 2006). The age of exposure in isolation may not explain the ultimate attainment of linguistic competence or native-like processing in L2 learners (Steinhauer et al., 2009).

One of the main cognitive components in L2 acquisition and development influenced by aging is WM. The current emphasis on attention has highlighted the importance of WM in bilingual processing. Miyake and Friedman (1999) presumed that language proficiency and WM capacity may interact such that the higher the level of language proficiency, the more effective the storage of information in WM. However, it remains unclear whether an age-related increase or decrease in WM capacity has differential effects on SLA.

L2 WM is closely related to L2 comprehension (Miyake & Friedman, 1998; McDonald, 2006). The processing of a less dominant/fluent L2 is more likely to drain WM resources. Hence, L2 learners with a low level of L2 proficiency tend to have a low level of WM capacity in terms
of their L2, which can limit their ability to quickly and accurately decode and process L2 linguistic inputs (McDonald, 2006). However, because of some correlation with language proficiency, WM’s independent contribution to L2 comprehension remains unclear.

L1 speakers with a high level of WM capacity tend to make accurate interpretations and show diverse ERP patterns. If language proficiency can be controlled, then it is possible to hypothesize about whether L2 learners can perform sentence processing in the same way as a subgroup of L1 speakers (e.g., native speakers with a low level of WM capacity). Assuming that online language processing occurs somewhere between fully specified parsing and shallow parsing even for the L1, shallow parsing may be (to some extent) caused by individual differences in WM capacity and selective attention between language learners (Sekerina & Brooks, 2006).

Although a number of studies have examined language proficiency, the age of acquisition, or the amount of exposure, few have investigated the relationship between individuals’ WM capacity and L2 parsing (Clahsen & Felser, 2006a), producing mixed results (Felser & Robers, 2007; Juffs, 2004, 2005).

By considering a sample of Germans who were advanced L2 learners of Dutch, Havik et al. (2009) conducted two self-paced reading experiments to test the processing of subject-object ambiguities in Dutch RCs (Dat is de vrouw die de meisjes heft/hebben gezien) in terms of L1 and L2 WM. relation to WM in L1 or L2. Native speakers of German as well as those of Dutch showed a strong preference for subject reading over object reading, and the two languages had similar linguistic properties for the test structure.

Havik et al. (2009) expected that when these L2 learners had to make use of only syntactic information (morphosyntactic means for the number agreement) for parsing, they
would behave differently from the native speakers of each language because they were supposed to rely more heavily on non-syntactic information for L2 (syntactic) processing. Havik et al. were concerned about the influence of WM capacity (for either the L1 or the L2) on L2 parsing.

Native speakers in both languages showed a preference for subject reading, whereas L2 learners with a high level of WM capacity had a similar processing advantage and showed a preference for subject reading for short sentences. The accuracy results also indicated a preference for subject reading, and in terms of object reading and long items, the learners behaved like native speakers, showing a low level of WM capacity.

In general, the L2 learners did not show the same structural preference (a preference for subject reading) as native speakers of Dutch. In the first experiment, which involved a semantic verification task with the subject/object interpretation of NPs in the active voice/past tense, only high-span learners showed a native-like preference for short items. In the second experiment, which involved a semantic verification task for only 6.25% of experimental targets (25% of subject/object interpretation items out of 25% of experimental sentences), there was no processing effect for the L2 learners and no WM effect. This suggests that L2 learners may perform native-like processing for certain language structures under certain linguistic conditions (Havik et al. 2009; Indefrey, 2006).

2.2.3. The Shallow Structure Hypothesis (SSH)

Clahsen and Felser (2006a, b) proposed the shallow structure hypothesis (SSH) based on their L1/L2 comparative research data. The SSH posits fundamental differences in online parsing between L1 speakers and adult L2 learners. For instance, according to the SSH, adult L2 learners tend to make heavy use of non-syntactic information during parsing, whereas L1 speakers can take advantage of syntactic as well as non-syntactic information. The underuse of syntactic information by adult L2 learners may be due to their underdeveloped L2 grammar, which in turn
may not provide them with a full parsing route. However, Clahsen and Felser (2006a) claimed that the full parsing route is not available to L2 learners regardless of their L2 proficiency.

There are two routes for language parsing for L1 sentence processing: full and shallow parsing. Full parsing is based on a fully specified syntactic analysis with native-like grammar, which can handle complicated syntactic representations. On the other hand, shallow parsing results from a prompt but coarse-grained syntactic representation with the use of various types of non-syntactic information, including world knowledge, pragmatic inferences, metalinguistic information, lexical/semantic information, and statistical patterns, among others. Adult native speakers can use either system when appropriate for their communication purposes, although some native speakers engage in shallow parsing in certain circumstances (Christianson et al., 2001; Christianson et al., 2006; Ferreira, 2003, 2003; Sanford & Sturt, 2002). Even for native speakers, there may be substantial differences in language proficiency when it is measured using different sets of language proficiency tests (Pakulak, 2008).

In theory, as in the case of young L1 speakers, both full and shallow parsing routes are available to L2 learners. However, because of their limited or underdeveloped L2 grammar, they tend to rely more on shallow parsing, which is readily available to them. As a result, their incomplete L2 grammar cannot lead them to native-like parsing. In this sense, the assumption underlying the SSH is that the learner has deficient L2 grammar or an inappropriate syntactic representation. As L2 learners’ proficiency increases, they may make a declining use of shallow parsing, which can eventually differentiate advanced L2 speakers from near-native ones (Sorace, 2006).

Previous studies of L2 sentence processing have found that that L2 learners make heavy use of non-syntactic information such as the verb argument structure and thematic/plausibility
information in native-like ways (Hopp, 2007). For comprehension purposes, using non-syntactic information without projecting a detailed syntactic representation can often lead to a sound understanding of the incoming input in both the L1 and the L2. Clahsen and Felser (2006a, b) noted that this shallow syntactic representation is optional in L1 sentence processing and thus can be employed on particular occasions, whereas it is an obligatory (although not desirable) component of bilingual processing.

Bilinguals who have insufficient L2 grammar are likely to have difficulty processing L2 inputs because of the L2’s additional processing demand and their limited cognitive capacity and resources. Bilingual processing is not automatized and thus requires more cognitive resources for processing lower level information. For online processing, even advanced L2 learners with nearly native-like grammar may have difficulty incorporating multiple information sources simultaneously because online sentence processing requires substantial cognitive resources and their processing routine for certain structures is not yet proceduralized like that of native speakers. Although Clahsen and Felser (2006a) did not consider the reader’s WM as a main source of differences in L1/L2 sentence processing (e.g., Felser & Roberts, 2007; Juffs, 2004, 2005), Dowen and Carreiras (2006) suggested that a mix of factors such as slow sentence processing, partial grammar acquisition, and the reader’s WM capacity may account for at least some part of observed differences in L1/L2 sentence processing.

However, despite of the SSH’s promising and reasonable contributions to theories of bilingual language processing, previous empirical studies of the SSH have typically considered only a narrow population of bilinguals, namely adult L2 learners. Clahsen and Felser (2006b) indicated a need for further research considering diverse linguistic structures and bilingual populations (e.g., simultaneous bilinguals, native-like L2 speakers, and L2 dominant speakers).
Most of the previous studies of bilinguals have typically focused on L2 learners with English as the L2 and a Romance language as the L1. In addition, grammatical similarities between the L1 and the L2 may influence bilingual sentence processing (Juffs, 1998; Williams et al., 2001). Dowens and Carreiras (2006) pointed out that Clahsen and Felser (2006a, b) did not address sentence processing by young bilinguals, although they provided applicable implications for research on L1/L2 sentence processing. This suggests a need for sentence-processing studies considering a wider range of bilingual populations such as L2 learners of non-Romance languages (Clahsen & Felser, 2006a, b; Pakulak, 2008).

2.2.4. **Semantic plausibility and WM in L2 sentence processing**

It remains unclear whether plausibility information is used during initial parsing by L2 learners, although L2 learners are assumed to show the same level of sensitivity to semantic plausibility information as adult L1 speakers of English for their final interpretation or syntactic revision. Some studies have proposed that plausibility information as well as syntactic information can play critical roles in L2 learners’ processing of filler-gap dependencies (Clashen & Felser, 2006a). On the other hand, other studies have indicated that L2 learners do not respond to plausibility information in the same way as adult L1 speakers (Pearlmutter & MacDonald, 1995). Previous L1 research has generally suggested that plausibility information is closely related to WM capacity in bilingual sentence processing.

Among various factors influencing L1 sentence processing, WM in bilingual sentence processing has attracted little attention from researchers (Havik et al., 2009; Omaki, 2005). Nevertheless, WM is considered a key component of L2 acquisition and an important predictor of L2 learning (Ardila, 2003).

L2 researchers have speculated that differences in WM capacity can influence L2 acquisition (Ellis, 2001). For instance, L2 learners with a high level of WM capacity are more
likely to acquire some linguistic parts of the L2 (e.g., phonology or morphosyntax) than those with a low level of WM capacity because their L2 input storage is larger and their processing capability is better. Mackey, Philip, Egi, Fuji, and Tatsume (2002) found that, particularly in the early stages of L2 learning, L2 learners with a high level of WM capacity are likely to a higher notice rate than those with a low level of WM capacity and suggested that subcomponents of WM (e.g., the phonological loop) play an important role in L2 learning. Bilingual sentence processing, which is less automatic and proficient than L1 sentence processing, may consume more WM, and thus, WM may play a less important role in bilingual sentence processing than in L1 sentence processing (Omaki, 2005). Even those native speakers with some forced limitations in terms of their cognitive resources may show L2-like processing performance (Hopp, 2007).

Hummel (2002) found a positive correlation between WM capacity and L2 proficiency, and Harrington and Sawyer (1992) noted a significant correlation between the L2 RST developed by Daneman and Carpenter (1980) and scores on the grammar section of the Test of English as a Foreign Language (TOEFL). L2 grammaticality judgment scores have been found to be closely related to RST scores, which is likely due to their strong correlation with language proficiency (Robinson, 2002). Miyake and Friedman (1998) found an analogous correlation between L2 RST scores and listening comprehension. In addition, WM, when measured by a speaking span test, may influence L2 production (Mota, 2003). Mota (2003) employed a picture description task and a narrative task for 13 ESL learners and found a positive correlation between WM span scores and fluency, accuracy, and structural complexity but a negative correlation between the scores and lexical density.

Comparative studies of WM effects on L1 and L2 sentence processing have found that L2 learners’ RST scores are generally lower than those of L1 speakers. Some L1 studies have suggested
that high-span L2 learners show processing patterns that are similar to those of low-span L1 speakers (Felser & Roberts, 2007; Juffs, 2005; Omaki, 2005; Williams et al., 2006).

However, in a series of self-paced reading studies of WM effects on bilingual sentence processing, Juffs (2004, 2005, 2006) found no correlations between WM and L2 sentence processing for L2 learners with different L1s (Chinese, Japanese, and Spanish). L1 speakers and L2 learners showed similar GP effects on reduced RCs/wh-questions and analogous reading profiles in many cases (Juffs, 2004, 2005), but there was no WM effect on the ambiguity resolution process for L2 learners.

L2 learners and some L1 speakers with a low level of WM capacity tend to be less sensitive to information in the intermediate gap than those with a high level of WM capacity. This may be due to their limited cognitive resources such as WM (Indefry, 2006). A number of studies have examined the WM effects on bilingual sentence parsing in terms of filler-gap dependency phenomena (Dussias & Pinar, 2010; Felser & Roberts, 2007; Marinis et al., 2005; Williams, 2006; Williams et al., 2001). Dussias and Pinar (2010) examined the WM effect on the use of plausibility information in L2 reading by considering Chinese ESL learners’ processing of long-distance wh-questions. In a self-paced reading experiment with an English RST and with the level of proficiency held constant between the two WM groups, only those ESL learners with a high level of WM capacity showed native-like processing patterns, whereas those with a low level of WM capacity did not show native-like sensitivity to plausibility information.

In a reading time study with Chinese, Korean and German ESL learners, Williams, Mobious, and Kim (2001) found that all the ESL learners were sensitive to plausibility information during initial parsing as well as during the reanalysis. In addition, they investigated whether L2 learners would show a native-like plausibility effect on a possible gap structure
during online processing with a wh-movement structure and found that both L1 and L2 learners of English showed a plausibility effect regardless of differences in L1s and WM capacity. They also found a significant interaction between plausibility and the word position. In addition to the WM span test, they compared the reading time for ESL learners with different L1s (i.e., Korean, Chinese, and German) with that for native speakers of English by using a stop-making-sense task in which the participants were supposed to press a button immediately after noticing any implausibility then continue reading the rest of the sentence. They manipulated plausibility for a filler NP or an MV in the wh-question and employed the following experimental sentences:

(27a) *Which machine did the mechanic fix (plausible gap: fix the machine) the motorbike with (real gap: disambiguating point) two weeks ago?* (plausible condition)

(27b) *Which customer did the mechanic (implausible gap: fix the customer) fix the motorbike for (real gap: disambiguating point) two weeks ago?* (implausible condition)

For the stop-making-sense decision task, both groups showed higher response rates in the implausible condition than the plausible condition at and immediately after the verb. On the other hand, at and immediately after the post-verbal noun, they showed higher response rates in the plausible condition than the implausible condition. In terms of reading times, the two groups showed similar reading profiles regardless of their L1 and WM capacity, although L1 speakers showed larger movements in the reading pattern for critical regions. L2 learners showed a delayed response in the use of plausibility information. For instance, L2 learners did not show an early plausibility effect until the post-verbal noun. In general, both groups showed a strong tendency to interpret the filler as the direct object of the verb and simultaneously process its plausibility as the direct object. L2 learners had more difficulty with their reanalysis in the plausible condition. However, no WM effect was found for reading times.
To examine the extent to which plausibility information influences syntactic processing, Williams (2006) conducted another experiment with ESL learners whose L1 was Chinese or a Romance language. In addition to the experimental materials in Williams et al. (2001), they considered two additional words (intensifier + adjective) as a semantic cue for the reanalysis to see the clearer plausibility effect rather than simple competition between arguments.

(28) *Which machine/friend did the mechanic fix the very noisy motorbike with two weeks ago?*

In the first experiment, which employed the same stop-making-sense task as Williams et al. (2001), both native and non-native groups showed plausibility effects in the plausible and implausible conditions, but there was a delayed effect for the non-native group. However, the stop-making-sense task is thought to require the reader to take a close and intentional look at the meaning (plausibility) of the sentence. Hence, a self-paced reading measure without a stop-making-sense part was employed in the second experiment in Williams (2006) for all plausible sentences. In addition, a memory probe test for every two sentences was employed to check the comprehension of experimental sentences, and for the analysis, the participants were divided into two groups based on their WM scores.

In general, according to the plausibility test without a stop-making-sense task, the critical post-verbal region showed a significant plausibility effect only for the native group with high WM scores. In other words, L1 speakers with high WM scores showed plausibility effects that were similar to those in the first experiment even for a different task. However, L1 speakers with low WM scores exhibited no plausibility effects. Interestingly, only L2 learners with high WM scores showed a delayed plausibility effect on the PP region, whereas those with low WM scores showed no plausibility effects. On the other hand, the filler as a direct object interpretation was
found regardless of the plausibility effect for both groups, which indicates their strong preference for syntactic information over plausibility information.

Williams (2006) claimed “less efficient and more task-dependent” processing in the L2 through the use of plausibility information, although both groups showed fairly similar reading profiles. He also suggested individual differences dependent on cognitive resources for semantic processing in the L2 because previous studies have found WM effects on the use of plausibility information in bilingual sentence processing, including the reduced use of plausibility information in attachment decisions (Just & Carpenter, 1992) and reduced sensitivity to plausibility information (Pearlmutter & MacDonald, 1995). Although L1 speakers showed immediate plausibility effects, L2 learners showed delayed or no plausibility effects because, as a result of less proficient processing by L2 learners, bilingual sentence processing requires more cognitive resources for processing insufficient information. There were differences in the pattern of the plausibility effect based on WM scores even for native speakers (Williams, 2006).

Recently, some studies have examined the plausibility effect on bilingual sentence processing. Felser and Roberts (2004, 2011) investigated plausibility effects on the process of resolving L2 subject/object ambiguities through a self-paced reading study with Greek L2 learners of English and native speakers of English. They manipulated the plausibility of post-verbal NPs with subsequent verbs such that the post-verbal NPs could be either the direct object of the embedded verb or the subject of the subsequent MV. They assessed two syntactic structures (complement clauses vs. adjunct clauses). The comprehension question was accompanied by the following sentences.

(29a) The man read the book/girl had upset very many people. (complement clause)

(29b) While the band played the song/beer pleased all the customers. (adjunct clause)
Felser and Roberts (2004, 2011) found differences in sentence processing between L1 speakers and L2 learners. L2 learners showed stronger plausibility effects on both sentences, indicating longer reading times at and one word after the disambiguating word in the implausible condition. Interestingly, they found some differences in the reanalysis effect and the accuracy rate for comprehension questions between the two sentences. L2 learners showed shorter reading times one word after the disambiguating region for implausible sentences. Felser and Roberts suggested that this reflects the process of recovering immediately from the initial misanalysis through the use of plausibility information. However, there was no plausibility effect in terms of accuracy rates for complement clauses. In contrast, L2 learners showed reanalysis effects on no region with respect to adjunct clauses, but their accuracy rate was lower for plausible sentences than for implausible ones. Felser and Roberts interpreted these results as an indication of strong plausibility effects on L2 sentence processing. This suggests that, in terms of reading times and accuracy, this plausibility effect for the immediate recovery from the initial misanalysis may vary according to the syntactic construction. Felser and Roberts also suggest that differences in reanalysis patterns between the two types of sentences may result from differences in the processing cost of reanalysis across clausal boundaries in the adjunct clause condition.

Felser and Roberts (2004, 2011) found a plausibility effect on an immediate reanalysis but no plausibility effect on accuracy rates for complement clauses. In this regard, examining Korean sentences with different RC structures involving possible reanalyses/revisions only within clausal boundaries for a similar resolution process for local ambiguities (an object noun for embedded RC verbs vs. MVs) should shed some light on whether there similar plausibility effects can be found for reading time data but no plausibility effect for interpretation data, as indicated in Felser and Roberts (2004).
Montag and MacDonald (2010) compared passive RCs with object RCs in Korean, which correspond with each other in terms of their meaning and word order, and found differences between the two types of RCs only when aspect markings showed different plausibility effects. They generated 24 quadruples (30) in both English and Korean and normed them for plausibility.

(30a) Object RC, Plausible:

[경찰이] 경찰이, 체포한 경찰이 사람을, 지니고 있는 것으로 알려졌다.

‘The thief that the policeman arrested was known to carry a knife.’

(30b) Object RC, Neutral:

[경찰이] 경찰이, 놀라게 한 경찰이 사람을, 지니고 있는 것으로 알려졌다.
kengchal-i, kengchal-ey, nollakeyha-n, cini-koy, cini-koy, kal-ul, kal-ul, iss-nun, iss-nun, kes-ulo, kes-ulo, allye-ss-ta, allye-ss-ta, known-PAS-PAST-DEC

‘The thief that the policeman scared was known to carry a knife.’

(30c) Passive RC, Plausible:

[경찰에] 경찰에, 체포한 경찰에 사람을, 지니고 있는 것으로 알려졌다.

‘The thief that was arrested by the policeman was known to carry a knife.’

(30d) Passive RC, Neutral:

[경찰이] 경찰이, 놀라게한 경찰이 사람을, 지니고 있는 것으로 알려졌다.
kyengchal-i, kyengchal-i, nollakeyha-n, cini-koy, cini-koy, kal-ul, kal-ul, iss-nun, iss-nun, allye-ss-ta, allye-ss-ta, known-PAS-PAST-DEC

‘The thief that was scared by the policeman was known to carry a knife.’

Both English (n=22) and Korean (n=23) participants read the sentences in a self-paced reading task that followed comprehension questions. In English, object RCs tend to have stronger plausibility effects than subject RCs (King & Just, 1991). In general, objective RCs in Korean are more challenging to process than passive RCs. According to their findings, only the passive RCs in Korean showed no significant plausibility effect in terms of reading times. Reading times at the head noun showed a significant difference between plausible object RCs
and passive-neutral RCs. In this regard, Montag and MacDonald suggested that the plausibility effect may vary according to the type of RC.

In his dissertation study, Rodriguez (2008) tested some hypotheses about L2 learners’ online processing based on the SSH. He conducted three experiments to compare processing behaviors of L2 learners and native speakers of English. In the first experiment, which included Spanish-English bilinguals whose English proficiency ranged from high-intermediate to advanced levels, he compared the use of word category information for the resolution of local ambiguities in adverbial subordinate clauses in English between these bilinguals and native speakers of English through a self-paced reading paradigm.

(31a) *When the tiger appeared the bird flew away.* (intransitively-biased subordinate verbs: overall implausible combination)

Question: *Did the tiger appear? Y*

cf. No condition: *Did the tiger fly away?*

(31b) *After the maid cleaned the stove began to heat up.* (transitively-biased subordinate verbs: plausible combination)

Q: *Did the stove heat up? Y*

cf. No condition – *Did the maid clean the stove?*

Both native and non-native speakers employed syntactic preference and plausibility information but not lexical information (verb transitivity information) during initial parsing in the first experiment. However, the focus of the first experiment was on the role of subcategorization information, not on the role of plausibility information. Based on accuracy scores from comprehension questions, for which both groups achieved higher scores in the intransitively biased subordinate verb condition than in the transitively biased subordinate verb
condition, Rodriguez (2008) determined that incorrect thematic role assignments remained, which is consistent with the good-enough representation (GER) (Ferreira et al., 2002; Christianson et al., 2001, Sanford & Sturt, 2002). He agreed that parsing often leads to incomplete representations because of time, cognitive capacity, and ambiguity constraints, among others. In this regard, this may be the reason for the efficient management of WM resources during parsing (Rodriguez, 2008). Rodriguez (2008) conducted additional correlation analyses of WM capacity and L1/L2 reading times and found a correlation between L1 WM capacity and L2 grammatical proficiency. In general, however, his WM findings do not provide support for the effect of WM (L1) on bilingual sentence processing.

In terms of reading times, Rodriguez (2008) found that L2 learners and L1 speakers processed local ambiguities in a manner similar, although there were differences in reading times between the two groups. Both groups prioritized syntactic preferences such as the MA principle for initial structure building without considering verb subcategorization information, which is inconsistent with the SSH’s main assumptions.

On the other hand, Kim (2008) conducted a series of online and offline reading experiments with globally ambiguous RC attachment preference constructions by using a task involving paraphrased sentences in English and Korean and by employing native speakers of English, native speakers of Korean, and Korean-English bilinguals and found that sentence complexity and WM capacity were closely related to disambiguation strategies of all three groups. Bilinguals depended more on lexical factors, whereas both groups of native speakers, more on syntactic factors (Kim, 2008), which is consistent with the SSH.

Swets, Hambrick, Ferreira, and Desmet (2007) tested the WM effect on RC attachment preferences by conducting separate offline surveys of English speakers and Dutch speakers for
the RC attachment in English and Dutch and employing two types of WM measures: reading span and spatial span tests. In the first experiment, they presented all the experimental sentences, and in the second experiment, they divided the sentences into three groups to minimize the variance from differences in their chunking strategies. In general, native speakers of English were less likely to show attachment preferences than native speakers of Dutch. However, according to these two experiments, low-span readers were more likely to show attachment preferences than high-span ones regardless of their L1. To explain this finding, Swets et al. (2007) applied Fodor’s (1998, 2002) implicit prosody hypothesis and found that high-span readers held more information for sentence processing than low-span readers and thus processed longer phrases and expressions simultaneously, whereas low-span readers paused between the RC and its modifying NP and thus showed stronger attachment preferences. Based on this finding, Swets et al. suggested that the interaction between the chunking strategy and WM with individual differences may influence the ambiguity resolution process during sentence processing. However, this conclusion is inconsistent with the findings of previous studies considering children (Felser et al., 2003) and adults with different L1 backgrounds, which have generally shown that low-span readers are less likely to show attachment preferences than high-span ones. This difference may be due to Swets et al.’s use of offline surveys, which typically allow respondents more time for post-interpretive processing. Other studies using offline tasks have shown stronger attachment preferences regardless of the L1.

L2 learners with a high level of WM capacity are likely to better store and process L2 information than those with a low level of WM capacity, which can facilitate their L2 acquisition and development. Previous studies of bilinguals have generally assumed WM to be confounded with actual L2 proficiency (Rodríguez, 2008). In addition, few studies have examined L1
sentence processing in the context of WM use, and thus, there is a need for studies investigating this topic through the use of diverse populations and methods. The lack of conclusive findings on WM effects on L2 sentence processing may be due in part to the fact that previous studies have not employed systematic research and analysis methods for L1 and L2 sentence processing (Conway et al., 2005; Friedman & Miyake, 2004). Thus, there is a need for further psycholinguistic research on the relationship between WM and L1/L2 sentence processing, which should provide a better understanding of the nature of sentence processing and individual differences in parsing performance (Juffs, 2005).

2.2.5. Heritage speakers’ language processing

In the era of globalization, it is more common to see multilinguals than true monolinguals. On the other hand, it is also difficult to find multilinguals who have native-like proficiency in the languages they speak. Multilinguals often show diverse levels of proficiency in their languages based on the age of exposure, language use, purposes/motives, cultural environments, and socioeconomic factors, among others (Moreno & Kutas, 2005).

One of the most interesting but overlooked populations in this field is early simultaneous bilinguals. Because of globalization, the term “bilingual” is difficult to define. In this regard, the present study later provides an operational definition of early simultaneous bilinguals. Early simultaneous bilinguals are expected to have a good command of both languages, but realistically, it is very difficult to strike an exact balance between two languages. The present study focuses on early Korean English bilinguals. In some cases, these bilinguals are referred to as speakers or learners of their “heritage language” (HL) because their second language dominates their HL in social and educational contexts. “Heritage speakers” (HSs) are generally defined as those who speak ethnic minority languages around the world. However, in the context of the United States, HSs refer to people who spoke a non-English language as their HL after
immigrating to the United States at an early age and thus speak English as their dominant language (Montrul, 2011).

In general, previous studies of HSs with different L1s have typically focused on their linguistic profiles and competence to provide better language education in school as an integral part of bilingual education. HSs generally have sufficient exposure to their HL early in their lives, but access to it often becomes less frequent and limited when they start school. Accordingly, their acquisition of the HL is likely to be incomplete or partial (Kim, 2005; Montrul, 2008; O’Grady, Lee & Lee, 2008).

An interesting aspect of HSs’ linguistic competence found in previous research (Montrul, 2008) is that they appear to have no problem with their daily communication, although they lack some linguistic competence in reading, writing, and meta-linguistic knowledge of L1 grammar. However, little is known about potential differences in parsing mechanisms between native speakers and HSs.

For a better understanding of how the HL (which is first acquired but less dominant) is processed, we need to examine their processing as well as their proficiency measures (Montrul, 2008). It remains unclear whether HSs exhibit native-like proficiency in the ambiguity resolution process or whether the SSH can be applied to HL processing. Birdsong (2006) suggested that sentence processing is slower and less efficient for the HL, the less dominant language. However, is HSs’ L1 sentence processing qualitatively similar to that of monolingual speakers if HSs have had limited inputs but made constant use of their HL? In the present study, to investigate this question, we considered Korean HSs who made active use of both English and Korean on a daily basis.
In the present study, the term “heritage speakers” are used to refer to early simultaneous Korean-English bilinguals whose HL is Korean. All of the HSs recruited for the experiment had lived immersed in English before puberty, and many had been exposed to both English and Korean at home. Although the level of proficiency in the HL is likely to vary according many factors, we controlled for their HL proficiency through proficiency tests and self-reports such that only those HSs above the intermediate level (which was high enough for reading and understanding all experimental sentences) were selected. More detailed linguistic profiles of these HSs are provided in the Participant section.

Briefly, the present study defines “Korean heritage speakers” as early Korean-English bilinguals 1) who have acquired Korean as their first language (either monolingually or simultaneously with English since birth); 2) whose reading proficiency in Korean is above the intermediate level; 3) who have continued to use Korean (although probably within a very limited set of contexts); 4) who feel more comfortable with English; and 5) who has native-like or better English proficiency.

The HL often becomes the less dominant language of early bilinguals once they start socializing in the dominant language. Accordingly, the HL is generally observed to undergo either fossilization (i.e., incomplete acquisition despite continuous (passive) exposure) or attrition (the loss of proficiency with little or no use) (Montrul, 2005, 2008). A number of studies of HSs have focused on L1 attrition, including the effect of attrition on various types of linguistic knowledge and L2 transfer effects on the HL, among others. In this regard, it is possible to assume that the dominant L2 may influence the HL in terms of sentence processing in the current study (Cook, 2003). In addition, language dominance is often intertwined with other issues in L2
acquisition and development, including the age of acquisition\(^1\) (AOA – early exposure to the target language) and L2 proficiency.

However, the L1 attrition of previously acquired linguistic knowledge or the L2 effect on this language group is not the primary focus of the present study. In this study, HSs’ acquisition of the HL might have been incomplete, but they did not suffer from the attrition of their HL knowledge and processing ability. First, they continued to use Korean on a daily basis since birth, although they were less likely to use Korean than English and their use of Korean was generally limited to family members and friends. Second, their proficiency (measured by a cloze test, a WM test, and self-reporting) was controlled to be above the intermediate level (high–intermediate/advanced) such that they were competent enough to participate in the reading experiment. Based on the screening process, this study’s HSs were not expected to show a substantial loss of their HL competence and processing ability.

In addition, sociocultural variables such as the sense of ethnic identity, motivation, and attitudes toward learning, and parents’ socioeconomic status, which are mainly about linguistic processing aspects of HSs’ language learning, were not controlled for in the present study.

As mentioned earlier, previous studies have found that the AOA and proficiency effects on L2 acquisition are often highly correlated, and thus, it is difficult to differentiate the AOA effect from the proficiency effect in sentence-processing research with bilinguals (Pakulak, 2008). Both the proficiency level and the AOA are related to native-like processing. Montrul (2011, 2008) claimed that HS research is pertinent to bilingual research in that it provides opportunities for a better understanding of ongoing issues in the field from another angle. Such issues include the role of the AOA, input and learning environments, language transfer, linguistic

\(^1\) The present study uses Montrul’s (2008, p. 502) definition of the term “age”: “a macro-variable including not only biological age but also all the other variables with age such as maturational state, cognitive development, degree of L1 proficiency, and L1 and L2 usage.”
competence, and parsing mechanisms. In addition, there is a need for further research on HLs with diverse linguistic phenomena for a better understanding of L1 acquisition (O’Grady et al., 2008).

Some researchers have found similarities between HL and L2 acquisition. However, Montrul (2008, 2011) indicated a need for a better understanding of HSs’ linguistic skills for more meaningful comparisons. HSs share some common features with L1 speakers (e.g., early exposure in a natural environment before the critical period) and adult L2 learners (e.g., limited access to inputs, underdeveloped linguistic knowledge, and transfer effects) (Kim, 2005; Montrul, 2008). Montrul (2008) noted several aspects of HSs that may be applicable to linguistics, psycholinguistics, and SLA research: i) the effect of the critical period on L1/L2 competence development, ii) the effect of inputs and experiences on various types (implicit and explicit) of knowledge acquisition, iii) the distinction between linguistic competence and performance, iv) innate or environmental factors influencing the achievement of native competence, v) the transfer effect, and vi) the development and linguistic properties of adult bilinguals (p. 501). Because of these multifaceted aspects of HSs, she suggested that HSs are interesting and valuable candidates for exploring these constantly contentious issues. For instance, she proposed one hypothesis about the AOA effect for HSs and L2 learners, both of which are language learners whose linguistic knowledge and skills are likely to be underdeveloped. In addition, both often have limited opportunities for accessing L1 inputs. Montrul (2008) suggested that a comparison of linguistic knowledge and skills between these two groups (with the difference in the proficiency level controlled for) should provide some evidence of the AOA effect on L1/L2 language acquisition.
In addition, recent studies of bilingual sentence processing have focused on adult learners and found that data from this L2 group do not provide a complete picture of bilingual processing. Fleige, Mackay and Piske (2002) found that early and late L2 learners show different levels of sensitivity in bilingual processing. In this sense, research considering HSs may be a good way to test these factors separately. For HSs, early exposure to the HL provides an advantage in the acquisition of some aspects of the HL, such as syntax and lexico-semantics (Moreno & Kutas, 2005).

Clahsen and Felser (2006a, b) suggested that the full parsing route may not be available to adult L2 learners. Thus, we can easily hypothesize that the full parsing route for L2 parsing may be not achievable after the critical period. In this regard, HSs, together with native speakers and adult L2 learners, may be good candidates for testing a hypothesis about whether neurobiological changes around puberty influence the partial availability of full parsing in L2 acquisition (Clahsen & Felser, 2006b). By controlling the AOA for HSs, it may be possible to determine whether full parsing is available for HSs with a relatively high level of HL proficiency. If full parsing is not available for HSs, then it may be due solely to their HL proficiency because they have been exposed to the HL since birth. They also suggested a need for further research considering diverse learner populations, languages, and linguistic constructions based on various languages for a better understanding bilingual sentence processing (Clahsen & Felser, 2006a, b).

Consistent with Montrul’s (2008) suggestion for a comparative analysis of HSs and L2 speakers for a better understanding of AOA effects (and possibly implicit/explicit knowledge acquisition) on L1/L2 acquisition, a comparative study of the ambiguity resolution process through the use of three groups of Korean speakers—native speakers of Korean, Korean HSs, and L2 learners of Korean—should shed some light on language processing for several reasons.
First, Korean HSs have been exposed to Korean since birth. Although it is well known that there are substantial differences in the level of HL proficiency among Korean HSs, they are likely to be truly “balanced” bilinguals if they were exposed to Korean at an early age and have continued to use it fairly frequently.

In reality, however, many early bilinguals tend to show ”incomplete” knowledge of their HL, particularly when their L2 becomes the dominant language in their daily lives (Montrul, 2008). L1 children and HSs show some similarities in terms of their linguistic profiles, including a lack of fully developed linguistic competence, fully developed parsing mechanisms, underdeveloped vocabulary and morphosyntactic rules, slow lexical retrieval, and limited WM capacity (Clahsen & Felser, 2006a). On the other hand, HSs and L1 children are different in that HSs are more likely to show general world knowledge and fully developed parsing mechanisms.

Unlike adult L2 learners, HSs pose no challenge to the maturational time constraint on the development of their basic linguistic competence. In addition, it would be interesting to compare sentence processing and interpretations between HSs and native speakers/adult L2 learners by carefully designing the screening process and controlling for the proficiency level for both languages of the HSs.

Because HSs often did not have formal instruction in their HL and their L1 development is not likely to be perfect, there may be some quantitative differences in the reading speed and comprehension accuracy between HSs and native speakers, as indicated by sentence processing among adult L2 learners. Nevertheless, addressing the question of whether HSs and native speakers show qualitatively similar or different processing patterns should provide a better understanding of L1 sentence processing.
In general, semantic processing is less likely to vary for L2 proficiency groups than for native speakers, whereas syntactic processing is likely to vary across proficiency levels, particularly for adult L2 learners with a low level of L2 proficiency (Hahne & Friederich, 2001). Previous studies have provided some evidence that syntactic processing and semantic processing show different ERP patterns. For instance, adult L2 learners are likely to show slower semantic processing than native speakers, and their syntactic processing is likely to be qualitatively different from that of native speakers (e.g., more bilateral ERP patterns and attenuated P600 waveforms) (Hahne & Friederici, 2001).

Moreno and Kutas (2005) examined age effects on L2 semantic processing by considering Spanish HSs and found that the age of exposure and language proficiency independently influence L2 semantic processing. They employed two groups of early Spanish-English bilinguals (Spanish-dominant groups) and an English-dominant group. Both of the bilingual groups displayed significantly weaker N400 effects and were more likely to show semantic anomalies in their HL than in their dominant language. Vocabulary knowledge and the AOA for the language to be tested were found to be key predictors of this delayed effect of semantic incongruity.

HSs tend to have a wide range of language exposure and experience for their HL. Nevertheless, HSs are generally known to have some advantages in receptive skills, including listening, reading, pronunciation, and vocabulary. However, previous studies comparing HSs with L2 learners in terms of various linguistic aspects have produced somewhat mixed results. Some studies have found that HSs have no advantage over L2 learners in terms of grammar, syntax, or L1 interferences (Kim, 2001; O’Grady et al., 2001). Kim (2001) found no HL advantage in the acquisition of parametric values (e.g., the location of the wh-phrase in a wh-
question) for four written production tasks with two test constructions: the null subject (pro-drop parameter) and wh-movement constructions. Kim (2005) claimed that previous findings based on HSs and L2 learners provide support for the assumption that although HSs are good at understanding linguistic inputs, have native-like pronunciation, and have a wider vocabulary than L2 learners, these advantages may not guarantee their morphosyntax superiority in terms of the target language.

On the other hand, for some aspects of linguistic knowledge (and processing), such as syntax, lexical semantics, and inflectional morphology, Montrul (2005, 2008) suggested some age and exposure advantages for HSs. In a series of comparative studies of Spanish HSs and English L2 learners of Spanish, Montrul (2005, 2008) found that Spanish HSs with an intermediate or even low level of HL proficiency acquired certain aspects of Spanish L1 syntax that were not fully learned by L2 learners. Montrul (2008) posited that certain features of syntax are impervious to attrition once they are acquired by HSs at a young age.

Au and Romo (1997) found that bilingual participants with early exposure to their HLs (Korean and Spanish) during childhood had some advantages in morpho-syntactic intuitions (subject vs. object case marker distinction) and phonological perceptions (tense vs. lax consonant distinction) in their HL in college. Later, Au, Oh, Knightly, Jun, and Romo (2008) conducted a similar study with three groups (native speakers, childhood overhearers, and typical late L2 learners of Spanish) and found that overhearing provided bilingual participants with more native-like phonology acquisition, whereas frequent speaking during childhood provided them with considerable benefits in terms of morphosyntax as well as phonology when they re-learned their HL.
Montrul (2005) examined linguistic knowledge by comparing adult late L2 learners’ L2 with early bilinguals’ HL and provided empirical evidence of similar linguistic properties for these two groups in terms of unaccusativity. With the proficiency level controlled for, HSs showed slightly better linguistic knowledge than adult L2 learners because of their early exposure to the HL. On the other hand, by focusing on phonology and morphology, Au, Knightly, Jun, and Oh (2002) investigated the long-term effects of childhood overhearing by comparing Spanish HSs with late L2 learners of Spanish. They found that HSs had a considerable advantage over L2 learners in terms of phonology but received no benefit from childhood overhearing in terms of morphosyntactic performance (either their perception or production).

O’Grady, Lee, and Choo (2001) found no difference between college-aged HSs and non-HS learners of Korean in processing Korean RCs. More specifically, they examined processing differences between these two groups of learners (second- and fourth-semester college students taking Korean classes) in terms of their use of morphosyntactic cues to understand RC structures in Korean. Different types of Korean RCs (five each for DO/subject RC and nine types of sentences as fillers) were used by controlling for the effect of pragmatic and contextual information on sentence processing. In this way, they examined how the three groups (the second-semester HS group, the second-semester non-HS group, and the fourth-semester non-HS group) made use of morphosyntactic cues in a picture selection task with audio inputs. In the picture selection task, the participants were required to select one of the three pictures from the booklet after listening to a sentence describing the picture. Interestingly, according to the error analysis, even those HSs in advanced classes could not make use of semantic, pragmatic, or
contextual information but employed only structural cues for processing RCs. This suggests that HSs are more like native speakers and less like L2 learners with respect to the SSH.

Sekerina (2004) conducted an eye-tracking experiment with fluent Russian-English bilinguals to examine how they processed GP sentences in English—e.g., *Put the frog on the napkin in the box* (ambiguous condition) vs. *Put the frog on the napkin and in the box* (unambiguous condition)—in two contexts (one referent and two referents). The bilingual group showed non-native-like processing patterns in terms of English PP attachment preferences. In contrast to the findings of previous studies considering native speakers of English and the same linguistic constructions, bilingual participants were more likely to make errors (22%) than native speakers, particularly those errors in the unambiguous and one-referent conditions (36.7%). However, there was no ambiguity effect on error rates. Sekerina suggested that the high error rates did not reflect some deficiencies in the referential principle but the application of speedy and simpler heuristics, that is, shallow processing strategies.

Montrul (2006) conducted an empirical analysis with Spanish HSs to compare their linguistic competence and processing of unaccusative and unergative verbs with those of native speakers of English and Spanish. Based on self-assessment and proficiency test results for the two languages, she observed that Spanish HSs showed fairly similar linguistic competence in both languages despite the apparent dominance of their English over their Spanish. In addition, Spanish HSs had strong knowledge of the syntactic effect of unaccusativity (unaccusative and unergative verbs) in both languages, which was determined through proficiency tests and four experiments involving grammaticality judgment tasks and online visual probe recognition tasks. According to her findings, the overall response patterns for Spanish HSs were similar to those for both groups of native speakers, although there was some differences for some constructions and...
some latency differences in response times between HSs and native speakers.

Given the mixed results from previous studies of heritage language processing, one of the main research questions in the present study is whether Korean HSs and adult L2 learners of Korean can perform native-like processing for temporarily ambiguous Korean RCs. Native-like processing requires the appropriate use of syntactic as well as non-syntactic information throughout the processing.

Non-native-like HS processing may be due to the HSs’ non-native-like competence. That is, without native-like linguistic knowledge and grammar, HSs cannot process language inputs like native speakers. If HSs have native-like linguistic knowledge but show non-native-like processing, then their limited processing capacity may have limited their ability to incorporate their knowledge into online processing. This raises the question of whether non-native-like processing is like L1 child’s processing (i.e., the use of syntactic information during parsing because of limited cognitive resources) or like adult L2 learners’ reliance on non-syntactic information because of their incomplete linguistic knowledge and shallow parsing strategies (Kim, 2005; O'Grady et al., 2001). In this regard, an analysis of Korean HSs should contribute to the literature by providing a better understanding of the nature of sentence processing by a wider range of language learners.
CHAPTER 3

TEMPORARY AMBIGUITY RESOLUTION IN KOREAN

Chapter 3 provides an overview of the structure and processing of Korean RCs with temporary ambiguities and a brief summary of Korean linguistics. In addition, some research questions and predictions are proposed for each group of participants based on the theories discussed in previous chapters.

3.1. Sentence Processing in Korean

To investigate the roles of plausibility and syntactic preference in Korean sentence processing, this study employs two types of RC constructions. The next section provides some background information on the Korean language, focusing on temporarily ambiguous RC constructions. Then the grammatical properties of the types of RCs examined in this study are described and analyzed.

3.1.1. Korean Syntax

Most of the existing sentence-parsing models, including syntax-first serial models and parallel multiple-constraint models, are based on data from native speakers of English, a head-initial language (Kim, 2004; Kim, 2005; Koh, 1997). Given that one of the main goals of psycholinguistics is the discovery of universal vs. language specific aspects of language processing (Kim, 2004; Kim, 2005; Papadopoulou, 2006), recent cross-linguistic approaches to the identification of universal principles of language processing and language specific features of a wider variety of languages represent a welcome phenomenon. Accordingly, it is necessary and meaningful to examine various languages that are not closely related to English, particularly those languages that have attracted little attention from researchers, for verifying or generalizing previous findings based on native speakers of English.
In English, a language in which verbs play a major role in the sentence structure, verbs are one of the most heavily investigated topics in sentence processing (Garrett, 1990). In general, verbs with different information about subcategorization and thematic roles are more likely to facilitate the determination of the sentence structure than other elements of the sentence in English. A major research question for L1 sentence processing in English is when and how certain types of verbs are applied during sentence processing (Garrett, 1990).

Korean is typologically different from English. One distinctive syntactic feature of Korean relative to English is the location of head components in the formation of larger phrases. A head in linguistics is the word that determines the syntactic category of the phrase it belongs to. For instance, English consistently places the head before other arguments in the phrase and has the SVO word order. On the other hand, the head is placed at the end of the phrase in Korean. Accordingly, English structures are prepositional, whereas Korean structures tend to be postpositional and also agglutinative because a series of particles (e.g., cases, delimiters, and suffixes) follow the words that they modify (Ha, 2005). In spite of the fact that verbs are located at the clause-final position in Korean, parsing in Korean tends to be incremental (Inoue & Fodor, 1995; Kwon, 2008) just as in head-initial languages.

Korean is known to have a rich morphological system and it allows flexible ordering of nominal arguments and adjuncts within sentences (Koh, 1997), although its default order is SOV. Case is more influential than word order in L1-Korean sentence interpretation (O’Grady, Kwak, Lee & Lee, 2011) since it allows the projection of possible syntactic structures before any predicates has yet appeared in the sentence, by interacting with other sources of information (Kim, 2005; Kim, 1999; Kwon, 2008). Unlike in English, it is very difficult to determine the structure of a sentence in Korean by relying only on just word order.
Models of incremental processing of head-final structures maintain that the argument structure of a verb can be predicted from the arguments preceding it (Kamide & Mitchell, 1999). NPs are considered to be attached to a partial structure without waiting for the head to appear. Korean speakers are thought to incrementally process sentences by generating probable structural information even before verbs appear by relying on other available information such as lexical and semantic information, case marking information, the word order, and plausibility, among others (Kim, 2005; Kwon, 2008).

Another important property of Korean is that it is a pro-drop language. Subjects and objects are frequently elided when the meaning of the sentence can be inferred from the context or relevant pragmatic information, which can result in many temporary ambiguities and consequent processing difficulties (Kim, 2005).

According to Sohn (2002), English-speaking learners of Korean tend to have difficulty learning Korean as a result of both its inherent linguistic complexity and its differences from English. In particular, Korean’s SOV order and its agglutinative morphology with a high degree of inflectional complexity pose serious challenges to native speakers of English. As a result, English speakers learning Korean are especially likely to have difficulty learning its syntactic properties compared to its lexical and pragmatic properties of Korean (Kim, 2005).

There are two common types of syntactic ambiguities in Korean: morphological and attachment ambiguities. Morphological ambiguities are easy to resolve, whereas attachment ambiguities in predicates and NPs such as NP + VP or VP + NP are more difficult. Many Korean sentences have more than two predicates and NPs and adverbials might be attached to either predicate in many such sentences. That is, there is often temporary ambiguity about which predicate NPs and adverbials should be attached to.
For example, Korean uses many center-embedded RCs that are thought to be laborious to process (Kwon & Kim, 2001). The word order in RCs differs between English and Korean, consistent with the basic word order differences between the two languages. English RCs are head-initial, i.e., the RC follows the head noun it modifies, which Korean RCs are head-final, i.e., the RC precedes the head noun it modifies. Another important difference is that Korean does not have relative pronouns. It has instead a relativizer suffix (REL) –(u)n that appears at the end of the predicate (-un after consonant, -n after vowel) (Sohn, 1999). The relativizer can form a relative clause ender possibly with preceding tense/aspect and/or mood suffix. A consequence of these differences between the languages is that in Korean it is often not clear that there is a RC until its end, while English RCs are clearly marked from the beginning. Thus, in Korean, there can be temporary ambiguity about whether particular NPs belong to the main clause or instead to an embedded RC.

The examples below illustrate some simple Korean sentences that have nominative (-i/ka) and accusative (-lul) case-marked NPs and also more complex versions that contain RCs.

(32a) Simple SOV (canonical word order)

아기가 모유를 마신다.
aki-ka moyu-lul masi-n-ta
baby-NOM breast milk-ACC drink-IN-DEC

The baby drinks breast milk.

(32b) Simple OSV (an unusual but possible scrambled word order)

모유를 아기가 마신다.
moyu-lul aki-ka masi-n-ta
breast milk-ACC baby-NOM drink-IN-DEC

The baby drinks breast milk.

(33a) Subject-extracted RC

[모유를 마시는] 아기
[_i moyu-lul] masi-nun aki
breast milk-ACC drink-REL baby
_The baby who drinks breast milk..._

(33b) Object-extracted RC

The baby who drinks breast milk...
be in casual speech, the listener can sometimes employ other kinds of information such as animacy to help figure out the sentence structure (Kim, 1999).

Despite the existence of extensive temporary structural ambiguity in Korean sentences, native speakers of Korean have no difficulty immediately projecting a possible sentence structure from information so far (e.g., Kim, 1999; Koh, 1997), and instant revisions/reanalyses often appear locally with no major disruption to the online processing of the sentence (Kwon, 2008). On the other hand, heritage speakers (HSs) tend to have difficulty processing RCs (O’Grady et al., 2001).

3.1.2. Temporary ambiguities in sentence processing in Korean

In Korean, NPs are syntactically analyzed immediately based on all available information, and this often leads to the need for reanalysis (Koh, 1997). NPs’ roles can be analyzed incrementally using information from case markers, but attachment across clausal boundaries may not be incremental because of the need to wait for some appropriate information (Kwon, 2008).

In an eye-tracking study of the effect of topic markers “(는)은” –(n)un and delimiters “만”-man (only) on revision processes for Korean RCs, Kim (2001) found some support for serial models.

(36a) RC beginning with a nominative marker

<table>
<thead>
<tr>
<th>그 운전수가</th>
<th>청소부를</th>
<th>설득한</th>
<th>건축가를</th>
<th>비판한다.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ku wunnchunsoo-ka</td>
<td>chunksopu-lul</td>
<td>seltuk-han</td>
<td>kunchuka-lul</td>
<td>bipanhanda.</td>
</tr>
</tbody>
</table>

The driver criticized the architect who persuaded the janitor.

(36b) Complex NP beginning with a nominative marker

<table>
<thead>
<tr>
<th>그 운전수가</th>
<th>청소부를</th>
<th>설득한</th>
<th>사실이</th>
<th>알려졌다.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ku wunnchunsoo-ka</td>
<td>chunksopu-lul</td>
<td>seltuk-han</td>
<td>sasil-i</td>
<td>allyeiciessta.</td>
</tr>
</tbody>
</table>

The driver criticized the architect who persuaded the janitor.

was known
The fact that the driver persuaded the janitor was known.

(36c) RC beginning with a delimiter marker

그 운전수만 청소부를 설득한 건축가를 비판한다.
ku wunnchunsoo-man chunksopu-lul seltuk-han kunchuka-lul bipanhanda.
The driver-DEL janitor-ACC persuade-REL architect-ACC criticized

Only the driver criticized the architect who persuaded the janitor.

(36d) RC beginning with a topic marker

그 운전수는 청소부를 설득한 건축가를 비판한다.
ku wunnchunsoo-nun chunksopu-lul seltuk-han kunchuka-lul bipanhanda.
The driver-TOP janitor-ACC persuade-REL architect-ACC criticized

The driver criticized the architect who persuaded the janitor.

The results suggested that readers followed the MA principle without any delay when the topic marker “는” –nun, was present as in example (36d). Readers used the topic marker in the revision process as a cue for the start of a revision or an effective revision process (Kim, 2001). That is, the initial sequence of NPs was interpreted as co-arguments of the main clause, but when the disambiguating region (persuade-REL) made it clear there was a RC, there were reanalysis effects shown by extended re-reading times and an elevated probability of regressive eye movements.

Kwon and Kim (2001) examined the effects of thematic role and topic markers on reanalysis processes in Korean and found that information from thematic role markers played an important role.

(37a)

영희가 [아이를 늑이터에서 본] 소녀에게 손짓했다.
Younghee-ka [ai-lul nolithe-eyse po-n] sonye-eykey soneishae-ss-ta
Younghee-TOP [child-ACC playground-LOC see-REL] girl-DAT wave-PAST-DEC

Younghee waved to the girl who saw a child at the playground.

(37b)

영희가 아이를 [늑이터에서 본] 소녀에게 소개했다.
Younghee-ka ai-lul [nolithe-eyse po-n] sonye-eykey sokayhae-ss-ta
After reading the first phrase Younghee-ka ai-lul nolithe-eyse po-n (Younghee-TOP child-ACC playground-LOC see-REL) in both (37a) and (37b), readers started computing a simple sentence structure such as Younghee-ka ai-lul nolithe-eyse po-n-ta (Younghee-TOP child-ACC playground-LOC see-DEC). However, when they then encountered the dative noun sonye-eykey (to the girl), they restructured the phrase ai-lul nolithe-eyse po-n (child-ACC playground-LOC see-REL) in both (37a) and (37b) as one RC modifying the subsequent noun sonye-eykey (to the girl). That revision worked in example (37a) since the MV soncishae-ss-ta (wave-PAST-DEC) confirmed it. In contrast, in example (37b), the MV sokayhae-ss-ta (introduce-PAST-DEC) required another revision to change the first object-marked noun ai-lul (child-ACC) from the object of the embedded RC verb po-n (see-REL) to the object of the MV sokayhae-ss-ta (introduce-PAST-DEC), since the latter is an obligatorily transitive verb. This second revision, which was based on thematic information, was considered more difficult to perform than the first revision. Because Korean is a head-final language with the verb at the end of the clause/sentence, these types of multiple revisions can become necessary, and thematic role information can play a major role in these revisions (Hirose & Inoue, 1998).

(37c)

영희가 [아이를 놀이터에서 본] 그네에 태웠다.
Younghee-ka [ai-lul nolithe-eyse po-n] kuney-ey taewe-ss-ta
Youngee put the child in a swing that she saw at the playground.

In (37c), from a purely syntactic standpoint, “swing-LOC” can be either a subject or an object of the embedded RC, just like the word “girl” in (37a) and (37b). However, it is less likely than the word “the girl” in (37a) and (37b) to require a revision. In a self-paced reading experiment, Kwon and Kim (2001) found significantly shorter reading times at the sixth word in
(37a) than in (37b) and (37c) and significantly shorter reading times at the sixth word in (37c) than in (37b). The accuracy of responses to comprehension questions was also higher for (37a) than for (37b) and (37c). Sentences like (37b) were the most difficult, presumably because they required multiple revisions. Kwon and Kim (2001) claimed that these findings provide evidence for the use of thematic information in structural revision in Korean.

Kwon and Kim (2001) did not consider the relative plausibility of different interpretations and thus did not control for it for their experimental materials, although they did mention the possibility that other types of non-syntactic information might be used in revision processes, including plausibility and referential resolution preferences. In addition, the last word and the disambiguating point were the same in their sentences, so they could not determine whether non-syntactic information preceding the verb really influenced initial parsing or revision at the end of the sentence. The inclusion of another word after the disambiguating word would allow for the examination of at least two word positions: one at the disambiguating point and the other at the end of the sentence in order to avoid conflating disambiguation and end-of-sentence wrap-up effects (Just & Carpenter, 1980).

Information from case markers and the thematic roles they identify has been found to be employed during revision processes in the later stages of sentence processing in Korean (Kim, 2001; Kwon & Kim, 2001) and in Japanese (Hirose & Inoue, 1998). The verb-finality of Korean and Japanese makes it possible to hypothesize that multiple local revisions could also be influenced by such information even before verbs have appeared (Kwon, 2000; Yamashita, 1994).

Kim (2004) examined spoken sentence processing in Korean using both online and offline measurement methods and found that three types of constraints play important roles in
both online and offline measurements of both initial and ultimate parsing decisions. The three types of constraints were 1) purely syntactic constraints such as preferring the simplest construction, 2) prosodic boundaries that facilitated the placement of syntactic boundaries, and 3) semantic plausibility.

Kim’s (2004) studies used spoken sentences containing one of two types of temporarily ambiguous RC: 1) Subject + Relative Clause (SR) vs. 2) Subject + Object + Relative Clause (SOR). Cross-modal naming was used to measure responses to disambiguating words that were consistent with one of the possible structures but not the other. She administered several pretests to carefully generate experimental materials that crossed syntactic, semantic, and prosodic constraints. Eighteen conditions were created by crossing two syntactic constructions (SR vs. SOR) with three kinds of semantic bias (SR-bias vs. Neutral-bias vs SOR-bias) and three prosodic biases (also SR-bias vs. Neutral-bias vs SOR-bias).

Kim (2004) used SR sentences like (38) and SOR sentences like (39). It is the fifth word that disambiguates the structure of these sentences. In (38), the fact that the fifth word is a noun-ACC means that the previous noun-ACC must have been the object of the embedded RC verb, while the obligatorily transitive verb in the fifth position in (39) means that the previous noun-ACC must be its object rather than the object of the RC verb.

(38) SR (Subject + RC)

영기는 개구리를 싫어하는 친구에게 청개구리를 던졌어.
Youngki threw a green frog [to the friend who dislikes frogs].

(39) SOR (Subject + Object + RC)

영기는 개구리를 싫어하는 친구에게 닭치라고 다가갔어.
Youngki nun kaykwuli-lul [silheha-nun] chinkwu-eykey chengkaykwuli-lul tencyesse

영기는 개구리를 싫어하는 친구에게 청개구리를 던졌어.
Youngki-nun kaykwuli-lul [silheha-nun] chinkwu-eykey chengkaykwuli-lul tencyesse

Youngki threw a green frog [to the friend who dislikes frogs].
Youngki approached to throw a frog [to the friend whom (Youngi) dislikes].

After hearing the first four words, participants saw a visual word that was either a noun leading to the SR completion or a verb leading to the SOR completion. Their task was to read the visual word aloud and complete the sentence. The logic was that people should read the word aloud faster when it was consistent with their analysis of the sentence structure so far. Examples (38) and (39) are taken from the Neutral-bias conditions so there is no semantic information to support one analysis over the other. In other conditions semantic bias supported one or the other structure, and the location of prosodic boundaries in the spoken fragment were also manipulated to be neutral or to support one or the other structure.

There was no significant overall difference between SR and SOR naming times, but only (1) in the prosodically neutral conditions and (2) in both prosodically and semantically neutral conditions, naming times were faster for continuation words consistent with the SR structure than for words consistent with SOR structure. Thus, in the absence of prosodic cues, there was a preference for the simpler SR structure. When semantic and prosodic cues were both present and agreed with one another, naming times were faster for continuation words that were consistent with the structure supported by the cues. When semantic and prosodic cues disagreed, naming times were especially slow. Kim claimed that these findings provided support for parallel multiple constraint-based models with a one-stage parser. Interestingly, there were no significant syntactic preferences in the semantically neutral condition.

Based on the results of the online cross-modal naming experiment, Kim (2004) suggested that semantic plausibility and prosody might start working together with syntactic preference constraints from the earliest stages of sentence processing. A limitation of the cross-modal naming task is that it taps into processing at just one point in the sentence, i.e., the
A task like word-by-word self-paced reading would measure processing at each word and thereby provide more information about how the processing of sentences like these unfolds over time.

Kim did not consider possible individual differences such as WM capacity in the processing of sentences containing temporarily ambiguous RCs. Some previous work in English has found that people with higher WM capacity are better able to hold onto multiple interpretations in parallel longer than those with lower WM capacity (MacDonald et al., 1992) and also to make better use of plausibility cues in resolving temporary structural ambiguities (Pearlmutter & MacDonald, 1995). It thus seems fruitful to collect WM capacity measures in studies of the processing of Korean SR/SOR ambiguities under different plausibility manipulations. Individual differences may be especially evident in L2-Learners and Heritage Speakers.

The present study builds on Kim’s (2004) work on the resolution of temporary SR/SOR ambiguities in Korean in several ways. First, the self-paced reading task is used to provide more information about processing at different points in the sentences. Second, individual differences in WM capacity are measured to determine whether some people are better able than others to make use of semantic constraints when they are present. Third, Korean speakers with different proficiency levels, including native speakers, L2-learners, and heritage speakers will be compared.

3.1.3. Structural ambiguities in Korean relative clauses

The present study employs sentences in six conditions created by crossing two sentence structures (SR vs. SOR) with three levels of plausibility (SR-biased, Neutral, and SOR-biased), as illustrated in the examples in (40) below.
(40a) SR-SR (SR syntactic construction + SR-biased plausibility)


I presented a bunch (of flowers) to my lover, who values flowers.

In (40a), the sentence has the SR structure and plausibility supports that interpretation throughout the sentence. *Flower* is a plausible object of the RC verb *value* after “(my lover+SUBJ) flower+ACC value +REL” *my lover valued flower* and that analysis is confirmed at the disambiguating point “a bunch+ACC.” No GP effect is expected in this condition.

(40b) SOR-SR (SOR syntactic construction + SR-biased plausibility)

나는 꽃을 [아끼는]애인에게 자랑하려고 갔다.

I went to show off flower to my lover, who I valued.

In this condition, the syntactically incorrect initial analysis “(my lover+SUBJ) flower+ACC value+REL” *my lover values flower*, as in (40a), is much more plausible than the syntactically correct analysis “I+SUBJ flower+ACC (lover+DAT) show off+to go+PAST+DEC” *I went to show off flower (to my lover).* Because of the plausibility of the initial misanalysis, it may take longer to revise the initial structure at the disambiguating region. Thus, structural reanalysis is expected at the disambiguating word “show off+to” in this condition.

(40c) SR-SOR (SR syntactic construction + SOR-biased plausibility)

입주자는 [열쇠를 청소하고 있는]주인에게 불편사항을 얘기했다.

The dweller told the landlord who was cleaning a key about the inconvenience.

In this condition, although the initial direct-object analysis “(landlord+SUBJ) key+ACC is cleaning+REL” *the landlord is cleaning a key* is syntactically correct, it is less plausible than “complaints+ACC tell+PAST+DEC” *tell about the inconvenience*. No structural revision is anticipated at the disambiguating point, by the MA while there may be some processing cost for
reconstructing the meanings of the first objects and embedded RC verbs before and at the disambiguating region expected by parallel models.

(40d) SOR-SOR (SOR syntactic construction + SOR-biased plausibility)

The dweller went to the landlord who was cleaning to return the key.

In this condition, the initial direct-object analysis “(landlord+SUBJ) key+ACC is cleaning+REL” the landlord is cleaning a key is less plausible than “key+ACC return+to” return the key, and thus, it is syntactically mianalyzed. As in (40c), there may be some processing cost for reconstructing the meanings of objects and verbs before the disambiguating region, but the restructuring cost at the disambiguating region can be lower than that in (40c) because of the plausible relationship between the object “key+ACC” and the MV “return+to.” In other words, at the disambiguating region or possibly earlier than that, a new syntactically correct analysis with “dweller+SUBJ key+ACC return+to go to+PAST” the dweller went to return the key is expected to be easily constructed because of its plausibility.

(40e) SR-N (SR syntactic construction + Neutral plausibility)

The department head gave a bonus to the employee who left his job.

Kim (2004) created a plausibility-neutral condition to compare the relative strengths of different constraints such as syntactic preference and prosody with or without biased plausibility constraints. In a similar vein, the present study compares the structural difference between SR and SOR analyses under three plausibility conditions—SR-biased, SOR-biased, and plausibility-neutral conditions—to determine whether the simpler SR analysis is preferred over the SOR analysis with or without plausibility biases. The plausibility-neutral condition is intended to
determine whether there is any structural preference and serves as a baseline that conditions with plausibility manipulations can be compared to.

In the SR-N condition, the syntactically correct initial analysis “(the employee+SUBJ) job+ACC leave+REL” the employee left the job is plausible, and thus, the analysis is kept at the disambiguating region. No GP effect is expected for this condition, similar to (40a).

(40f) SOR-N (SOR syntactic construction + Neutral plausibility)

부원님은 직장을 떠나는 사원에게 알아봐주기로 했다.

The department head promised to find a job for the employee who was leaving.

In the SOR-N condition, both the syntactically incorrect initial analysis “(the employee+SUBJ) job+ACC leave+REL” the employee left the job and the possibly later revised and syntactically correct analysis “department head+SUBJ job+ACC find+to promise+PAST” are plausible. Thus, it takes longer at the disambiguating region for a syntactic reanalysis here than in (40e). However, the disruption at the disambiguating region should not be as severe as that in (40b), where plausibility information is in conflict with syntactic preference.

A paraphrase judgment response was required following each sentence to examine participants’ final interpretations. The paraphrases consisted of two sentences that targeted the role of the first NP as the object of the embedded RC verb or the MV (Kim, 2008). Participants were instructed to decide whether the meaning of the paraphrase was identical to that of the sentence they had just read.

A full set of example sentences plus their paraphrases from each experimental condition and also from the filler trials are presented below (41-53).

(41) SR-SR

 나는 [꽃을 아끼는] 애인에게 한바탕을 선물했다.
The dweller went to the landlord to return the key. The landlord was cleaning.

Paraphrase (Yes):

나는 애인에게 자랑하려고 갔다. 내 애인은 꽃을 아꼈다.

I presented a bunch of flower to my lover who valued flower.

Paraphrase (No):

나는 애인에게 자랑하려고 갔다. 내 애인은 꽃을 아꼈다.

I went to show off flower to my lover who I valued.

(42) SOR-SR

The dweller told the landlord who was cleaning a key about the inconvenience.

Paraphrase (Yes):

The dweller told the landlord about the inconvenience. The landlord was cleaning keys.

(44) SOR-SOR

The dweller went to the landlord who was cleaning in order to return the key.

Paraphrase (Yes):

The dweller went to the landlord to return the key. The landlord was cleaning.

(45) SR-N
My brother bought lunch for his junior who was saving money.

Paraphrase (Yes):
My brother bought lunch for his junior. The junior was saving money.

(46) SOR-N

My brother decided to lend money to his friend, whom he (i.e., my brother) trusted.

Paraphrase (Yes):
My brother decided to lend money to his junior. My brother trusted the junior.

(47) Filler sentence: RC with ambiguities (SOR construction)

I returned after I gave a present to the friend that was sick.

Paraphrase (No):
I gave a present to a friend. I stayed with my friend.

(48) Filler sentence: RC with no ambiguity (SR construction)

Mi-Young bought peaches from the man who sold fruits cheap.

Paraphrase (No):
Mi-Young bought peach from a man. He sold fruit at a high price.
(49) Filler sentence: RC with no ambiguity (SOR construction)

Paraphrase (No):

Dad caught the thief who was climbing over a wall.

Paraphrase (Yes):

Dad ran away after throwing a pencil case to his classmate who was sitting.

(50) Filler sentence: RC with no ambiguity (SR construction + Neutral)

The police greeted the teacher who guided the delinquent students.

Paraphrase (No):

The police greeted a teacher. The teacher worked with the best students only.

(51) Filler sentence: RC with no ambiguity (SOR construction + Neutral)

The uncle disposed the inheritance that he took from his young cousin.

Paraphrase (No):

The uncle took the inheritance of his young cousin. The young cousin could get it back.

(52) Filler sentence: Adverbial phrase with scope ambiguities

Dad caught the thief who was climbing over a wall with a bat.

Paraphrase (Yes):

Dad caught the thief who was climbing over a wall.
Dad caught a thief with a bat. He was climbing over a wall.

(53) Filler sentence: Non-RC with and, as, when, etc.

Mr. Kim started the car and took the madam to her home.

Paraphrase (Yes):

Mi-Mr. Kim started the car. He drove the madam home.

The paraphrases were designed so that “yes” answers were correct if the sentence was parsed correctly for all of the experimental conditions except for the SOR-SR condition. The reason the SOR-SR condition was treated differently was that it was the one condition in which plausibility cues early in the sentence and structural preference both supported the SR structure but then the sentence turned out to have the SOR structure instead. Sentences in this condition should have had the strongest bias toward the SR interpretation, so it should be the condition in which people were most likely to maintain that interpretation in spite of a contradictory disambiguating cue later in the sentence. This condition therefore provided a good chance to test of the Good Enough Processing (GEP) idea. If the SR interpretation was usually constructed first because it was simpler and plausibility also initially supported that interpretation, then according to the GEP model, people should sometimes hold onto that interpretation even when later disambiguating evidence contradicted it. This was the condition in which that seemed most likely to happen. For that reason, participants were given the SR paraphrase in the SOR-SR condition so that a “yes” response would indicate that they had held onto the SR interpretation in spite of contradictory disambiguating evidence. In the other conditions with the less preferred SOR structure,
plausibility was either neutral or supported it, so people were less likely to maintain the SR interpretation in those conditions. (In retrospect, however, it was not the best idea to set things up so that the correct response was “yes” in 5 out of 6 conditions and “no” in the one remaining condition because people often have response biases for one or the other type of response and because “yes” responses generally tend to be faster than “no” responses.)

3.2. Research Questions and Predictions

The present study examines how different types of information such as plausibility and syntactic preferences are integrated during the process of resolving temporary ambiguities in Korean. In addition, the study tests whether individual differences in WM influence the use of plausibility information during the ambiguity resolution process and whether there are differences in the ability to use different information sources between native speakers of Korean and L2-Korean learners and Korean HSs. Although some previous work has investigated the use of plausibility and syntactic preferences (and their interaction) as well as the role of WM in structural ambiguity resolution in both L1 and L2 processing in English (Juffs, 1998; Juffs & Harrington, 1996), possible interactions between the two information sources and the role that WM may play in such interactions have not been thoroughly explored across languages. The present study attempts to replicate Kim’s (2004) results for native speakers of Korean and also addresses the following research questions by employing two groups of non-native speakers of Korean (L2 learners of Korean and Korean HSs):

(i) Is there a general structural preference for SRs over SORs in Korean?

(ii) When and how do Korean speakers at different levels of language proficiency make use of plausibility information to help resolve structural ambiguities? Does this plausibility influence initial structural decisions as well as final interpretations?
(iii) Does WM capacity influence readers’ ability to process information plausibility information during the resolution of structural ambiguities?

Three different proficiency groups of Korean-language speakers (native speakers, Korean HSs, and L2-learners of Korean) were recruited. Reading time and paraphrase judgement data for native speakers of Korean were used to determine which major sentence processing models (i.e., serial models vs. parallel models) would better explain the processing of these types of ambiguities in Korean. They were also used as a baseline to compare with the results for each of the two groups of non-native speakers.

In terms of inter-group differences in reading behavior, it is to be expected that the two groups of non-native speakers will read more slowly overall than the native group. It is also likely that they will find the structural ambiguity of the experimental sentences more difficult than native speakers. If there is a general preference for the SR structure because it is simpler than the SOR structure, that preference should be even stronger in the non-native groups than in the native group. However, if one reason the SR structure is preferred is that it occurs more frequently, native speakers may show the preference more strongly because they have more experience with both kinds of sentences. Alternatively, though, it could be predicted that the non-native speakers will show the preference more strongly because in their more limited experience they have encountered many more SR sentences than SOR sentences because of the frequency disparity between them.

The sentences employed in the present experiment were the two types of temporarily ambiguous Korean RC constructions discussed above and exemplified in examples (40a-f) and (41-53), which crossed sentence structure (SR vs. SOR) with plausibility (SR vs. Neutral vs. SOR). When plausibility was biased toward the SR interpretation, the meaning of the first object-
marked noun made it an especially plausible object of the embedded RC verb, which was consistent with the SR structure. In contrast, when plausibility was biased toward the SOR interpretation, the first object-marked noun was not plausible as the object of the embedded RC verb but it was quite plausible as the object of the later-appearing disambiguating main clause verb. In the Neutral plausibility conditions, the first object-marked noun was approximately equally plausible as the object of either of the two verbs.

The expected results for each group are as follows, first for the native Korean speakers. Both syntax-first serial models and parallel multiple constraint-based models expect longer reading times for sentences that are implausible because plausibility is ultimately relevant in all sentence-processing models. However, these models differ in terms of their predictions about the effects of plausibility as the sentence unfolds over time. Syntax-first serial models predict that plausibility does not determine initial parsing decisions but instead influences later interpretation stages. In contrast, parallel multiple constraint-based models predict that plausibility influences initial parsing decisions from the beginning.

If readers initially process the first object-marked noun in the sentences as the object of the immediately following first verb in the sentence, which is the embedded RCV, then the continuation of the sentence in the SR-SR and SR-N conditions supports that interpretation and there is no need for any structural reanalysis when the second object-marked noun appears at word 5. In these conditions, the meaning of the RCV supports the initial structural decision that the first object-marked noun is its object, and so does the rest of the sentence. In contrast, in the SR-SOR condition, the first object-marked noun is not plausible as the object of the immediately following RCV, which should lead to some processing cost there because of the implausibility of the structurally preferred relationship between it and the preceding noun. It is possible that the
implausibility of the preferred structure could lead to a switch to the structurally less preferred SOR interpretation at that point, if plausibility can influence structural choices quickly enough. If it then turns out at word 5 that the sentence actually has the SR structure after all, in spite of its implausibility, there should be additional processing costs at that position. However, if readers’ early structural decisions are not influenced by plausibility, then there should be no processing cost at word 5 because it confirms the SR analysis the reader has maintained despite its implausibility. Thus, a plausibility effect at the RCV is likely because it is implausible for the object-marked verb preceding it. However, despite this early implausibility, there should be no GP effect at words 5 and 6.

In contrast, in the SOR constructions, at the main clause verb in the 5th word position it turns out that the earlier object-marked noun has to be its object rather than the RCV’s object and thus that reanalysis is required. Serial syntax-first models predict that all three conditions with the SOR construction should require reanalysis and thus be approximately equally difficult at word position 5. In contrast, parallel constraint-based models predict that if the context supports an SOR interpretation, there may be little or no processing cost at word 5. In the SOR-biased plausibility conditions, if the first object-marked noun is not plausible as the object of the immediately following RCV, then there should be an expectation that another verb will appear later for which it is a plausible object, providing support for the SOR structure over the SR structure. If the sentence is consistent with that expectation, as it is in the SOR-SOR condition, then no reanalysis is required at word 5. Table 2 below clarifies the predictions of the two types of models about the circumstances under which they predict long reading times at words 3 and 5. The effects are likely to spill over to subsequent word positions, as well. Table 2 is reused here for reference to the expected reading time results at each critical word position.
Table 2

*Expected Reading Time Patterns at Critical Positions*

<table>
<thead>
<tr>
<th>Critical Word Position</th>
<th>Serial Models</th>
<th>Parallel Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word 3</td>
<td>Word 5</td>
<td>Word 3</td>
</tr>
<tr>
<td><strong>SR construction</strong></td>
<td>SR bias &lt; SOR bias</td>
<td>(SR-SR=SR-N)=SR-SOR</td>
</tr>
<tr>
<td><strong>SOR construction</strong></td>
<td>SOR-SR=(SOR-N=SOR-SOR)</td>
<td>SR bias &lt; SOR bias</td>
</tr>
<tr>
<td></td>
<td>SR-SOR &lt; SOR-SOR</td>
<td>SR-SOR ≥ SOR-SOR</td>
</tr>
</tbody>
</table>

Parallel models’ predictions about whole-sentence RTs for both SR and SOR constructions are similar to those of serial models, but they expect different RT patterns across the three plausibility conditions for each construction. Like serial models, they expect longer RTs at words 3 and 4 in SR-SOR and SOR-SOR conditions as well as longer whole-sentence RTs. However, parallel models predict some processing cost at word 5 in the SR-SOR condition but little or no processing cost for reanalysis for the SOR-SOR condition if plausibility can influence structural choices quickly enough. In terms of the structural reanalysis for SOR-SR and SOR-N conditions, parallel models predict that the processing cost is higher at word 5 in the SOR-SR condition than in the SOR-N condition because of the plausible relationship between the object noun at word 2 and the embedded RC verb at word 3 as well as the MV at word 5 in the SOR-N condition.

Neither of these sentence-processing models necessarily makes predictions about post-sentence paraphrase judgment times or accuracy because both focus mainly on the on-line process of resolving temporary ambiguities while reading the sentences. However, it is easier to
derive clear predictions about paraphrase judgments from the Shallow Parsing and GEP models, and those are what the paraphrases were primarily intended to test.

The plausibility effect may be applied for accuracy of the comprehension questions (QA) in both constructions. The investigation of final interpretation via a paraphrased comprehension question was intended to check whether speakers of Korean do care about the real meaning of the sentences they read or not. Regarding final interpretations of the resolved ambiguity, plausibility is likely to affect more strongly the whole meaning of the sentence as shown by some L1 research in English (i.e., partial analysis or lingering wrong analysis from the previous one). Thus, for the SR construction, different QA rates are expected for the SR-SOR condition because of initial implausibility.

For the SOR construction, different QA rates are expected for each plausibility condition based on the plausibility bias and the employment of a full syntactic analysis. Paraphrased questions with the SOR construction can help determine whether readers reanalyze their initial misinterpretation in their final interpretation. For SOR-SR with conflicting plausibility and syntactic construction information, the questions were constructed such that the participants would answer “yes” more often if they relied more on plausibility information in shallow parsing. For SOR-SOR with SOR-biased plausibility facilitating an SOR syntactic analysis (full syntactic parsing with reanalysis), the participants were expected to answer “yes” more often because both plausibility and syntactic analyses work in the same direction. On the other hand, for SOR-N, where plausibility is in the neutral condition with an SOR syntactic analysis, the participants were expected to answer “yes” (“no”) more often if they employed structural reanalysis in their final interpretation (if they were satisfied with their shallow parsing).
Table 3

*Expected Responses to Comprehension Questions*

<table>
<thead>
<tr>
<th>Sentence Structure</th>
<th>Plausibility Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR Bias</td>
<td>SR-Biased</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| SOR                  | Yes: Shallow parsing\(^2\) | Yes | Yes *
|                    | *No: Shallow parsing |         | *No: MA |

In terms of the interaction between plausibility and structural preferences for a simple structural analysis for SR and SOR constructions, serial models predict that SR constructions (SR-SR, SR-N, and SR-SOR) have a processing advantage over SOR constructions (SOR-SR, SOR-N, and SOR-SOR) because of a possible structural reanalysis for SOR constructions regardless of the plausibility condition. SOR constructions may lead to longer RTs and QTs and lower accuracy rates for comprehension questions because these constructions require structural reanalysis in the later stages of parsing. SR syntactic constructions are generally preferred by serial models. In terms of RT patterns across positions between two constructions, serial models expect longer RTs at word 5 of the SOR construction regardless of the plausibility bias information.

On the other hand, parallel models expect diverse RTs and RT patterns across the three plausibility conditions for each construction. Depending on the degree of plausibility agreement between the noun and the subsequent verb in the three plausibility conditions, overall RTs can vary across SR and SOR constructions. In general, when both the semantic plausibility and

\(^2\) Shallow parsing in the comprehension question processing & interpretation refers to a partial analysis of the syntactic structure as a more preferred SR-analysis without projecting a full syntactic structure as in the Good-Enough approach. Depending on the plausibility biasness in each condition, it can have a SR-biased plausibility support or no support from the SOR-biased plausibility condition. It is different from Clahsen & Felser’s (2006) Shallow Parsing by L2 learners with little use of syntactic information.
syntactic conditions point toward a similar direction (i.e., SR-SR and SR-N), overall RTs and QTs are expected to be shorter for either construction. In addition, parallel models expect diverse RT patterns across positions between two constructions. Longer RTs at word 5 in the SOR construction may or may not be observed across the three plausibility conditions depending on the combination of the plausibility bias and syntactic construction information. For example, the largest RT difference is more likely at word 5 in SOR-SR than in SOR-N and SOR-SOR because of incompatible plausibility and syntactic information in SOR-SR.

In terms of the comparison of QTs and QAs between the two constructions, longer QTs and lower QA rates are generally more likely for SOR constructions than for SR constructions based on SR constructions’ structural advantage over SOR constructions. On the other hand, the structural advantage of SR constructions may or may not be reflected in the QT and QA comparisons in the three plausibility conditions depending on the combination of the plausibility bias and syntactic construction information.

In terms of the effect of WM capacity on the ambiguity resolution process, readers with a high level of WM capacity may be able to integrate more information and allocate attention appropriately to the various information sources than those with a low level of WM capacity and are thus expected to read sentences faster. However, in conflicting conditions, they may read sentences slower because they may pay close attention to all available information. On the other hand, it may take longer for high-span readers to read sentences because holding more information is likely to incur some processing cost.

For the plausibility effect with respect to WM span differences, as shown in previous L1 research, native speakers of Korean with a high level of WM capacity are expected to show stronger plausibility effects. On the other hand, they may also show weaker plausibility effects.
because they are better at reanalyzing sentences upon encountering implausibilities. Nevertheless, low-span native speakers are expected to be less sensitive to plausibility information than high-span ones.

In general, non-native speakers are expected to spend more time reading sentences and questions because of their limited linguistic proficiency and rely more on non-syntactic information during parsing. In this regard, the nativeness of HSs’ parsing with that of native speakers of Korean and L2 learners of Korean was compared to examine whether early and continued exposure to the language is an important factor in native-like processing even when the exposure to the language is limited and constrained. RTs and RT patterns for comprehension questions were analyzed to determine whether the plausibility of comprehension questions influences the two groups of non-native speakers relatively more than native speakers.

If native speakers show some WM span effect for plausibility information, then processing patterns for high-span HSs and/or L2 learners may be the same as or similar to those for native speakers. On the other hand, low-span HSs and L2 learners may be least likely to show native-like processing patterns or any span effect for plausibility information.
CHAPTER 4

METHOD

4.1. Participants

There were three groups of participants for the main experiment: native speakers of Korean (NSs), Korean heritage speakers (HSs), and L2 learners of Korean with different L1s (K2s). There were 40 NSs, 40 HSs (20 for the SR syntactic list and 20 for the SOR syntactic list for each group), and 42 K2s (21 for the SR syntactic list and 21 for the SOR syntactic list for each group). In addition, to norm the materials, 26 NSs were recruited for a web-based survey. WM capacity is known to deteriorate with age (Park, Welsh, Marschuetz, Gutchess, Mikels, & Polk, 2003), and the processing speed is known to be higher for those in their mid-twenties and above (Oberauser, Wendland & Kliegl, 2003; Sagarra, 2008). Hence, 50 was set as the maximum age.

NSs and HSs were composed mainly of undergraduate/graduate students at the University of Illinois at Urbana-Champaign and family members of students/visiting scholars from Korea. These participants were recruited from psychology/educational psychology subject-pools, Korean-language classes offered at the University of Illinois at Urbana-Champaign, and Korean churches in Korean-American communities through the use of emails, phone calls, and fliers (posted in campus buildings).

K2s were those in Korea learning Korean as a second or foreign language at the time of the experiment. They were either international students attending a Korean-language institute at Kangwon National University and Hankuk University of Foreign Studies or foreign spouses of Korean in Kangwon province. International students were recruited from Korean-language classes offered by the two universities through fliers, which were posted on bulletin boards.
inside the institution. Foreign spouses were mainly from East Asian countries such as China and Japan and were recruited from the Hong Cheon-Gun Multicultural Family Support Center in Kangwon-do, Korea, which offered various classes in the Korean language and culture to multicultural families in the region.

Each participant received either PSYC/EPSY course credit or cash ($10-$20 for NSs and HSs and $20-$25 in the Korean won for K2s) for participating in the study. The cash amount varied according to the amount of time they spent.

To norm the materials in a pilot study in 2008, a total of 26 NSs were recruited for a web-based survey. Most of these NSs were short-term visiting scholars from Korea who were participating in a workshop during their summer break. Given that the author resided in the U.S. and had limited access to “pure” monolingual speakers of Korean in the area, these NSs, who were in the U.S. for less than three years, were recruited for the NS group. All the participants in the NS group ($M = 24.78$ years old, $SD = 6.62$; $Range = 18 - 50$; 15 females) were Korean-English bilinguals but spoke Korean as their first and dominant language. None had been diagnosed with a reading disability, and all had normal or corrected-to-normal vision. These participants started learning English as a foreign language in Korea through formal instruction ($M: 10.77$ years, $SD = 3.83$; $Range = 0 - 20$). Their parsing performance was examined to investigate their use of multiple information sources with respect to WM during the ambiguity resolution process and thus test the two sentence-processing models. Their parsing performance was used as the baseline condition for a comparison between NSs and HSs/K2s.

In addition, 40 HSs ($M = 20.92$ years old, $SD = 2.78$; $Range = 18 - 35$; 27 females) and 42 K2s ($M = 28.33$ years old, $SD = 9.19$; $Range = 19 - 48$; 35 females) with good reading skills in

---

3 http://www.eyeinhc.kr/
Korean participated in the experiment as the target groups. All had normal or corrected-to-normal vision. HSs were early Korean-English bilinguals who started learning English since the early stages of their lives ($M = 5.19$ years old, $SD = 4.35$; $Range = 0 - 13$), and their current dominant language was English. Those HSs who spoke Korean regularly since childhood and who showed at least an intermediate level of reading proficiency in Korean were recruited.

The participants in all three groups were asked about their language background and current language use. In addition, they provided a self-assessment of their overall proficiency in their languages on a Likert-type scale ranging from “very poor” (1) to “native command” (5). Self-assessment reports about language use and proficiency provide an index of language dominance (Flege et al., 2002).

HSs were asked to complete two proficiency tests—one in Korean (20 items) and the other in English—whereas NSs and K2s took only the Korean-proficiency test. Although both proficiency tests were cloze tests in their format, they were different in terms of the level of difficulty. The English-proficiency test was a cloze test with 40 blanks and different categories of linguistic knowledge and was used to evaluate whether HSs had native-like English proficiency. This test has been used in other L2 studies (e.g., Kim, 2008). On the other hand, the Korean-proficiency test was a cloze test with 20 blanks and 3 answer choices for each blank, given that Korean was not the dominant language for HSs and K2s (Kim, 2008).

HSs tend to have wide range of language experience and use. Despite their regular Korean use with their family members and others, HSs (early simultaneous bilinguals and Korean Americans) typically vary widely in terms of their fluency in Korean (Kim, 2005). In the present study, most of the HSs immigrated to the U.S. before puberty (13 years old) or were born in the U.S. To ensure that they had at least an intermediate level of reading proficiency in
Korean, several screening measures were employed, including a cloze test of reading proficiency, a WM test, and a self-reported proficiency rating. Based on the self-assessment and language proficiency test results, those participants with scores that were too low were excluded from the analysis.

To participate in the study, HSs had to rate themselves at least 2 out of 5 in the self-assessed Korean-proficiency survey, get at least 12 out of 20 items correct in the cloze test for Korean proficiency \((M = 16.65, SD = 2.92)\), and get at least 21 out of 42 items correct in the WM RST semantic acceptability test \((M = 37.25, SD = 3.84)\).

HSs self-rated their Korean and English to be good or advanced \((M = 4.02 \text{ in Korean and } 4.40 \text{ in English})\) and reported that they used English more frequently \((M = 71.25\% \text{ of the time})\) than Korean \((28.75\% \text{ of the time})\). In addition, they spoke Korean only with their family members and friends at the time of the test. Their English proficiency was also tested through a cloze test \((M = 35.68 \text{ out of } 40 \text{ items, } SD: 2.40)\). Although Korean was their first language, they were exposed to English early in their lives \((M = 5.19, SD = 4.35; \text{ Range } = 0-13)\). At the time of the test, they felt more comfortable with English, which they reported was their dominant language \((M: 4.40 \text{ out of } 5 \text{ as native-like, } SD: 0.71)\), and used it more frequently than Korean in daily communication.

Foreigners living in Korea who showed high-intermediate/advanced levels of reading proficiency in Korean were recruited through those proficiency-screening measures used for HSs. A total of 32 out of 42 K2s (25 Chinese, 6 Japanese and 1 Bengali) were international students who had been taking Korean-language classes at the language institute, and their proficiency levels were either 5 or 6 (levels ranged from 1-6, with 6 being the highest level). The remaining 10 K2s comprised immigrants from East Asian countries (7 Japanese, 2 Chinese, and 1 Thai) and
were married to Korean men. These foreign spouses had lived in Korea for at least five years ($M$: 12.11 years, $SD$: 3.99). International marriages between women from Southeast Asia and men from rural farming communities in Korea have recently become a major pattern of international marriages in Korea (Kim & Shin, 2007). Unlike international students at the language institute, these foreign spouses learned Korean informally by communicating with their family members and neighbors, watching TV shows, and taking vocational classes at community centers.

K2s self-rated their Korean to be good or advanced ($M = 3.52$, $SD = 0.71$). The screening criteria were the same as those for HSs (except for lower scores on the Korean-proficiency cloze test). Because of their low proficiency in Korean, they had to get at least 10 out of 20 items correct in the cloze test of Korean proficiency ($M = 14.29$, $SD = 2.80$). They reported that they used Korean more frequently ($M = 57.62\%$ of the time, $SD = 27.99$) than their first language. In addition, they were exposed to Korean in their early twenties ($M = 21.64$, $SD = 7.66$, Range: 1-44). On average, they had been learning Korean for approximately seven years at the time of the test ($M = 5.56$ years, $SD = 5.26$).

Those who did not meet the criteria described above were excluded from the analysis. Originally, 48 and 52 participants were recruited for the HS group and the K2 group, respectively, but only 40 and 42, respectively, met the study criteria. Table 4 provides descriptive statistics for the participants.
Table 4

Descriptive Statistics for Participants by Group (Means)

<table>
<thead>
<tr>
<th></th>
<th>Native Speakers</th>
<th>Heritage Speakers</th>
<th>L2 Learners</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (total: 102)</td>
<td>40</td>
<td>40</td>
<td>42</td>
</tr>
<tr>
<td>Syntax List (SR vs. SOR)</td>
<td>SR: 20</td>
<td>SR: 20</td>
<td>SR: 21</td>
</tr>
<tr>
<td></td>
<td>SOR: 20</td>
<td>SOR: 20</td>
<td>SOR: 21</td>
</tr>
<tr>
<td>Age (SD, Min-Max)</td>
<td>24.78 (6.59, 18-51)</td>
<td>20.92 (2.78, 18-35)</td>
<td>28.33 (9.19, 19-48)</td>
</tr>
<tr>
<td>L1</td>
<td>Korean</td>
<td>Korean</td>
<td>Chinese, Japanese, Thai and Bangla</td>
</tr>
<tr>
<td>L2</td>
<td>English or others</td>
<td>English or others</td>
<td>Korean</td>
</tr>
<tr>
<td>Age of L2 Exposure (SD, Min-Max)</td>
<td>10.77 (3.83, 0-20)</td>
<td>5.19 (4.35, 0-13)</td>
<td>21.64 (7.66, 1-44)</td>
</tr>
<tr>
<td>Length (year) of L2 Learning (SD, Min-Max)</td>
<td>NA⁴</td>
<td>5.64 (3.64, 0-14)</td>
<td>5.56 (5.26, 0.60-18)</td>
</tr>
<tr>
<td>Korean Proficiency (SD) - 20 items</td>
<td>19.47 (0.93)</td>
<td>16.65 (2.92)</td>
<td>14.29 (2.80)</td>
</tr>
<tr>
<td>Self-Report on Korean Proficiency (5= native-like (SD))</td>
<td>5 (0.00)</td>
<td>4.02 (1.00)</td>
<td>3.52 (0.71)</td>
</tr>
<tr>
<td>Percentage of Korean Spoken (SD)</td>
<td>65.88% (16.00)</td>
<td>28.75% (15.80)</td>
<td>57.62% (27.99)</td>
</tr>
<tr>
<td>English Proficiency (SD) - 40 items</td>
<td>NA</td>
<td>35.68 (2.40)</td>
<td>NA</td>
</tr>
<tr>
<td>Self-Report on English Proficiency (5= native-like (SD))</td>
<td>NA</td>
<td>4.40 (0.71)</td>
<td>NA</td>
</tr>
<tr>
<td>WM RST – Memory (SD)</td>
<td>33.58 (5.08)</td>
<td>30.87 (5.49)</td>
<td>29.09 (6.91)</td>
</tr>
<tr>
<td>WM RST – Semantic Judgment (SD)</td>
<td>40.46 (1.78)</td>
<td>37.25 (3.52)</td>
<td>32.24 (4.51)</td>
</tr>
</tbody>
</table>

Tables 5 and 6 show the means of Korean proficiency test results and the RST scores, respectively, by group (NSs/HSs/K2s) as well as by syntactic construction (an SR construction with two objects vs. an SOR construction with one object). The participants in all groups took an RST as a measure of their WM capacity. The RST was conducted in Korean based on a self-paced reading paradigm. The participants read sentences aloud and retained sentence-final one-

⁴ Although it is possible that L2 (English for Korean native speakers) could influence processing in Korean, I did not report their length of L2 learning here because neither L2 influence on Korean processing nor L1 attrition by Korean native speakers was the focus of the present study. On average, Korean native speakers start English learning at school at the age of 10 in the elementary school.
syllable letters in Korean for subsequent recall in the RST with 42 items. Approximately half of the sentences in the RST were plausible, and the remaining sentences, semantically implausible.

Table 5

*RST Scores by Group/Syntactic Construction*

<table>
<thead>
<tr>
<th>Syntactic Construction</th>
<th>Native Speakers (n=40)</th>
<th>Heritage Speakers (n=40)</th>
<th>L2 Learners (n=42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR (n=20)</td>
<td>SR (n=20)</td>
<td>SR (n=20)</td>
<td>SOR (n=20)</td>
</tr>
<tr>
<td>Mean</td>
<td>32.55</td>
<td>30.40</td>
<td>30.04</td>
</tr>
<tr>
<td>SD</td>
<td>5.94</td>
<td>5.43</td>
<td>5.16</td>
</tr>
<tr>
<td>Range</td>
<td>22-40</td>
<td>21-40</td>
<td>17-37</td>
</tr>
</tbody>
</table>

Table 6

*Cloze Test Scores for Korean Proficiency by Group/Syntactic Construction*

<table>
<thead>
<tr>
<th>Syntactic Construction</th>
<th>Native Speakers (n=40)</th>
<th>Heritage Speakers (n=40)</th>
<th>L2 Learners (n=42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR (n=20)</td>
<td>SR (n=20)</td>
<td>SR (n=20)</td>
<td>SOR (n=20)</td>
</tr>
<tr>
<td>Mean</td>
<td>19.60</td>
<td>17.15</td>
<td>14.57</td>
</tr>
<tr>
<td>SD</td>
<td>0.82</td>
<td>2.70</td>
<td>2.73</td>
</tr>
<tr>
<td>Range</td>
<td>17-20</td>
<td>11-20</td>
<td>10-19</td>
</tr>
</tbody>
</table>

There were significant differences between the three groups of participants in terms of their scores on the Korean-proficiency cloze test (*F* (2, 114) = 58.71, *p* < .01). HSs scored significantly lower than NSs (*p* < .01), and K2s scored significantly lower than HSs and NSs (*p* < .01). However, there was no significant difference in Korean-proficiency scores in terms of the two syntactic lists (the SR construction vs. the SOR construction) in the three groups (*F* (2, 114) <1).

---

5 Only 12 of 40 Korean native speakers took the Korean cloze test. These scores are calculated assuming that those who did not take the test would have received perfect scores of 20.

6 For multiple pairwise comparisons, the Bonferroni correction was applied.
There were significant differences between the three groups in terms of WM semantic judgment scores (processing component) \( (F(2, 114) = 65.09, p < .01) \), but no difference in terms of syntactic lists \( (F(1, 114) < 1) \). NSs (accuracy rate: 96.88%) scored higher than HSs and K2s (91.07% for HSs and 87.95% for K2s) \( (p < .01) \), and HSs scored higher than K2s \( (p < .01) \). In addition, all three groups showed different levels of WM capacity in terms of WM RST memory scores \( (F(2, 114) = 5.42, p < .01) \), but there was little difference in terms of the syntactic list \( (F(1, 114) < 1) \). K2s scored lower than NSs and HSs \( (p < .01) \).

![Figure 1. WM RST memory score distribution by Group.](image)

There was some correlation between self-assessed ratings and Korean-proficiency test scores \( (r = .46, p < .01) \). There was no correlation between WM capacity—measured through the RST (memory component)—and Korean-proficiency test scores \( (r = .23, p > .05) \), but there was a substantial correlation between WM semantic judgment scores (processing component) and Korean-proficiency scores \( (r = .65, p < .01) \). Although there was a significant group difference in WM semantic judgment scores, the participants’ overall accuracy rates for WM semantic judgments exceeded 87%, indicating that they paid attention to the meaning of the sentence. This
suggests that the processing component of the WM RST may be closely related to language proficiency but that the storage component of the WM RST may not be strongly language-specific.

4.2. Materials

To investigate the roles of syntactic preferences and semantic plausibility in sentence processing, two types of RC constructions in Korean—SR and SOR constructions (Hirose, 2000; Hirose & Inoue, 1998; Kim, 2004; Mazuka & Itoh, 1995)—were considered under three semantic plausibility bias conditions (SR-biased, Neutral, and SOR-biased). Nine sentences for each condition (3 x 9: a total of 27) and some filler sentences were obtained from Kim (2004), who specified very detailed information on the process of plausibility norming. A small number of items from Kim (2004) had to be adjusted for their plausibility, and six new sentences for each experimental condition were added through pretests. The Appendix lists these sentences.

To create additional experimental materials and check the plausibility of some of the items from Kim (2004), separate norming procedures were employed by using sentence fragments composed of an object noun and a verb and rating the plausibility of the object noun as the theme of the verb. Two norming studies and an internet corpus search were conducted as pretests to control for the degree of semantic plausibility of the experimental items. Sentence fragments with an object noun and a verb pair were collected by using internet search engines in Korean (e.g., http://www.google.co.kr). Blair, Urland, and Ma (2002) outlined some advantages of using internet-based corpora for conducting a norming study and how it can be done. The authors also claimed that internet search engines with large databases can provide accurate frequency estimates for phrases as well as individual words. Thus, Google was used to select noun-verb pairs for the plausibility-norming study. Approximately 300 noun-verb pairs in
Korean were searched via Google and their frequencies were compared. As a result, 228 noun-verb pairs (half-higher hits and half-lower hits) were selected for a personal rating study to determine their plausibility.

In the rating study (a web-based survey for convenience), the participants were asked to evaluate the extent to which the 228 noun-verb pairs were plausible and natural on a six-point Likert-type scale ranging from “sounds implausible and awkward” (1) to “sounds plausible and natural” (6). A higher score implied that the pre-verbal noun was a more plausible direct object of its following verb. Nouns were classed into plausible and implausible direct objects according to the ratings. However, those nouns with the lowest ratings were excluded to maintain a certain level of ambiguity balance. This approach provided an approximation of how plausible the noun-verb pair sounded. A total of 26 NSs (16 visiting scholars from Korea for three weeks in the U.S. for the first norming and 10 Korean graduate students from the University of Illinois for the second norming) participated in the experiment and were paid $10 for their time. The web version of the norming study facilitated data collection in terms of accessibility and participant anonymity. Table 7 shows the mean plausibility scores from the norming tests for the noun-verb pairs in each condition.

Table 7

<table>
<thead>
<tr>
<th>Norming results</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR-biased</td>
</tr>
<tr>
<td>1st Obj + RCV</td>
</tr>
<tr>
<td>2nd Obj + MV</td>
</tr>
<tr>
<td>SR construction</td>
</tr>
<tr>
<td>5.61</td>
</tr>
<tr>
<td>5.72</td>
</tr>
<tr>
<td>Obj + RC Verb</td>
</tr>
<tr>
<td>Obj + Main Verb</td>
</tr>
<tr>
<td>SOR construction</td>
</tr>
<tr>
<td>(5.3 in the 1st norming)</td>
</tr>
</tbody>
</table>
In the SR-biased condition, the first object and the embedded RCV in the SR syntactic construction had a plausible relationship (5.61), but the object noun and the MV in the SOR syntactic construction had an implausible relationship (2.76). According to the first rating study, the noun-verb pairs from the SOR-SR condition, including the sentences from Kim (2004), showed a 5.3 plausibility rating, indicating a highly plausible relationship. Thus, another rating study with new noun-verb pairs from internet searches was conducted with sentences from only the SOR construction and obtained a 2.76 rating for the SOR-SR condition. Thus, the lower plausibility rating between the first noun object and the MV in the SOR-SR condition was expected to lead the participants to the SR analysis under the constraint-based account.

In contrast, in the SOR-biased plausibility condition, the first object and the embedded RCV in the SR syntactic construction had a less plausible relationship (3.69) than the object and the MV in the SOR syntactic construction (5.61). This less plausible combination of object + RCV was expected to increase reading times around the RCV and induce the participants to the SOR analysis under the constraint-based account.

In the neutral condition, the plausibility scores for all condition were similar (5.62, 5.77, and 5.77 for 1st object + RCV, 2nd object + MV, and 1st object + MV, respectively). In Kim (2004), syntactic preferences for the minimal structure appeared only in the neutral condition, and sentence processing in the neutral condition was faster than that in all other conditions.

Based on the results of the norming tests (excluding extreme cases, i.e., a 6 rating), three semantic plausibility bias sets for each of the two syntactic constructions were constructed: a set semantically biased toward the SR analysis, a set biased toward the SOR analysis, and a semantically neutral set. 15 sentences were employed for each condition (nine sentences from
Kim (2004) with slight modifications based on the plausibility rating results plus six new sentences). Two versions of each sentence were written, which differed only with respect to the last two words (the second object + the MV in the SR construction and two MVs in the SOR construction). Because the syntactic construction was manipulated as a between-subject factor and plausibility conditions as a within-subject factor, each participant read sentences with either SR or SOR constructions. Given that there were more filler sentences with relative and non-RC constructions than those sentences for the main experiment, the syntactic construction (either SR or SOR) was expected to have less distinctive priming effect on experiment results. In other words, many RC and non-RC constructions included in the experiment were expected to reduce the likelihood of finding structural effects for the types of RCs of the experimental targets. The two versions of each sentence did not have exactly the same length of syllables, and the frequency of each word was not controlled for. A one-way ANOVA was conducted with data on the frequency and number of syllables in the disambiguating region and no significant difference was found in word frequency at word 5 between the second noun object in the SR construction and the verb in the SOR construction ($F(1, 88) = 2.95, p = .09$). However, there was a significant difference in the number of syllables at word 5 between the two constructions ($F(1, 88) = 25.65, p < .01$). As a result, averaged residual RT data were calculated and (instead of original RTs) used for the analysis of RT patterns to adjust for the length effects of syllables.

Two presentation lists (List A for SR constructions and List B for SOR constructions) were constructed based on syntactic condition information and were presented visually to the three groups of participants. Each participant was required to read sentences with only one construction (either SR or SOR) under three plausibility conditions. A total of 174 sentences (45
for each set of constructions and 129 filler sentences) were employed for the main experiment.

Table 8 shows the numbers of sentences used for the main experiment.

**Table 8**

*Number of Sentences for Each Condition*

<table>
<thead>
<tr>
<th>Condition</th>
<th>SR-biased</th>
<th>Neutral</th>
<th>SOR-biased</th>
<th>Subtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR syntactic construction (List A)</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>45</td>
</tr>
<tr>
<td>SOR syntactic construction (List B)</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>45</td>
</tr>
<tr>
<td>Filler 1 (RC without ambiguities)</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>45</td>
</tr>
<tr>
<td>Filler 2 (RC with an ambiguous adverbial phrase)</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>27</td>
</tr>
<tr>
<td>Filler 3 (RC without an ambiguous adverbial phrase)</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>27</td>
</tr>
<tr>
<td>Filler 4 (no RC without ambiguities)</td>
<td>30</td>
<td></td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

129 filler sentences were constructed by using various syntactic forms. Approximately 75% of these filler sentences (Fillers 1, 2, and 3) contained RCs, and approximately 30% (Filler 1) were structurally similar to the experimental ones. However, RCs in the filler sentences were not ambiguous in terms of plausibility.

The participants in all three groups saw only one set of sentences from the syntactic construction (either A or B) in addition to all the same fillers. The main focus was to test individual variations in the use of plausibility information in the same syntactic construction, not to compare processing behaviors through different syntactic constructions under the same plausibility condition. For this, two syntactic lists of items were developed for the self-paced reading task, each of which contained three plausibility conditions. Semantic plausibility was a within-subject condition, whereas SR/SOR syntactic construction were a between-subject condition.
All experimental sentences were RCs that were temporarily ambiguous up to the appearance of the disambiguating word (word 5). Three features of Korean syntax are particularly relevant to this study. First, the MV occurs in the clause/sentence-final position. Second, RCs pre-modify NPs. Third, scrambling is allowed in Korean, which is a pro-drop language for both subjects and objects.

In this study, SR and SOR syntactic constructions had the same sentence format until word 5 at the disambiguating region (total of six word positions). In the SR construction with two objects, the second object, which was an accusative NP, was followed by an MV, whereas the first element of the double-verb construction as an MV appeared at word 5 in the SOR construction. In Korean, the verb in the embedded RC can be both transitive and intransitive, but all MVs are transitive, requiring a direct object. All the experimental sentences had an embedded reduced RC. Thus, except for the first two word positions, the remaining four word positions, including the disambiguating region at word 5, played key roles in the detection of earlier or later plausibility effects. The Appendix lists all the sentences used in the experiment (including fillers).

4.3. Procedures

The experiment was conducted in three phases: i) surveys for the language background and the proficiency test, ii) the RST, and iii) the main reading time experiment. The surveys were administered first for all three groups, which included questions about their language background and current language use and a self-assessment of their overall proficiency in their languages. Approximately half of the participants in the NS and HS groups took the RST before the main experiment, and the rest participated in the main experiment before the RST. All the participants
in the K2 group took the RST before the main experiment because of the pre-screening process through the RST.

Based on the survey and proficiency test results, the participants were assigned to three groups: the NS, HS, and K2 groups. Only HSs took both Korean- and English-proficiency tests in the cloze format. The English-proficiency questionnaire and the Korean version of the RST for WM measurement were from Kim (2008)\textsuperscript{7}.

In the main experiment, a self-paced moving-window reading task with end-of-sentence comprehension questions was administered as well as an RST for measuring WM capacity. Self-paced reading is one of the most widely used online techniques in psycholinguistics for investigating what information is used during sentence parsing (Garrod, 2006).

All the participants in the study took a Korean RST as a measure of their WM capacity. The RST scores were employed as an indicator of their cognitive capacity relative to their reading skills in Korean (Korean version created by Kim (2008)), which have been shown to influence the use of non-syntactic information during sentence processing. As mentioned in Conway et al. (2005), there are various scoring methods for this test. For scoring, partial credit was given for incomplete sets, as suggested in Conway et al. (2005). For example, if only 2 out of 5 items were recalled correctly, then .40 would be given to this set, not 0, to obtain a better reflection of the actual storage capacity (Conway et al., 2005; Rodriguez, 2008). RST storage scores were included in statistical models as a covariate.

The RST, a WM task, was administered by using a PC station running on experimental software E-prime (version 1.2) (Rhode, 2001: Psychology Software Tools, Inc., Pittsburgh, PA). In the RST, a total number of 42 sentences were presented on the computer screen in sets of

\textsuperscript{7} Gratitude must be extended to Ji Hyon Kim (Kim, 2008) for these two tests.
varying sizes (2-5). The session started with instructions and three practice sessions. The participants were provided with an answer sheet so that they could write down the words they recalled. The participants each sentence aloud with an irrelevant Korean syllable (one consonant + one vowel from the basic alphabet system) and determined its logical accuracy. During the test, the researcher sat next to each participant and recorded his or her semantic responses to each item. There were 12 sets whose sentence length varied from 5-10 words. The RST scores were calculated as the total number of sentence-final letters recalled during the whole test (maximum of 42 letters) instead of using the set/unit size calculation method to obtain a wider range of RST scores. Equal span distribution in each syntactic condition was not pre-controlled because of time constraints but was confirmed through posterior statistical analyses. Total administration time for the RST varied from 15 minutes (for NSs) to 30 minutes (for HSs and K2s).

A self-paced moving-window reading task with end-of-sentence comprehension questions was administered in the main experiment by using Presentation Software V14.0 (NeuroBehavioral Systems Inc., Albanay, CA). Each sentence was presented with one word at a time in the position it would typically occupy in the sentence on the computer screen. A total of 174 sentences (45 experimental sentences for each construction and 129 filler sentences) were used. Each sentence was preceded by a sentence number appearing in the center of the computer screen. The participants pressed the spacebar to reveal each word of the sentence, which included six words for experimental sentences and 5-6 words for filler sentences.

When each new word appeared, the preceding word disappeared. The amount of time spent on each word was recorded as the amount of elapsed time between keystrokes. The keystroke following the final word of the sentence initiated the display of a question about the content of the sentence, which the participants were required to answer using the key labeled
“yes” (the shift key on the left) or “no” (the shift key on the right). The participants were told to read each sentence and answer this question as quickly and accurately as possible. Comprehension questions were designed to keep the participants focusing on the meaning of sentences while reading and to prevent them from rapidly passing over words just to complete the task.

For the main experiment, a desktop computer in a lab (room 2437 of the Beckman Institute) or a laptop in a quiet location on campus were used. Both computers had the same screen resolution (1024 x 768) and a screen refresh rate of 60 Hz, although the screen size was different. The experiment (including the surveys, cloze tests, practice trials, instructions, trials, and de-briefing sessions) took approximately 1-2.5 hours depending on the participants’ reading skill. Each session was individually administered by the researcher. The participants were given short breaks between sessions upon their request.

4.4. Data Analysis

To address this study’s research questions, two main analyses were employed: one for analyzing structural preferences and the other for the plausibility effect. In these analyses, the reading time (RT: overall and for each position), the response time for comprehension questions (QT), and the accuracy of comprehension questions (QA) were the dependent variables, and the WM RST score was a covariate.⁸

In terms of RTs for critical words, different trimming criteria were employed for the three groups: 200/2000 msec (data between 200 ~ 2000 msec included) for NSs and 200/3000

---

⁸ Group specification was a strong predictor of Korean proficiency ($r = -.68, p < .00$). The participants with a high level of Korean proficiency tended to be NSs (0), and those with a low level of Korean proficiency tended to be K2s (2). Approximately 50% of the change in the level of Korean proficiency was explained by the change in the specification of group ($R^2 = .47$). Group and Korean-proficiency scores were strongly related to each other, and thus, we omitted Korean-proficiency scores from the statistical analysis to better predict models. On the other hand, group accounted for only 3% of the variance in WM scores ($R^2 = .027$), although it had a negative correlation with WM scores ($r = -.19, p = .054$).
msec (data between 200 ~ 3000 msec included) for HSs and K2s. This influenced less than .8% of the data. Two participants (one NS from the SOR construction list and one HS from the SR construction list) lost more than 70% of their RT data in most of the test conditions after the trimming process and were thus excluded from further analysis. RTs were then corrected for differences in the word length (the number of characters, which equals the number of syllables in Korean) across conditions. Residual RTs were calculated by estimating a separate linear regression for each participant, with their raw RTs as the dependent variable and the number of syllables in words as the independent predictor variable. Only the experimental items were included in these regressions. Predicted RTs for each word length were derived for each participant and then subtracted from original RTs, producing length-corrected residual RTs (Ferreira & Clifton, 1986; Trueswell, Tanenhaus & Garnsey, 1994). Here a positive residual RT meant that the RT was longer than predicted (given the length of the word), whereas a negative residual RT meant that the RT was shorter than predicted.

To compare the performance of the three groups in terms of the interaction effect of plausibility and structural preferences on early processing and final interpretations (NSs vs. HSs and NSs vs. K2s), a repeated measures ANOVA was conducted using RT data (RT and RT at each position) organized by participants. Here syntactic constructions (SR vs. SOR syntactic construction lists) and groups (NSs vs. HSs and NSs vs. K2s) were between-subject factors, and plausibility (SR-biased plausibility, SOR-biased plausibility, and Neutral for both syntactic conditions) and word positions (6 word levels) were within-subject factors. In addition, the WM RST score (a measure of WM capacity) was a covariate.

A repeated measures ANOVA was conducted by using RT organized by items, with plausibility as a between-subject factor and syntactic constructions/word positions as within-
subject factors. QTs and QA rates for comprehension questions for the three groups were compared through a repeated-measure ANOVA. In addition, generalized estimating equations (GEEs) were employed to address the dependency in the RTs and QA data organized by participants resulting from repeated measures. For brevity, however, only those results of subject-based analyses based on the repeated measures ANOVAs are reported in the following sections when the results are similar.
CHAPTER 5

RESULTS AND DISCUSSION

5.1. Native Speakers

The results are discussed for several whole-sentence measures before moving on to RTs at critical word positions. These whole-sentence measures include accuracy (QA) and response times (QTs) for the post-sentence paraphrase judgment task as well as whole-sentence RTs generated by summing trimmed RTs for all words in the sentence. The manipulation of the plausibility bias was expected to have no main effect on these whole-sentence measures because plausibility was manipulated with respect to particular structures and all sentence-processing models assume that plausibility ultimately influences comprehension. For all the whole-sentence measures, the plausibility bias interacted with the structure, and thus, to streamline the reporting of results, it is not reported whether or not the main effect of plausibility bias was significant in the analyses of the global measures because any main effect would have been due to some elements of the interaction. However, in a later section discussing RTs at words 3 and 4, i.e., the critical words preceding the word disambiguating the structure, the plausibility bias was expected to have a main effect and, thus, whether there was such an effect is reported.

**Accuracy of Paraphrase Judgments and Response Times**

In all but one of the conditions (SOR-SR), the post-sentence paraphrase was consistent with the structure that the sentence turned out to have, which was determined completely by whether word 5 was a second object-marked noun (in the SR structure) or an obligatorily transitive verb (in the SOR structure). If the participants responded to paraphrases on the basis of the ultimate sentence structure, they were expected to provide “yes” answers in most of the experimental trials. However, for the SOR-SR condition, which was expected to be the most
difficult condition because the sentence started out to be consistent with the preferred SR 
structure but then became disambiguated to a more difficult SOR structure, we decided to use the 
paraphrase that would have been correct if the sentence turned out to have an SR structure. Thus, 
in that one condition, the syntactically appropriate response to the paraphrase was a “no” answer. 
The idea was to specifically probe whether the participants maintained the SR structure in that 
condition, i.e., a “yes” answer.

Table 9 shows the accuracy of paraphrase judgments (QA). NSs were generally more 
accurate for paraphrases with the SR construction than with the SOR construction ($\chi^2 = 7.58$, $df = 1$, $p < .01$), which provides support for the argument that there is an overall preference for the SR 
structure. Many (but not all) of the incorrect responses for the SOR construction were for the 
SOR-SR condition, which was the only condition whose correct response was a “no” answer. 
When a participant incorrectly provided a ”yes” answer for the SOR-SR condition, he or she was 
not only being generally wrong but also specifically endorsing the paraphrase that would have 
been correct if the sentence had the SR structure, which provides additional support for a general 
preference for the SR structure. In addition, there was an interaction between plausibility and the 
structure ($\chi^2 = 77.49$, $df = 2$, $p < .01$). In the Neutral and SR-biased plausibility conditions, the 
participants showed higher QA rates for the SR construction than for the SOR construction ($p < .01$), whereas they showed similar QA rates for SR and SOR constructions in the SOR-biased 
plausibility condition ($p < .01$). For the SOR structure, the participants showed higher QA rates 
in the SOR-based plausibility condition than in the other two plausibility conditions ($p < .05$) and 
higher QA rates in the Neutral condition than in the SR-biased plausibility condition ($p < .05$). 
This indicates that the SOR-biased plausibility of the SOR construction made it easier for the 
participants to understand the SOR analysis in their final interpretation.
For both constructions, accuracy was lowest when the structure did not match the plausibility bias, but this was especially true when SR-biased plausibility contexts were followed by disambiguation to the SOR structure. (However, care must be taken in interpreting the relative sizes of the mismatch effects, since one of the mismatching conditions (i.e., SOR-SR) required a “no” response while the other (i.e., SR-SOR) required a “yes” response.)

Table 9

Accuracy of Paraphrase Judgments (%)

<table>
<thead>
<tr>
<th>Sentence Structure</th>
<th>SR-Biased</th>
<th>Neutral</th>
<th>SOR-Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR Construction</td>
<td>93.67</td>
<td>96.00</td>
<td>87.00</td>
</tr>
<tr>
<td>SOR Construction</td>
<td>12.63</td>
<td>73.00</td>
<td>80.70</td>
</tr>
</tbody>
</table>

Table 10 shows the results for QTs for the paraphrase judgment task. There was no significant difference in QTs between SR and SOR constructions ($F(1, 37) = 2.95, p = .09^9$), although the participants were 760 msec faster for the SR structure. This nonsignificant difference can be explained by three factors. First, there were substantial differences in paraphrase QTs, particularly for SOR-structure sentences. Second, the sentence structure was manipulated as a between-subject factor, and thus, all these differences contributed to the error term in the analysis of the effect of the sentence structure. Finally, the direction of the difference between SR- and SOR-structure sentences was reversed for one of the plausibility bias conditions, which led to an interaction between plausibility and the syntactic structure ($F(2, 74) = 15.61, p < .01$). For QTs, this interaction arose because bias effects were stronger for the SR structure than for the SOR structure ($F(2, 37) = 26.14, p < .01$ for the SR construction vs. $F < 1$).

---

^9 All $p$-values were adjusted using the Greenhouse-Geisser (G-G) epsilon correction factor for the violation of the sphericity assumption for the repeated measures ANOVA. The corrected $p$-values, together with uncorrected degrees of freedom, are reported.
for the SOR construction), whereas for QA rates, the SOR structure showed stronger bias effects. The longest QTs as well as the lowest QA rates were observed in the condition in which the SOR bias was followed by the disambiguation of the SR structure.

The different bias effect patterns in accuracy and response times make sense, given an initial general preference for the SR structure and differences in the points in the sentences where expectations were violated. In the SOR-SR condition, everything was consistent with the generally preferred SR structure until the appearance of an obligatorily transitive verb at word 5, signaling that the only object-marked noun in the sentence thus far had to be its object rather than the object of the earlier embedded RCV at word 3. Under these circumstances, the participants tended to maintain their SR analysis instead of fully revising to yield the correct SOR structure. The situation was different for the SR-SOR condition, in which the participants found out earlier in the sentence that the sentence was not going to be simple because word 3 was a verb that could not plausibly take the object-marked noun preceding it as its object, thus signaling that another verb should be coming later. Then at word 5, a second object-marked noun appeared, signaling that the earlier object-marked noun had to be the object of the verb at word 3 after all, although it was implausible (though not impossible) in that role. The sentence turned out to have the originally preferred SR structure, and the participants ended up with that interpretation (i.e., high accuracy rates). However, it took them longer to get there because their expectations were violated twice: first at word 3 and again at word 5.

Table 10

*Response Times for Paraphrase Judgments*

<table>
<thead>
<tr>
<th>Sentence Structure</th>
<th>SR-Biased</th>
<th>Neutral</th>
<th>SOR-Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR Construction</td>
<td>3969 (1083)</td>
<td>3900 (784)</td>
<td>5759 (1815)</td>
</tr>
<tr>
<td>SOR Construction</td>
<td>5287 (2032)</td>
<td>5373 (1737)</td>
<td>5245 (1531)</td>
</tr>
</tbody>
</table>

*Mean reading times (SD) in msec.*
Whole-Sentence Reading Times

Table 11 shows the mean whole-sentence RTs, which were generated by summing trimmed RTs for all words in the sentence. In both QTs and the whole-sentence RTs, times were slower in the SR-SOR condition than in the SR-SR and SR-N conditions. On the other hand, whole sentence RTs were slower in the SOR-SOR condition than in the SOR-SR and SOR-N conditions ($F(2, 72) = 5.34, p < .01$), whereas there was no significant RT difference in the QTs for the SOR construction. QTs and total RTs agreed in the absence of a main effect of the sentence structure ($Fs < 1$). They differed, however, in that the QTs were a non-reliable 759 msec faster for SR-structure sentences, whereas total RTs were only 22 msec faster for SR structures.

The results for the global measures indicate that the manipulation of the plausibility bias was successful. The longest QTs and the lowest QA rates were observed for sentence structures that contradicted what was suggested by the plausibility bias earlier in the sentence. For both constructions, the participants took longer to read whole sentences in the SOR-biased plausibility condition than in the Neutral and SR-biased plausibility conditions. The initial implausibility influenced the overall RT for sentences regardless of the difference in the sentence structure.

There was somewhat mixed evidence about whether the SOR structure was more difficult overall. In general, accuracy rates were higher for the SR structure, but neither of the global time measures showed reliably faster times overall for the SR structure. This indicates that the participants did not always revise their incorrect initial interpretation of the sentence. If they tended to go with the SR structure, they would be likely to respond “yes” to paraphrases matching that structure, both when that was correct and when it was incorrect, which is the pattern observed in the accuracy measure. When they failed to revise their initial analysis, they
were not likely to slow down even when they obtained evidence contradicting the SR analysis because it did not trigger reanalysis. In other words, the pattern of results for the global measures is quite consistent with the GE processing and shallow parsing proposals.

Table 11

*Whole-Sentence Reading Times*

<table>
<thead>
<tr>
<th>Sentence Structure</th>
<th>SR-Biased</th>
<th>Neutral</th>
<th>SOR-Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR Construction</td>
<td>3276 (736)</td>
<td>3181 (711)</td>
<td>3519 (837)</td>
</tr>
<tr>
<td>SOR Construction</td>
<td>3303 (695)</td>
<td>3194 (671)</td>
<td>3546 (790)</td>
</tr>
</tbody>
</table>

* Mean reading times (SD) in msec.

*Word-by-Word Reading Times*

Residual RTs were calculated by using trimmed RT data (200~2000 msec) in order to correct for the effect of the word length. Table 12 shows the mean uncorrected RTs for each word, and Table 13 shows the mean length-corrected residual times.

Table 12

*Mean Word-by-Word Reading Times*

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subject</td>
<td>1st Object</td>
<td>RCV</td>
<td>Dative NP</td>
<td>2nd Object/MV</td>
<td>MV</td>
</tr>
<tr>
<td>SR-SR</td>
<td>566.32</td>
<td>543.29</td>
<td>537.22</td>
<td>546.27</td>
<td>551.03</td>
<td>535.34</td>
</tr>
<tr>
<td>SR-N</td>
<td>557.98</td>
<td>552.80</td>
<td>551.88</td>
<td>534.97</td>
<td>487.72</td>
<td>511.12</td>
</tr>
<tr>
<td>SR-SOR</td>
<td>553.05</td>
<td>557.02</td>
<td>589.18</td>
<td>666.82</td>
<td>582.64</td>
<td>563.76</td>
</tr>
<tr>
<td>SOR-SR</td>
<td>586.54</td>
<td>527.63</td>
<td>524.53</td>
<td>511.02</td>
<td>582.01</td>
<td>549.99</td>
</tr>
<tr>
<td>SOR-N</td>
<td>598.23</td>
<td>542.97</td>
<td>511.80</td>
<td>498.22</td>
<td>507.02</td>
<td>516.90</td>
</tr>
<tr>
<td>SOR-SOR</td>
<td>601.03</td>
<td>556.57</td>
<td>577.84</td>
<td>639.07</td>
<td>578.59</td>
<td>566.80</td>
</tr>
</tbody>
</table>

* Mean reading times in msec.
Table 13

Mean Length-Corrected Residual Word Reading Times*

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>1st Subject</th>
<th>1st Object</th>
<th>RCV</th>
<th>Dative NP</th>
<th>2nd Object/MV</th>
<th>MV</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR-N</td>
<td>5.45</td>
<td>16.40</td>
<td>-0.56</td>
<td>-37.74</td>
<td>-58.65</td>
<td>-52.77</td>
</tr>
<tr>
<td>SR-SOR</td>
<td>15.39</td>
<td>19.96</td>
<td>21.17</td>
<td>93.32</td>
<td>29.14</td>
<td>5.70</td>
</tr>
<tr>
<td>SOR-SR</td>
<td>40.31</td>
<td>-10.48</td>
<td>-35.78</td>
<td>-61.04</td>
<td>14.61</td>
<td>-0.07</td>
</tr>
<tr>
<td>SOR-N</td>
<td>45.82</td>
<td>-0.70</td>
<td>-40.62</td>
<td>-63.59</td>
<td>-57.78</td>
<td>-27.86</td>
</tr>
<tr>
<td>SOR-SOR</td>
<td>55.50</td>
<td>11.53</td>
<td>15.23</td>
<td>74.85</td>
<td>14.19</td>
<td>19.26</td>
</tr>
</tbody>
</table>

* Mean reading times in msec.

Figure 2. Mean length-corrected residual word RTs at six word positions for NSs.

Words 3 and 4: The Detection of Implausibility

Plausibility was manipulated at word 3, which was a verb that ultimately turned out to be the verb in the RC in all experimental items. The nature of the plausibility manipulation was that the verb could (SR bias and Neutral bias), or could not (SOR bias), plausibly take the object-marked noun preceding it as its direct object. The participants presumably started out with the assumption that an object-marked noun would be the object of the verb that immediately followed it, as was the case in sentences with a simple SOV structure and also in more complex
SR-structure sentences. When the verb was not plausible in that relationship, as in the SOR-biased condition, the participants would be expected to slow down at the verb and possibly at the next word as well, reflecting the detection of implausibility and possibly the start of a revision of any structural commitment made thus far. Under the self-paced reading paradigm, effects are sometimes delayed a word and often extend to the next word.

Figures 2 and 3 show the mean length-corrected residual RTs at words 3 and 4, respectively. As shown in the figures, RT patterns across the conditions were similar for the two words, although differences grew larger at word 4. This is a typical RT pattern in the self-paced reading paradigm. RTs were significantly longer at word 3 in the SOR-biased condition than in the other two conditions ($F (2, 76) = 8.53, p < .01$), and the same was true at word 4 ($F (2, 76) = 36.79, p < .05$). Thus, the participants slowed down when they detected the implausibility between the object-marked noun at word 2 and the verb at word 3 in the SOR-biased condition.

There was a WM main effect on the reading times at both word 3 and 4, $F (1, 37) = 6.73, p < .05$ and $F (1, 37) = 3.50, p = .07$. The participants with a low level of WM capacity took longer to read words at both positions in the SOR-biased condition than those with a high level of WM capacity ($\beta = -5.77, t (37) = -3.10, p < .01$ at word 3 and $\beta = -5.94, t (37) = -2.08, p < .05$ at word 4). This provides some evidence of WM’s sensitivity to plausibility information.

The next question is what use the participants made of the implausibility they detected at word 3 when they reached word 5, the disambiguating point.
Words 5 and 6: Disambiguation and Revisions

Word 5 was the first different word between the two sentence structures, and it fully disambiguated whether the sentence had an SR structure or an SOR structure. In the SOR structure, word 5 was the MV, and in the SR structure, it was the second object-marked noun.

Figures 4 and 5 show the mean length-corrected residual RTs at words 5 and 6, respectively. As shown in the figures, the pattern of RTs across the conditions was generally similar for the two words. Both subject-based and item-based statistical analyses were performed on each word separately and also for a region measure with both words.
There were no reliable differences between SR and SOR constructions at words 5 and 6 ($F (1, 36) < 1$), although the participants were slower by 29 msec at word 6 when it disambiguated the sentence to the SOR structure. However, the plausibility bias had a significant main effect at both words (word 5: $F (2, 74) = 13.87, p < .01$; word 6: $F (2, 74) = 4.76, p < .01$). This effect was also reliable for the measure combining the two words ($F (2, 74) = 13.58, p < .01$); however, there was no significant interaction between the syntactic structure and the plausibility bias at words 5 and 6 ($F (2, 72) < 1s$).

In terms of the plausibility bias effect, RTs at word 5 and the combined measure were significantly shorter in the Neutral condition than in SR- and SOR-biased plausibility conditions. For instance, in the SR construction, RTs were slower for SR-SR and SOR-SOR than for SR-N ($p < .01$), while they were slower for SOR-SR and SOR-SOR than for SOR-N in the SOR construction ($p < .01$). On the other hand, no such significant reductions in RTs were observed at word 6 ($p > .05$).

When the sentence started out biased toward the SR structure (Neutral and SR-biased plausibility conditions), the obligatorily transitive verb at word 5 in the SOR-structure sentences was the first indication that the sentence was not what it seemed, i.e., that the earlier object-marked noun (word 2) could not be the object of the earlier RCV (word 3) as it had seemed to be, but must instead be the object of the newly arriving MV (word 5). On one hand, the situation was the same at this point for the Neutral and SR-biased plausibility conditions. In both cases, the arrival of an obligatorily transitive verb at word 5 meant that it was no longer possible for the earlier object-marked noun to be the object of the earlier RCV, and in both cases, it was the first indication of that. On the other hand, there was a difference in the situation between the Neutral and SR-biased plausibility conditions. In the SR-biased plausibility conditions, in addition to
making it impossible to maintain the SR analysis, word 5 in the SOR structure made the whole sentence less plausible. That was because the earlier object-marked noun had to be the object of the newly arriving MV, but it was much less plausible in that role than it would have been (and had seemed to be until now) as the object of the earlier verb in the RC. In this condition, participants slowed down substantially at word 5 ($p < .05$). Some of that slowing at word 5 was no doubt due to the decrease in overall sentence plausibility as a result of revising to a less preferred structure when the sentence thus far had been perfectly consistent with the preferred structure, as well as the revision process itself.

In contrast, in the Neutral condition, the newly arriving MV could plausibly take the earlier object-marked noun as its object (just as plausibly as the earlier RCV could), and thus, the sentence remained generally plausible when the structure was revised. Given that the participants had preferred the SR interpretation until word 5, making it necessary to revise their initial analysis in the SOR structure condition, the Neutral and SR-biased plausibility conditions required the same revision. However, in the Neutral condition, a revised analysis was just as plausible as the abandoned one, whereas in the SR-biased condition a revised analysis was less plausible. Under these circumstances in the SOR construction, the participants took less time at word 5 in the Neutral condition than in the SR-biased plausibility condition ($p < .05$). One possible explanation for these differences between the Neutral and SR-biased plausibility conditions is that in the SR-biased plausibility condition (SOR-SR), the revision was harder because it reduced the overall plausibility of the sentence and thus led to larger RT effects than the revision in the Neutral condition (SOR-N), which maintained its plausibility.

When the sentence started out biased against the SR structure because word 3 was a verb that could not plausibly take the object-marked noun preceding it as an object, there was no
reliable difference in RTs at words 5 and 6 between SR and SOR structures \((p < .05)\), although RTs at word 6 were longer by 22 msec for the SOR structure than for the SR structure. This minimal difference between these conditions can be explained in many ways. First, there is one factor that should make the SR-SOR condition slow: The sentence started out biased against the SR structure but then turned out at word 5 to have that structure after all, even though that made the whole sentence less plausible. There is some evidence in the RTs against overall sentence plausibility being the only factor here. The overall more plausible SOR-SOR sentences were about 12 msec faster than the overall less plausible SR-SOR at word 5, whereas the SOR-SOR sentences were about 22 msec slower than the SR-SOR sentences at word 6. With the structural revision cue at word 5, the participants slowed down at word 5 for both SOR-SR and SOR-SOR, but they kept slowing down at word 6 only for SOR-SOR. This suggests that even if the SOR structure really is more difficult, the supporting context may facilitate RTs such that it allows shorter RTs than the simpler SR structure at word 5. This confirms the findings from the global response time and response patterns, which present evidence of strong preferences for the SR analysis as suggested in the GE processing and shallow parsing proposals.

In terms of the WM effect, the participants with a high level of WM capacity showed some sensitivity to plausibility by processing the implausibility faster than those with a low level of WM capacity at earlier words. However, in terms of NSs, individual differences in WM had no effect on any of the other global RTs and accuracy measures.

5.2. Heritage Speakers

**Accuracy of Paraphrase Judgments and Response Times**

Table 14 shows the accuracy of paraphrase judgments. HSs were generally more accurate for paraphrases with the SR construction than with the SOR construction \(\chi^2 = 177.01, df = 1, p\)
which provides support for the argument that there is an overall preference for the SR structure. There was a significant interaction between plausibility and the syntactic structure ($\chi^2 = 100.33, df = 2, p < .01$). In the Neutral and SR-biased plausibility conditions, the participants showed higher QA rates for the SR construction than for the SOR construction ($p < .01$), whereas they showed similar QA rates for the two constructions in the SOR-biased plausibility condition ($p < .01$). For the SR structure, the participants showed the lowest QA rate in the SOR-biased plausibility condition (SR-SOR) compared to the other two plausibility conditions ($p < .05$). On the other hand, for the SOR structure, the participants showed the lowest QA rate in the SR-biased plausibility condition (SOR-SR) compared to the other two plausibility conditions ($p < .05$). This indicates that the SOR-biased plausibility in the SOR construction made it easier for the participants to understand the SOR analysis in their final interpretations (SOR-SOR) while the same plausibility made it more difficult for them to understand the SOR analysis in the SR construction (SR-SOR).

There was a significant two-way interaction of plausibility and WM capacity on the accuracy of paraphrase judgments ($\chi^2 = 15.91, df = 2, p < .01$). High-span HSs were slightly less likely to provide a “yes” answer in the SOR-bias plausibility condition than low-span HSs ($\beta = -.20, t (37) = -3.23, p < .01$, indicating that high-span HSs were more sensitive to the implausibility, similar to native speakers.

Table 14

<table>
<thead>
<tr>
<th>Sentence Structure</th>
<th>Plausibility Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SR-Biased</td>
</tr>
<tr>
<td>SR Construction</td>
<td>84.33</td>
</tr>
<tr>
<td>SOR Construction</td>
<td>12.00</td>
</tr>
</tbody>
</table>
Table 15 shows QTs for the paraphrase judgment task. There was no significant
difference in QTs between SR and SOR sentences ($F (1, 37) < 1$), although QTs were shorter by
520 msec for the SR structure. Regardless of the structural difference, the participants took
longer to respond to paraphrased comprehension questions in the SOR-biased plausibility
condition for both constructions than in the other two plausibility conditions ($F (2, 76) = 23.18, p
< .01$). There was no significant two-way interaction between plausibility and the syntactic
structure ($F (2, 76) = 1.13, p = .33$).

The lowest QA rates were found for the condition in which the SR bias was followed by
the disambiguation to the SOR structure (SOR-SR), whereas the longest QTs were observed for
the condition in which the SOR bias was followed by the disambiguation to the SOR structure
(SOR-SOR). The participants had some difficulty processing initial implausibility for
comprehension questions in the SOR-biased plausibility condition, particularly for SOR-SOR
(even with contextual support for the SOR analysis), although the QA rate for SOR-SOR was
higher than that for SOR-SR and SOR-N ($p < .01$).

WM had significant main effects on QTs in all three plausibility conditions ($F (1, 37) = 12.77, p < .01$). HSs with a low level of WM capacity took longer to read and answer
comprehension questions than those with a high level of WM capacity regardless of the
condition.

Table 15

Response Times for Paraphrase Judgments*

<table>
<thead>
<tr>
<th>Sentence Structure</th>
<th>Plausibility Bias</th>
<th>Plausibility Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR Construction</td>
<td>SR-Biased</td>
<td>Neutral</td>
</tr>
<tr>
<td>SOR Construction</td>
<td>5498 (2501)</td>
<td>6105 (2005)</td>
</tr>
</tbody>
</table>

*Mean reading times (SD) in msec.
Whole-Sentence Reading Times

Table 16 shows the mean whole-sentence RTs. The syntactic structure had no significant main effect on total RTs ($F(1, 36) < 1$), although total RTs were shorter by 316 msec for the SOR structure than for the SR structure. There was a significant two-way interaction effect of plausibility and the syntactic structure on whole-sentence RTs ($F(2, 72) = 5.36, p < .05$), although the 567 msec difference in whole-sentence RTs between SR-SOR and SOR-SOR was not reliable ($p > .05$). Total RTs were longer for SR-SOR than for SR-SR and SR-N and were longer for SOR-SOR than for SOR-SR ($F(2, 72) = 4.39, p < .05$).

For SR and SOR constructions, WM had significant main effects on whole-sentence RTs in all three plausibility conditions ($F(1, 36) = 21.54, p < .01$). HSs with larger WM capacity tended to read sentences faster than those with smaller WM capacity.

The results for the global RT measures indicate that initial implausibility in the SOR-biased plausibility condition had negative effects on overall RTs as well as QTs. The longest QTs were found for the SOR-SOR condition, whereas the longest whole-sentence RTs were observed in the SR-SOR condition. Somewhat mixed results were found for the preference for the SR structure over the SOR structure. QA rates were generally higher for the SR structure, but neither global measure showed significantly shorter RTs for the SR structure. This indicates that HSs did not always revise their incorrect initial interpretation of the sentence. These results for the global measures again provide general support for the GE processing and shallow parsing proposals.
Table 16

*Whole-Sentence Reading Times*

<table>
<thead>
<tr>
<th>Sentence Structure</th>
<th>SR-Biased</th>
<th>Neutral</th>
<th>SOR-Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR Construction</td>
<td>4937 (2079)</td>
<td>4942 (2095)</td>
<td>5552 (2243)</td>
</tr>
</tbody>
</table>

* Mean reading times (SD) in msec.

**Word-by-Word Reading Times**

Residual RTs were calculated by using trimmed RT data (200~3000 msec) in order to correct for the effect of the word length. Table 17 shows the mean uncorrected RTs for each word, and Table 18 shows the mean length-corrected residual RTs.

Table 17

*Mean Word-by-Word Reading Times*

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>1 Subject</th>
<th>2 1st Object</th>
<th>3 RCV</th>
<th>4 Dative NP</th>
<th>5 2nd Object/MV</th>
<th>6 MV</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR-SR</td>
<td>697.65</td>
<td>852.92</td>
<td>853.44</td>
<td>837.46</td>
<td>828.56</td>
<td>810.64</td>
</tr>
<tr>
<td>SR-N</td>
<td>703.15</td>
<td>914.09</td>
<td>858.32</td>
<td>832.51</td>
<td>789.90</td>
<td>783.79</td>
</tr>
<tr>
<td>SR-SOR</td>
<td>723.74</td>
<td>900.69</td>
<td>958.78</td>
<td>1028.75</td>
<td>906.16</td>
<td>945.94</td>
</tr>
<tr>
<td>SOR-SR</td>
<td>674.62</td>
<td>797.53</td>
<td>807.27</td>
<td>811.54</td>
<td>799.67</td>
<td>774.87</td>
</tr>
<tr>
<td>SOR-N</td>
<td>699.91</td>
<td>830.66</td>
<td>769.52</td>
<td>844.76</td>
<td>813.51</td>
<td>762.17</td>
</tr>
<tr>
<td>SOR-SOR</td>
<td>684.20</td>
<td>845.71</td>
<td>836.65</td>
<td>932.77</td>
<td>853.57</td>
<td>769.59</td>
</tr>
</tbody>
</table>

* Mean reading times in msec.

Table 18

*Mean Length-Corrected Residual Word Reading Times*

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>1 Subject</th>
<th>2 1st Object</th>
<th>3 RCV</th>
<th>4 Dative NP</th>
<th>5 2nd Object/MV</th>
<th>6 MV</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR-SR</td>
<td>-119.04</td>
<td>69.60</td>
<td>-16.43</td>
<td>-62.95</td>
<td>-22.05</td>
<td>-65.82</td>
</tr>
<tr>
<td>SR-N</td>
<td>-131.04</td>
<td>105.16</td>
<td>24.19</td>
<td>-49.90</td>
<td>-45.14</td>
<td>-76.91</td>
</tr>
<tr>
<td>SR-SOR</td>
<td>-94.99</td>
<td>86.91</td>
<td>103.37</td>
<td>147.62</td>
<td>63.81</td>
<td>77.63</td>
</tr>
<tr>
<td>SOR-SR</td>
<td>-80.10</td>
<td>55.16</td>
<td>-4.72</td>
<td>-35.25</td>
<td>-35.29</td>
<td>-24.93</td>
</tr>
<tr>
<td>SOR-N</td>
<td>-75.58</td>
<td>69.85</td>
<td>-10.13</td>
<td>6.81</td>
<td>-16.23</td>
<td>-9.89</td>
</tr>
<tr>
<td>SOR-SOR</td>
<td>-70.02</td>
<td>91.69</td>
<td>32.85</td>
<td>104.81</td>
<td>8.72</td>
<td>-18.26</td>
</tr>
</tbody>
</table>

* Mean reading times in msec.
Words 3 and 4: The Detection of Implausibility

Figures 8 and 9 show the mean length-corrected residual RTs at words 3 and 4, respectively. As shown in the figures, RT patterns across the conditions were similar for the two words, although word 4 showed a larger difference. The participants were more likely to slow down at word 3 in the SOR-biased plausibility condition than in the other two plausibility conditions ($F(2, 76) = 6.96, p < .01$), and the same was true at word 4 ($F(2, 76) = 28.44, p < .01$). Thus, the participants slowed down when they detected the implausibility between the object-marked noun at word 2 and the verb at word 3 in the SOR-biased plausibility condition.
Words 5 and 6: Disambiguation and Revisions

Figures 10 and 11 show the mean length-corrected residual RTs at words 5 and 6, respectively. As shown in the figures, the RT pattern across the conditions was generally similar for the two words. Both subject-based and item-based statistical analyses were conducted for each word separately and also for a region measure with both words.

There were no significant differences between SR and SOR constructions at words 5 and 6 and for combined measures ($F(1, 36) < 1s$). However, the plausibility bias had a significant main effect at both words (word 5: $F(2, 74) = 5.48, p < .01$; word 6: $F(2, 72) = 5.88, p < .01$). RTs were generally longer in the SOR-biased plausibility condition than in the other two
plausibility conditions. Although the nature of the interaction between the syntactic structure and the plausibility bias was similar at words 5 and 6, this interaction was significant only at word 6 and on the combined RT for both words ($F (2, 72) = 7.42, p < .01$, word 6; $F (2, 72) = 6.01, p < .01$, combined RT for both words). At words 5 and 6, RTs were significantly longer in the SOR-biased plausibility condition for the SR construction (SR-SOR) than in the other two conditions (SR-SR and SR-N), whereas RTs were generally similar across the three plausibility conditions for the SOR construction.

RTs were longer by 60 msec at word 5 and 100 msec at word 6 for the SR-SOR condition than for the SOR-SOR condition, but neither difference was significant ($p > .05$). According to the results of item-based and combined RT analyses, RTs at word 6 in the SR-SOR condition were longer by 95 msec ($p < .05$) and 160 msec ($p = .06$) than those in the SOR-SOR condition.

In general, HSs did not engage in structural reanalysis at words 5 and 6 in the SOR construction. Unlike NSs, HSs took less time reading word 6 for more plausible SOR-SOR sentences than for less plausible SR-SOR sentences. This indicates that the earlier implausibility in the SOR-biased plausibility condition might have led HSs away from the SR analysis during their initial structure building, which in turn might have reduced the processing cost of restructuring at word 6 in the SOR-SOR condition. These results provide support for parallel multiple constraint-based models, while the results for QTs and QAs offer evidence of a strong preference for the SR analysis, as suggested in the GE processing and shallow parsing proposals.

In terms of the WM effect, there was a significant two-way interaction between the plausibility bias and WM capacity for only word 6 ($F (2, 72) = 3.91, p < .05$). High-span HSs, who read whole sentences faster than low-span ones, took longer to read at word 6 in the Neutral
and SR-biased plausibility conditions than low-span ones (Neutral condition: $\beta = 9.57, t (36) = 1.95, p = .06$; SR-biased condition: $\beta = 9.83, t (36) = 1.98, p = .05$).

5.3. Korean Learners

*Accuracy of Paraphrase Judgments and Response Times*

Table 19 shows the accuracy of paraphrase judgments. Korean learners (K2s) were generally more accurate for paraphrases with the SR construction than with the SOR construction ($\chi^2 = 7.23, df = 1, p < .01$), which provides support for the argument that there is an overall preference for the SR structure. There was a significant interaction between plausibility and the syntactic structure ($\chi^2 = 24.53, df = 2, p < .01$). In the Neutral and SR-biased plausibility conditions, K2s showed higher QA rates for the SR construction than for the SOR construction ($p < .01$), whereas they showed similar response patterns for SR and SOR constructions in the SOR-biased plausibility condition ($p = .17$). For the SOR construction, the lowest QA rates were found when the SR-biased plausibility condition was followed by the disambiguation to the SOR structure (SOR-SR) ($p < .01$). K2s showed a clear indication of structural preference for the SR construction in their final interpretations when there was no plausibility-biased consideration.

Table 19

*Accuracy of Paraphrase Judgments (%)*

<table>
<thead>
<tr>
<th>Sentence Structure</th>
<th>SR-Biased</th>
<th>Neutral</th>
<th>SOR-Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR Construction</td>
<td>72.06</td>
<td>84.73</td>
<td>78.37</td>
</tr>
<tr>
<td>SOR Construction</td>
<td>17.46</td>
<td>64.44</td>
<td>66.35</td>
</tr>
</tbody>
</table>

Table 20 shows QTs for the paraphrase judgment task. There was no significant difference in QTs between SR and SOR constructions ($F (1, 40) < 1$), although QTs were shorter by 390 msec for the SR structure. There was no significant interaction between plausibility and
the syntactic structure \((F(2, 80) < 1)\). QTs were longer in the SOR-biased plausibility condition than in the Neutral condition \((F(2, 80) = 3.99, p < .05)\), indicating that the initial implausibility had negative effects on QTs in the paraphrase judgment task regardless of the structural difference between SR and SOR constructions. The longest QTs were observed for the SOR-SOR condition, whereas the lowest accuracy rates, for the SOR-SR condition.

Table 2

*Response Times for Paraphrase Judgments*

<table>
<thead>
<tr>
<th>Sentence Structure</th>
<th>SR-Biased</th>
<th>Plausibility Bias</th>
<th>Neutral</th>
<th>SOR-Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR Construction</td>
<td>7925 (2246)</td>
<td>7558 (2743)</td>
<td>8548 (2504)</td>
<td></td>
</tr>
<tr>
<td>SOR Construction</td>
<td>7653 (1588)</td>
<td>8193 (1148)</td>
<td>9357 (1979)</td>
<td></td>
</tr>
</tbody>
</table>

* Mean reading times \((SD)\) in msec.

**Whole-Sentence Reading Times**

Table 21 shows the mean whole-sentence RTs. There was no significant difference in the whole-sentence RTs between SR and SOR constructions \((F(1, 40) = 1.21, p < .23)\), although whole-sentence RTs were shorter by 759 msec for the SR structure. There was some interaction between plausibility and syntactic structure \((F(2, 80) = 2.76, p = .07)\). Only in the Neutral condition, whole-sentence RTs were significantly shorter for the SR-N condition than for the SOR-N condition. Total RTs were longer in the SOR-biased plausibility condition than in the SR-biased and Neutral plausibility conditions \((F(2, 80) = 11.85, p < .01)\), indicating that the initial implausibility had negative effects on total RTs.

Both QTs and whole-sentence RTs were longer for the SR-SOR condition than for the SR-SR and SR-N conditions. On the other hand, QTs were longer for the SOR-SOR condition than for the SOR-N condition, whereas whole-sentence RTs were longer for the SOR-SOR condition than for the SOR-SR condition. The sentence structure had no main effect on QTs or total RTs \((Fs < 1)\), although both were shorter (but not significantly) for SR sentences (390 msec
for QTs, 759 msec for total RTs). WM had no main effect on both QTs or total RTs ($F$s < 1).

The analysis of whether the SOR structure was generally more difficult provided somewhat mixed results. Paraphrase judgment accuracy was generally higher for the SR structure, but neither of the global time measures showed reliably faster times for the SR structure. On the other hand, the results for the global measures indicate that the initial implausibility had negative effects on QTs as well as whole-sentence RTs. This indicates that K2s did not always revise their incorrect initial interpretation of the sentence. These results for the global measures are quite consistent with the GE processing and shallow parsing proposals.

Table 21

*Whole-Sentence Reading Times*

<table>
<thead>
<tr>
<th>Sentence Structure</th>
<th>SR-Biased</th>
<th>Neutral</th>
<th>SOR-Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR Construction</td>
<td>6464 (2066)</td>
<td>5997 (1300)</td>
<td>7070 (2063)</td>
</tr>
<tr>
<td>SOR Construction</td>
<td>7039 (1583)</td>
<td>7175 (1899)</td>
<td>7596 (2300)</td>
</tr>
</tbody>
</table>

* Mean reading times (SD) in msec.

*Word-by-Word Reading Times*

Table 22 shows the mean uncorrected RTs for each word, and Table 23 shows the mean length-corrected residual times.

Table 22

*Mean Word-by-Word Reading Times*

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>1 Subject</th>
<th>2 1st Object</th>
<th>3 RCV</th>
<th>4 Dative NP</th>
<th>5 2nd Object/ MV</th>
<th>6 MV</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR-SR</td>
<td>854.79</td>
<td>1159.14</td>
<td>1019.99</td>
<td>1122.14</td>
<td>1140.58</td>
<td>1033.70</td>
</tr>
<tr>
<td>SR-N</td>
<td>849.39</td>
<td>1229.36</td>
<td>1068.59</td>
<td>998.15</td>
<td>1045.40</td>
<td>997.55</td>
</tr>
<tr>
<td>SR-SOR</td>
<td>873.34</td>
<td>1158.22</td>
<td>1194.35</td>
<td>1189.44</td>
<td>1130.61</td>
<td>1137.87</td>
</tr>
<tr>
<td>SOR-SR</td>
<td>763.60</td>
<td>1144.16</td>
<td>1018.00</td>
<td>1147.84</td>
<td>1067.71</td>
<td>1356.31</td>
</tr>
<tr>
<td>SOR-N</td>
<td>731.63</td>
<td>1259.87</td>
<td>1040.04</td>
<td>1110.19</td>
<td>993.46</td>
<td>1228.42</td>
</tr>
<tr>
<td>SOR-SOR</td>
<td>765.44</td>
<td>1175.04</td>
<td>1113.30</td>
<td>1259.91</td>
<td>1062.27</td>
<td>1243.55</td>
</tr>
</tbody>
</table>

* Mean reading times in msec.
Table 23

Mean Length-Corrected Residual Word Reading Times*

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Subject</th>
<th>1st Object</th>
<th>RCV</th>
<th>Dative NP</th>
<th>2nd Object/MV</th>
<th>MV</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR-SR</td>
<td>-37.49</td>
<td>102.15</td>
<td>-5.35</td>
<td>89.06</td>
<td>52.11</td>
<td>13.70</td>
</tr>
<tr>
<td>SR-SOR</td>
<td>168.67</td>
<td>269.64</td>
<td>152.95</td>
<td>323.44</td>
<td>91.64</td>
<td>-4.56</td>
</tr>
<tr>
<td>SOR-SR</td>
<td>-37.10</td>
<td>142.36</td>
<td>7.64</td>
<td>90.19</td>
<td>106.93</td>
<td>34.17</td>
</tr>
<tr>
<td>SOR-N</td>
<td>353.22</td>
<td>250.86</td>
<td>47.76</td>
<td>-99.07</td>
<td>35.37</td>
<td>26.82</td>
</tr>
<tr>
<td>SOR-SOR</td>
<td>191.00</td>
<td>200.93</td>
<td>221.04</td>
<td>117.94</td>
<td>84.84</td>
<td>74.30</td>
</tr>
</tbody>
</table>

* Mean reading times in msec.

Figure 12. Mean length-corrected residual word RTs at six word positions for K2s.

Words 3 and 4: The Detection of Implausibility

Figures 13 and 14 show the mean length-corrected residual RTs at words 3 and 4, respectively. As shown in the figures, RT patterns across the conditions were similar for the two words, although word 4 showed a larger difference. The participants were more likely to slow down at word 3 in the SOR-biased plausibility condition than in the other two plausibility conditions \( F (2, 80) = 3.63, p < .05 \), and the same was true at word 4 \( F (2, 80) = 8.13, p < .01 \).
Thus, the participants slowed down when they detected the implausibility between the object-marked noun at word 2 and the verb at word 3 in the SOR-biased plausibility condition.

WM had some main effect on RTs at words 3 only \((F(1, 40) = 3.57, p = .06)\). K2s with more WM capacity took less time reading words at word 3 in the SOR-biased plausibility condition than those with less WM capacity \((\beta = -11.62, t(39) = -1.85, p = .06)\), which provides some evidence of WM’s sensitivity to plausibility information for adult L2 learners of Korean.

**Figure 13.** RTs at word 3.  
**Figure 14.** RTs at word 4.

**Words 5 and 6: Disambiguation and Revisions**

Figures 15 and 16 show the mean length-corrected residual RTs at words 5 and 6, respectively. Both subject-based and item-based statistical analyses were conducted on each word separately and also on a region measure with both words.
There were significant differences between SR and SOR constructions at word 6 only ($F(1, 37) = 3.85, p = .05$), although RTs were shorter by 32 msec at word 5 and 97 msec at word 6 for the SR construction. This indicates that K2s did not start restructuring at word 5, where they were supposed to detect the disambiguation to the SOR structure. At word 6, they showed some sign of structural reanalysis in the SOR construction.

The plausibility bias had a significant main effect at word 6 ($F(2, 80) = 3.32, p < .05$) and for the measure combining words 5 and 6 ($F(2, 80) = 5.56, p < .01$). The participants were more likely to read faster at word 6 in the Neutral plausibility condition than in the other two plausibility conditions ($p < .05$). However, there was no significant interaction between the syntactic structure and plausibility at either word position ($Fs < 1$). No WM effect was found for RTs at words 5 or 6 ($Fs < 1$).

Findings from the global response patterns and RT data at words 5 and 6 indicate that K2s generally had more difficulty with the SOR structure than with the SR structure and offer evidence of a strong preference for the SR analysis as suggested in the GE processing and
shallow parsing proposals. In other words, K2s had difficulty performing structural reanalysis at the word-final position in the SOR construction regardless of plausibility bias information.

5.4. Native Speakers vs. Heritage Speakers

The two groups showed different levels of Korean proficiency \( F(1, 74) = 51.18, p < .01 \), and there was no within-group difference between the two syntactic lists (SR vs. SOR) \( F(1, 96) < 1 \). In the cloze-type proficiency test, NSs showed higher accuracy rates (95.23%) than HSs (91.07%). The two groups had different level of WM capacity \( F(1, 74) = 4.29, p < .05 \), with little difference between the two syntactic lists \( F(1, 74) < 1 \)). In general, HSs took longer reading sentences than NSs \( p < .05 \). This is likely due in part to the difference in the level of Korean proficiency between the two groups.

**Accuracy of Paraphrase Judgments and Response Times**

Figure 17 shows the accuracy of paraphrase judgments by group. NSs were generally more accurate for paraphrases than HSs \( \chi^2 = 4.34, df = 1, p < .05 \). Both groups were more accurate for paraphrases in the SR construction than those in the SOR construction \( \chi^2 = 514.51, df = 1, p < .01 \), which provides support for the argument that there is an overall preference for the SR structure. There was a significant interaction between plausibility and the syntactic structure \( \chi^2 = 175.00, df = 2, p < .01 \).

Both groups were more likely to provide a “yes” answer for the SR-SR (SR-N) construction than for the SOR-SR (SOR-N) construction \( p < .01 \), and both were likely to answer “yes” in a similar way between the SOR-biased plausibility conditions (SR-SOR and SOR-SOR). Both groups showed a strong structural preference for a simpler structure in the Neutral and SR-biased plausibility conditions. On the other hand, they did not show the same type of preference in the SOR-biased plausibility condition, in which the plausibility bias
supported the SOR syntactic analysis.

In terms of plausibility effects for the SR construction, both groups were less likely to provide a “yes” answer for the SR-SOR condition than for the SR-N condition ($p < .05$). For the SOR construction, however, they showed the lowest accuracy rates for the SOR-SR condition, and lower accuracy rates for the SOR-N condition than for the SOR-SOR condition ($p < .05$). Both showed lower accuracy rates in the plausibility and syntactic information mis-match conditions for both constructions. SOR-biased plausibility information improved the accuracy of the SOR analysis for the SOR-SOR condition, that is, both groups showed the highest accuracy rates for the SOR-SOR condition among three plausibility conditions for the SOR construction. In general, NSs and HSs showed similar paraphrase judgment response (QA) patterns despite the overall group difference in accuracy rates between the two.

There was a significant two-way interaction between plausibility and WM capacity only for HSs ($\chi^2 = 15.91, df = 2, p < .01$). High-span HSs were less likely to provide a “yes” answer in the SOR-biased plausibility condition than low-span HSs ($\beta = -.20, t (37) = -3.23, p < .01$). In the SOR-biased plausibility conditions, high-span HSs were less accurate regardless of the syntactic structural difference between SR and SOR constructions. This indicates their heavy reliance on plausibility information, which led them away from a syntactically correct analysis in either case.
Figure 17. Accuracy of paraphrase judgments for NSs and HSs (%).

Figure 18 shows QTs by group. In general, NSs were slightly faster in reading and answering comprehension questions than HSs ($F(1, 74) = 3.15, p = .08$). There was a significant three-way interaction among plausibility, syntactic structure, and group ($F(2, 148) = 6.25, p < .01$). In addition, there was a significant two-way interaction between plausibility and syntactic structure ($F(2, 148) = 10.37, p < .01$), and syntactic structure had a main effect on QTs ($F(1, 74) = 4.63, p < .05$). The two groups showed different QT patterns for the two constructions across the three plausibility conditions.

There was little difference in the amount of time that HSs spent answering comprehension questions between SR and SOR constructions in all three plausibility conditions ($F(1, 37) < 1$), whereas QTs for NSs were longer for the SOR construction in the Neutral and SR-biased plausibility conditions ($F(2, 74) = 15.69, p < .01$). HSs showed longer QTs in the SOR-biased plausibility condition than in the Neutral and SR-biased plausibility conditions for both constructions ($F(2, 76) = 21.18, p < .05$), whereas NSs showed longer QTs in the SOR-biased condition than in the other two conditions only for the SR construction ($F(2, 74) = 12.88, p < .01$). When responding to paraphrased comprehension questions, NSs showed some structural preferences for the SR construction when the plausibility bias supported the SR
analysis in the Neutral and SR-biased plausibility conditions. On the other hand, HSs were not sensitive to structural differences.

In general, these two groups showed different QT patterns. HSs showed stronger plausibility effects on QTs regardless of the structural difference, whereas NSs showed plausibility effects only in the SR construction, which was easier to process than the SOR construction. Only HSs showed a significant main effect of WM on QTs ($F(1, 37) = 12.77, p < .01$). HSs with less WM capacity took longer to read and answer comprehension questions than those with more WM capacity, regardless of the structural differences.

![Figure 18](image)

**Figure 18.** Response times for paraphrase judgments for NSs and HSs.

**Whole-Sentence Reading Times**

Figure 19 shows the mean whole-sentence RTs by group. In general, NSs read sentences faster than HSs ($F(1, 73) = 17.04, p < .01$), which is likely due to the difference in their level of Korean proficiency. There was a significant three-way interaction among plausibility and the syntactic structure, and the group ($F(2, 146) = 4.74, p < .05$). In addition, there was a significant two-way interaction between plausibility and syntactic structure ($F(2, 146) = 3.16, p = .05$), and the plausibility bias had a significant main effect on total RTs ($F(2, 146) = 8.92, p < .01$). Syntactic structure had no main effect on total RTs ($F(1, 73) < 1$).
The two groups showed similar whole-sentence RT patterns. Each group showed similar whole-sentence RTs for SR and SOR constructions across the three plausibility conditions. Both groups showed longer whole-sentence RTs in the SOR-biased plausibility condition than in the other two conditions (NSs: $F (2, 72) = 5.34, p < .01$; HSs: $F (2, 72) = 4.39, p < .05$). Both groups showed longer whole-sentence RTs for the SOR-SOR condition than for the SOR-SR condition, which indicates that the initial implausibility in the SOR-SOR condition had more negative effects on whole-sentence RTs than the implausibility of the whole-sentence in the SOR-SR resulting from a mis-matched SOR analysis.

For both SR and SOR constructions, only HSs showed significant main effects of WM on whole-sentence RTs across the three plausibility conditions ($F (1, 36) = 21.54, p < .01$). HSs with more WM capacity tended to read faster than those with less WM capacity.

![Figure 19. Whole-sentence RTs for SR and SOR constructions for NSs and HSs.](image)

*Word-by-Word Reading Times*

For a comparison of RT patterns between NSs and HSs, Figure 20 shows the mean length-corrected residual times by group for six world positions.
Figure 20. RTs for SR and SOR constructions by position for NSs and HSs.

**Words 3 and 4: The Detection of Implausibility**

Figures 21 and 22 show the mean length-corrected residual RTs by group at words 3 and 4, respectively. There was little group difference in RTs at word 3 ($F(1, 75) = 2.30, p = .13$) or word 4 ($F(1, 75) = 1.50, p = .23$). The two groups showed similar RT patterns across the three conditions for these words. Both groups were slower at word 3 in the SOR-biased plausibility condition than in the other two conditions ($F(2, 152) = 13.84, p < .01$), and the same was true at word 4 ($F(2, 152) = 61.29, p < .01$). Thus, both groups slowed down when they detected the implausibility between the object-marked noun at word 2 and the verb at word 3 in the SOR-biased plausibility condition.

WM had a main effect on RTs at word 3 only for NSs ($F(1, 37) = 6.73, p < .05$). NSs with less WM capacity took longer to read words at word 3 in the SOR-biased condition than
those with more WM capacity, providing some evidence of WM’s sensitivity to the implausibility for NSs.

*Figure 21.* RTs at word 3 for NSs and HSs.  
*Figure 22.* RTs at word 4 for NSs and HSs.  

**Words 5 and 6: Disambiguation and Revisions**

Figures 23 and 24 show the mean length-corrected residual RTs at words 5 and 6, respectively. There was no significant group difference in RTs at the two words ($F (1, 73) < 1s$). As shown in the figures, the two groups showed generally similar RT patterns across the three conditions at words 5 and 6, although there was some significant group difference in RTs for the SOR construction.

There was no significant group differences between SR and SOR constructions at words 5 and 6 ($F (1, 73) < 1s$); however, the plausibility bias had a significant main effect on RTs for both groups at word 5 ($F (2, 148) = 14.09, p < .01$) and word 6 ($F (2, 146) = 5.23, p < .01$). There was a significant three-way interaction of plausibility and syntactic structure and group on RTs only at word 6 ($F (2, 146) = 4.70, p < .01$). Similarly, there was a significant two-way interaction effect of plausibility and syntactic structure on RTs only at word 6 ($F (2, 146) = 4.47, p < .05$).

For both groups, the plausibility bias had a main effect on RTs at word 5 (NSs: $F (2, 74) = 13.87, p < .01$; HSs: $F (2, 74) = 5.48, p < .01$). NSs showed significantly shorter RTs in the Neutral condition than in the other two conditions for both SR and SOR constructions. On the
other hand, HSs showed significantly longer RTs for the SR-SOR condition than for SR-SR and SR-N conditions, whereas they showed similar RTs across the three plausibility conditions in the SOR construction. NSs showed shorter RTs at word 5 for the SOR-N condition with a plausible reanalysis than for the SOR-SR condition (in which the initial SR analysis had to be restructured into an SOR analysis with an implausible reanalysis) and for the SOR-SOR condition (in which the initial SOR implausibility did not reduce their restructuring cost from the SR analysis to the SOR analysis). In contrast, HSs showed no sign of structural reanalysis at word 5 for the SOR construction.

The two groups showed slightly different RT patterns across the three plausibility conditions for SR and SOR constructions at word 6. The plausibility bias had a main effect on RTs at word 6 for NSs ($F(2, 74) = 4.76, p < .01$), although none of the individual multiple comparisons within each construction was significant at the $p = .05$ level. On the other hand, there was a significant interaction between plausibility and syntactic structure for HSs ($F(2, 72) = 7.42, p < .01$). HSs showed significantly longer RTs for the SR-SOR condition than for SR-SR and SR-N conditions, whereas there was little difference in RTs for the SOR construction across the three plausibility conditions ($p > .05$). HSs showed no sign of structural reanalysis at word 6, either.

For the SR construction, both groups showed longer RTs at words 5 and 6 for the SR-SOR condition than for SR-SR and SR-N conditions, which indicates some processing cost of overall implausible sentences in the condition with conflicting syntactic and plausibility information. For the SOR construction, however, both groups showed no sign of structural reanalysis at word 6, whereas HSs showed no such sign at word 5. In contrast, NSs showed longer RTs at word 5 for SOR-SR and SOR-SOR constructions than for the SOR-N construction,
which allowed for a compatible analysis of the object noun at word 2 as a direct object of the embedded RCV or the MV. There was no structural preference for the SR construction in the SOR-biased plausibility condition for both groups, although RTs tended to be longer for the SR-SOR condition than for the SOR-SOR condition by 60 msec at word 5 and 96 msec at word 6. HSs were more likely to be adversely influenced by the initial implausibility than NSs.

Figure 23. RTs at word 5 for SR and SOR constructions for NSs and HSs.

Figure 24. RTs at word 6 for SR and SOR constructions for NSs and HSs.

For HSs, there was a significant two-way interaction between plausibility and WM capacity only at word 6 ($F(2, 72) = 3.91, p < .05$), whereas for NSs, WM has no main effect on words 5 and 6. HSs with a high level of WM capacity took longer to read at word 6 in the SR-biased plausibility condition ($\beta = 9.57, t(36) = 1.95, p = .06$) and in the Neutral condition ($\beta = 9.83, t(36) = 1.98, p = .05$). Low-span HSs took less time to read words at the end of the sentence with initially plausible SR bias information than high-span HSs.
5.5. Native Speakers vs. Korean Learners

The two groups showed different levels of Korean proficiency ($F (2, 77) = 79.22, p < .01$), and there was no within-group difference between the two syntactic lists (SR vs. SOR) ($F (1, 77) < 1$). In the cloze-type proficiency test, NSs showed higher accuracy rates (95.23%) than K2s (71.45%). The two groups had different levels of WM capacity ($F (2, 77) = 7.40, p < .01$) with little difference between the two syntactic lists ($F (1, 77) < 1$). In general, K2s took longer reading sentences than NSs ($p < .05$), which is likely due to differences in their Korean proficiency ($p < .05$). L1 information of K2s was included as a covariate to analyze RT data in addition to the WM RST scores. However, it was excluded in the analysis report because it was not significant in any of the measures ($F (1, 75) < 1s$).

Accuracy of Paraphrase Judgments and Response Times

Figure 25 shows the accuracy of paraphrase judgments by group. NSs and K2s had overall similar accuracy rates for paraphrases ($\chi^2 = 2.05, df = 1, p = .15$). There were significant two-way interactions between plausibility and syntactic structure ($\chi^2 = 99.49, df = 2, p < .01$) and between group and syntactic structure ($\chi^2 = 4.99, df = 2, p < .05$).

In general, NSs and K2s were more accurate for paraphrases in the SR construction than in the SOR construction ($\chi^2 = 17.27, df = 1, p < .01$), providing support for the argument that there is an overall preference for the SR structure. Both groups were more likely to provide a “yes” answer for the SR-SR construction than for the SOR-SR construction ($p < .01$), and both were more likely to provide a “yes” answer for the SR-N condition than for the SOR-N condition ($p < .05$). However, both were likely to provide a “yes” answer for both SR-SOR and SOR-SOR conditions. Both groups showed a strong structural preference for a simpler structure in the Neutral and SR-biased plausibility conditions, whereas they did not show the same type of
preference in the SOR-biased plausibility condition, in which plausibility supported the SOR syntactic analysis.

In terms of plausibility effects for the SR construction, both groups were less likely to provide a “yes” answer for the SR-SOR condition than for the SR-N condition ($p < .05$). For the SOR construction, on the other hand, both groups showed the lowest accuracy rates for the SOR-SR condition ($p < .05$). In other words, both groups had the lowest accuracy in the plausibility and syntactic information mis-match conditions for both constructions. In general, the two groups showed similar paraphrase judgment response accuracy rates in terms of the plausibility effect and WM had no main effect on the QA rates for neither group ($\chi^2 = 1.46, df = 1, p = .23$).

Figure 25. Accuracy of paraphrase judgments for NSs and K2s (%).

Figure 26 shows QTs by group. In general, NSs took much less time for the paraphrase judgment task than K2s ($F (1, 56) = 61.37$, $p < .01$). There was a significant three-way interaction among plausibility, syntactic structure, and group ($F (2, 112) = 2.29$, $p < .01$). In addition, there was a significant two-way interaction between plausibility and syntactic structure ($F (2, 112) = 4.05$, $p < .05$), and the plausibility bias had a significant main effect on QTs ($F (2, 112) = 13.56$, $p < .01$). The two groups showed different QT patterns for SR and SOR constructions across the three plausibility conditions.
There were no significant differences in QTs between SR and SOR constructions for NSs ($F(1, 36) = 3.18, p = .08$) and K2s ($F(1, 19) < 1$), although NSs and K2s were faster for the SR structure by 760 msec and 390 msec, respectively.

In terms of the effect of the plausibility bias on QTs, there was a significant interaction between plausibility and syntactic structure for NSs. NSs showed longer QTs in the SOR-biased plausibility condition than in the other two conditions only for the SR construction ($F(2, 74) = 12.88, p < .01$). However, they showed similar QTs across the three plausibility conditions for the SOR construction. On the other hand, K2s showed longer QTs in the SOR-biased plausibility condition than in the Neutral and SR-biased conditions regardless of the structural difference ($F(2, 40) = 3.99, p < .05$). None of the individual multiple comparisons of QTs between the plausibility conditions were significant at $p = .05$ level for the L2 group. When reading and responding in the paraphrase judgment tasks, K2s showed less native-like response patterns and QTs in terms of plausibility processing in the SOR-biased plausibility condition than in the other two plausibility conditions. Neither group displayed a WM effect on the QTs ($F(1, 56) = 1.09, p = .30$).

![Figure 26](image_url)  
*Figure 26. Response times for paraphrase judgments for NSs and K2s.*

*Whole-Sentence Reading Times*
Figure 27 shows the mean whole-sentence RTs by group. In general, NSs read sentences much faster than K2s ($F(1, 75) = 26.77, p < .01$). There was no significant interaction between plausibility and the syntactic structure, and the syntactic structure had no main effect on total RTs ($F$s < 1). For both groups, there were no significant differences in whole-sentence RTs between SR and SOR constructions across the three plausibility conditions ($F$s < 1). Both groups showed longer whole-sentence RTs in the SOR-biased plausibility condition than in the other two plausibility conditions ($F(2, 72) = 5.34, p < .01$ for NSs, $F(2, 80) = 11.85, p < .05$ for K2s).

In terms of the global RT measures, neither group was not sensitive to structural differences between SR and SOR constructions, whereas they were sensitive to the implausibility in the SOR-biased plausibility condition. Neither group had any WM effects on whole-sentence RTs ($F$s < 1).

![Figure 27. Whole-sentence reading times for NSs and K2s.](image)

**Word-by-Word Reading Times**

For a comparison of RT patterns between the two groups, Figure 28 shows the mean length-corrected residual RTs at the six word positions.
Words 3 and 4: The Detection of Implausibility

Figures 29 and 30 show the mean length-corrected residual RTs by group at words 3 and 4, respectively. There was some group difference in RTs at word 3 ($F(1, 78) = 3.79, p = .05$), but not at word 4 ($F(1, 78) = 2.23, p = .14$). NSs were significantly faster at word 3 than K2s. Both groups were significantly slower at word 3 in the SOR-biased plausibility condition than in the other two conditions ($F(2, 156) = 6.21, p < .01$), and the same was true at word 4 ($F(2, 156) = 14.17, p < .01$). Thus, both groups slowed down when they detected implausibility between the object-marked noun at word 2 and the verb at word 3 in the SOR-biased plausibility condition.

WM had a significant main effect on RTs at word 3 for both groups (NSs: $F(1, 37) = 6.73, p < .05$; K2s: $F(1, 40) = 3.57, p = .06$). High-span participants in both groups took less
time at word 3 in the SOR-biased plausibility condition than low-span ones, which provides some evidence of WM’s sensitivity to plausibility for both NSs and K2s.

**Figure 29.** RTs at word 3 for NSs and K2s. **Figure 30.** RTs at word 4 for NSs and K2s.

### Words 5 and 6: Disambiguation and Revisions

Figures 31 and 32 show the mean length-corrected residual RTs by group at words 5 and 6, respectively. There was a significant group difference in RTs at word 5 ($F(1, 76) = 5.36, p < .05$), but no group difference in RTs at word 6 ($F(1, 76) < 1$), which may be due to significantly shorter RTs for K2s for the SR-N condition. NSs were significantly faster at word 5 than K2s.

There were significant differences in RTs between SR and SOR constructions at word 6 only in the case of K2s ($F(1, 76) = 4.31, p < .05$). NSs showed similar RTs at both words for the two constructions across the three plausibility conditions ($F(1, 36) < 1$s). On the other hand, K2s showed shorter RTs for the SR-N construction than for the SOR-N construction at word 6 ($p < .05$). Unlike NSs, who did not show a significant structural preference for the SR construction at words 5 and 6, K2s showed structural preferences for the SR construction at word 6. The results for K2s’ structural preference patterns indicate that they showed no sign of structural reanalysis until word 6, where they showed some sign of processing difficulty in the Neutral
condition in particular. K2s were sensitive to structural differences, but only when they arrived at the word immediately following the disambiguation point.

The plausibility bias had significant main effects on RTs at word 5 ($F(2, 154) = 8.20, p < .01$) and at word 6 ($F(2, 154) = 5.61, p < .01$). In terms of word 5, the plausibility bias had significant main effects on RTs for NSs ($F(2, 74) = 13.86, p < .01$), but not for K2s ($F(2, 80) = 2.54, p = .085$). Only NSs showed significantly shorter RTs in the Neutral condition than in the other two plausibility conditions regardless of the structural difference. On the other hand, the two groups showed similar RT patterns at word 6 across the three plausibility conditions for SR and SOR constructions. The plausibility bias had a significant main effect on RTs at word 6 for NSs ($F(2, 74) = 4.76, p < .05$) and for K2s ($F(2, 80) = 3.33, p < .05$). Both groups showed significantly shorter RTs in the Neutral condition than in the SOR-biased plausibility condition ($p < .05$).

In general, for both groups, WM had no significant effect on RTs at words 5 or 6 ($F(1, 76) < 1$).

Figure 31. RTs at word 5 for SR and SOR constructions for NSs and K2s.
5.6. General Discussion

5.6.1. L1 sentence processing in Korean

The effects of the plausibility bias and the structural preference on NSs’ processing of two types of syntactically ambiguous RC sentences in Korean was examined to determine which sentence processing models (the serial syntax-first models or the parallel constraint-based models) would better explain sentence processing in Korean. For this, two types of RC constructions (SR and SOR) in Korean were employed with three plausibility bias conditions (SR-biased, SOR-biased, and Neutral conditions).

The first research question addressed whether there would be a structural preference for the SR construction over the SOR construction and how the plausibility bias and the structural preference would interact for two syntactic constructions under different plausibility bias conditions. The SR construction has generally been considered to be more preferable to the SOR construction (Kim, 2004), which is structurally challenging with the need for reanalysis and thus has a less frequent usage. However, the results for RTs indicate that NSs showed no overall structural preference for the SR construction. Only the accuracy of paraphrase judgments (QA) provided some evidence of an overall preference for the SR structure. The results for all other RT measures, including QTs, whole-sentence RTs, and RTs at critical positions, indicate no
significant difference between SR and SOR constructions.

NSs were slower (although not significantly so) at words 5 and 6 by 22 msec and 88 msec, respectively, for the SR construction than for the SOR construction. In this regard, more sensitive RT measures such as the eye-tracking method may shed more light on the difference in RTs between the two constructions.

In terms of RTs at words 5 and word 6 for the structural reanalysis of the SOR construction, NSs showed no significant differences in RT patterns at and after the disambiguating point at word 5 between the two constructions across the three plausibility conditions. Serial models predict longer RTs at word 5 for the SOR construction than for the SR construction regardless of the plausibility condition. On the other hand, parallel models predict longer RTs for the SOR construction in the Neutral and SR-biased plausibility conditions (SOR-SR and SOR-N) than for the SR construction (SR-SR and SR-N). In addition, they predict similar or even shorter RTs for the SOR construction in the SOR-biased plausibility condition (SOR-SOR) than for the SR construction (SR-SOR) depending on the level of structural difficulty and the plausibility manipulation. The earlier SOR-biased plausibility of the SOR-SOR condition is expected to reduce the structural reanalysis cost at word 5. Surprisingly, NSs showed no differences in RTs at word 5 between SR and SOR constructions across all three plausibility conditions. In other words, they showed no sign of a restructuring cost for RTs at word 5 across the plausibility conditions, which provides some support for parallel models.

On the other hand, the global measures for paraphrased comprehension questions (i.e., QTs and QA rates) showed significant interactions between the plausibility bias and the syntactic structure. The results for QTs and QA rates indicate that NSs showed a stronger structural preference for the SR construction (i.e., the SR syntactic analysis following the MA) in the
Neutral and SR-biased plausibility conditions. In the SOR-biased plausibility condition, in which plausibility information was in conflict with the SR-syntactic analysis, NSs showed no preference for the SR construction in terms of QTs and QAs. This indicates that NSs engaged in structural reanalysis while responding to comprehension questions only in the condition in which the plausibility bias supported the SR analysis. However, they were not likely to slow down at these two positions for the structural reanalysis.

Table 24 shows the results for critical positions (the actual results for conflicting conditions at word 5 are underlined).

Table 24

<table>
<thead>
<tr>
<th>Reading Times at Critical Positions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Critical Word</td>
</tr>
<tr>
<td>Position</td>
</tr>
<tr>
<td>Serial Models</td>
</tr>
<tr>
<td>Word 3</td>
</tr>
<tr>
<td>SR bias &lt; SOR bias</td>
</tr>
<tr>
<td>SR construction</td>
</tr>
<tr>
<td>SR bias &lt; SOR bias</td>
</tr>
<tr>
<td>SOR construction</td>
</tr>
<tr>
<td>SR bias &lt; SOR bias</td>
</tr>
<tr>
<td>SR-SR-N=SOR-SOR</td>
</tr>
<tr>
<td>SOR-N=SOR</td>
</tr>
<tr>
<td>SOR-SOR</td>
</tr>
<tr>
<td>SR-SOR &lt; SOR-SOR</td>
</tr>
<tr>
<td>SR-SR &lt; SOR-SR</td>
</tr>
<tr>
<td>SR-N &lt; SOR-N</td>
</tr>
<tr>
<td>SR-SR &gt; SOR-SOR</td>
</tr>
<tr>
<td>SR-SR &lt; SOR-SR</td>
</tr>
<tr>
<td>SR-N &lt; SOR-N</td>
</tr>
<tr>
<td>SR-SOR &gt; SOR-SOR</td>
</tr>
<tr>
<td>SR-SR &lt; SOR-SOR</td>
</tr>
<tr>
<td>SR-N &lt; SOR-N</td>
</tr>
<tr>
<td>SR-SOR &gt; SOR-SOR</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

In terms of the use of plausibility information during the ambiguity resolution process, NSs immediately detected plausibility information in the SOR-biased plausibility condition. Two SOR-biased plausibility conditions, SR-SOR and SOR-SOR, showed an implausible relationship between the first object-marked noun at word 2 and the embedded RC verb at word 3. NSs showed significantly longer RTs at words 3 and 4 in the SOR-biased plausibility condition than the other two plausibility conditions.

For NSs, there were some plausibility-facilitating/deteriorating effects based on the
biasness of plausibility information in each syntactic construction. NSs showed significantly shorter RTs at word 5 in the Neutral condition than in the SR-biased and SOR-biased plausibility conditions for both syntactic constructions. Serial models predict similar RTs at word 5 across all three plausibility conditions for both SR and SOR constructions, whereas parallel models predict longer RTs at word 5 in the SOR-biased plausibility condition for the SR construction (SR-SOR) because the appearance of the second object noun at word 5 confirms the initially unfavorable SR analysis and the earlier SOR bias generates implausible sentences. These results indicate longer RT at word 5 for SR-SOR than for SR-N, which implies that the early detection of implausibility might have influenced the participants’ decision on the initial structure such that they incurred some processing cost at the disambiguating point at word 5 when they encountered an unexpected second-object marked noun.

For the SOR construction, however, serial models consider longer RTs at word 5 in all three plausibility conditions as a sign of structural reanalysis. In contrast, parallel models predict shorter RTs at word 5 in the Neutral and SOR-biased plausibility conditions than in the SR-biased plausibility condition. The difference between SOR-SR and SOR-N lies in the degree of the plausibility bias. The SOR-SR condition involves a strongly implausible relationship between the object noun at word 2 and the MV at word 5, whereas the SOR-N condition has an object noun at word 2 that is compatible with either the embedded RC verb at word 3 or the MV at word 5. The difference between SOR-SR and SOR-SOR lies in the initial implausibility. For the SOR-SR condition, readers generally start with a plausible SR analysis and then face an unexpected structural reanalysis at word 5. On the other hand, in the SOR-SOR condition, readers are led away from an SR analysis with the SOR-biased plausibility if the early implausibility is properly manipulated and end up with an SOR analysis.
However, in the present study, NSs showed similar RT patterns for SOR-SR and SOR-SOR conditions, and their RTs for these conditions were much longer than those for the SOR-N. This indicates that the initial implausibility of the SOR construction might not have induced NSs to parse the SOR analysis early or strongly enough, which led to some restructuring cost at word 5 for the SOR-SOR condition.

The results for the global measures indicate some plausibility effects, which implies that the plausibility influenced NSs’ final interpretations. In general, NSs preferred the SR analysis (the MA reading), and as a result, the SOR analysis produced longer RTs and lower accuracy rates in the paraphrase judgment task. For both constructions, NSs showed lower accuracy rates and longer RTs in the syntactic structure and plausibility conflicting conditions. NSs showed higher QA rates for SOR-SOR (which involved longer RTs and QTs) than for SOR-SR and SOR-N. Thus, the SOR-biased plausibility condition helped NSs to better comprehend sentences in the SOR-SOR condition.

In sum, for NSs, the plausibility bias influenced the process of resolving temporary syntactic ambiguities, which is consistent with the previous findings of other L1 studies, namely shorter RTs in the plausible condition but longer RTs in the implausible condition. NSs also made some use of plausibility information in their initial structure building, particularly for the SR construction. That is, the plausibility had some influence on NSs’ initial structural decisions as well as final interpretations in the process of resolving syntactic ambiguities. These results provide more support for parallel multiple constraint-based models than for syntax-first serial models.

In general, for NSs, WM had no influence on the most measures of sentence processing for either online processing or final interpretations in the present study. However, WM had a
main effect on RTs at words 3 and 4, indicating that NSs with a high level of WM capacity took less time to process the initial implausibility than those with a low level of WM capacity. Thus, WM had some effect on the online processing of plausibility information during the process of resolving temporary ambiguities at least for NSs.

Consistent with previous studies (Kim, 2004), semantic plausibility, together with the syntactic preference for a simpler analysis, had considerable influence on sentence parsing in Korean. The results for NSs generally provide support for both sentence-processing models. However, NSs showed some plausibility sensitivity in the SOR-biased plausibility condition, indicating their use of plausibility information in the subsequent structural analysis, particularly for the SR construction. Overall, the present study’s mixed results for sentence processing in Korean provide support more for parallel multiple constraint-based models than for syntax-first serial models. In addition, the results for QTs and QAs for NSs provide support for the GE and Shallow Parsing account of language processing for their final interpretations.

5.6.2. Bilingual sentence processing in Korean

NSs were compared with HSs and K2s separately in terms of whole-sentence RTs, RTs for critical positions, QTs, and QAs by considering six conditions with two Korean RC structures to examine how native-like the two non-native groups resolved temporary structural ambiguities in Korean. All the HSs were early Korean-English bilinguals living in the U.S. at the time of the experiment, whereas all the K2s lived in Korea at the time of the experiment and were late KSL (Korean as a second language) learners (i.e., they started learning Korean in their twenties) with various East Asian languages as their L1s, including Chinese and Japanese. Because of the differences in the linguistic profile between these two groups of non-native speakers, no direct comparison between these two groups was conducted.
HSs and K2s were selected based on a cloze-type test of Korean proficiency and a read-aloud RST in Korean. However, there was no test administered to determine their vocabulary knowledge of the words in the experimental sentences. About ¾ of the K2 participants were taking Korean-language classes at the Korean-language institutes, and the instructors at the institutes checked their students’ vocabulary proficiency. Thus, it is possible that the remaining 1/4 might have had some difficulty with the words in the experimental sentences.

In general, there were significant differences in RTs between these three groups, which may be due to differences in the level of their Korean proficiency. NSs read faster than HSs and K2s, and HSs read faster than K2s. In addition to this latency difference in RTs, HSs and K2s provided qualitatively mixed results for their native-like processing patterns, although HSs generally showed more native-like processing patterns than K2s.

The structural processing advantage of the SR construction over the SOR construction was examined across the three plausibility conditions, and it was found that HSs showed more native-like processing patterns than K2s. All three groups showed preferences for the SR construction in the paraphrase judgment task, which implies that the SR construction was easier to understand than the SOR construction when the plausibility information supported the SR analysis in terms of its final interpretation. On the other hand, like NSs, HSs did not show an overall processing advantage for the SR construction in terms of the RT measures. In addition, K2s did not show an SR preference in terms of whole-sentence RTs and QTs.

However, K2s were the only group who showed a structural preference for the SR construction in terms of RTs at word 6. They were sensitive to the structural difference between SR and SOR constructions and took more time reading at word 6 for the SOR construction than for the SR construction. This indicates that processing structural reanalysis at words 5 and 6 for
the SOR construction might have increased the processing cost for K2s, who had the lowest level of Korean proficiency. It is clear that, unlike NSs and HSs, K2s were sensitive to the structural difference between target constructions (at least in terms of processing times) at the end of the sentence.

In terms of the detection of plausibility information, both HSs and K2s generally showed native-like processing patterns. All three groups immediately responded to the implausibility in the SOR-biased plausibility condition. In terms of the effect of plausibility information on the initial structural decision, there were some structural differences between the groups. The initial implausibility biased toward the SOR analysis increased the processing cost at words 5 and 6 for the SR construction for NSs and HSs, but at word 6 only for K2s. That is, they were led away from the SR analysis initially but ended up with the SR analysis for the SR construction in the SR-SOR condition. The plausibility bias had the strongest effect on RTs at words 5 and 6 for HSs, whereas it had a stronger effect on RTs at word 6 than at word 5 for HSs. On the other hand, K2s were less sensitive to the implausibility for the SR construction in terms of the use of plausibility information during initial parsing than NSs and HSs.

On the other hand, for the SOR construction, all three groups showed no clear sign of the use of plausibility information in their initial structure building. This suggests that if SOR-biased plausibility information influences the reader’s initial structural analysis by leading him or her away from the SR analysis, then his or her processing cost for the ultimate SOR analysis at word 5 or 6 is lower for the SOR-SOR condition than for the SR-SOR condition (in which the SOR-bias plausibility information is in conflict with the ultimate SR analysis) or for the SOR-SR condition (in which the SR-biased plausibility information is in conflict with the ultimate SOR analysis).
The three groups showed similar RTs at words 5 and 6 for SOR-SOR and SOR-SR. In terms of SR-SOR and SOR-SOR, no group showed a significant RT difference between SR and SOR constructions, although only HSs showed longer RTs at word 6 for the SR-SOR condition than for the SOR-SOR condition ($p < .05$) in terms of the item-based analysis. This implies that the SOR-biased plausibility condition did not reduce the processing cost of structural reanalysis from the SR analysis to the SOR analysis for NSs and K2s.

All three groups immediately detected the initial implausibility in the SOR-biased plausibility condition. The use of plausibility information for initial structure building increased RTs at words 5 and 6, where they had to revert to their preferred but possibly initially deserted SR analysis for the SR construction (SR-SOR). However, it is unclear whether they used that information to prevent themselves from performing an SR-biased syntactic analysis (i.e., the MA principle) during their immediate structure building, particularly for the SOR construction.

The effect of plausibility bias on the final interpretation slightly varied according to group. The results for QTs and QAs indicate that both HSs and K2s generally showed native-like patterns for the SR construction, but less native-like processing patterns for the SOR construction. For the SR construction, all three groups showed longer QTs but lower QA rates in the SOR-biased plausibility condition (SR-SOR). They showed lower QA rates in the SR-biased plausibility condition (SOR-SR) for the SOR construction. On the other hand, HSs and K2s showed longer QTs in the SOR-biased plausibility condition (SOR-SOR) than in the other two plausibility conditions while NS showed similar QTs across the three plausibility conditions for the SOR construction. This indicates that non-native speakers showed a slightly stronger plausibility bias effect on QTs processing implausibility in the SOR-SOR condition than NSs.
In terms of the whole-sentence RTs, all three groups showed similar RT patterns. They showed longer total RTs in the SOR-biased plausibility conditions regardless of the structural difference. These results across the global measures for the three groups provide some evidence of GE processing and Shallow Parsing by all three groups. That is, they did not necessarily revise their initial analysis even when they encountered a structural cue triggering reanalysis.

In sum, HSs and K2s showed somewhat native-like processing patterns in terms of detecting the implausibility, but K2s were more sensitive to the structural difference between the two constructions than NSs and HSs. This is inconsistent with the SSH (Clahsen & Felser, 2006a, b). Unlike NSs and HSs, K2s, who had a lower level of Korean proficiency than HSs (who were exposed to Korean since birth), were sensitive to the structural difference and showed some sign of structural reanalysis for the SOR construction. Although K2s showed some sensitivity to the implausibility at word 3, that information did not modulate their reanalysis for their final interpretation. They seemed to rely more on syntactic information like L1 children do to resolve temporary syntactic ambiguity. Like young L1 children, adult K2s relied more on syntactic information to resolve temporary syntactic ambiguities. This provides no support for the SSH and suggests instead that adult L2 learners of Korean may have difficulty integrating plausibility information into syntactic parsing, more consistent with a GE-based processor (e.g., Christianson et al., 2010).

WM capacity influenced the participants’ ability to process plausibility information during the resolution of structural ambiguities, but there were some group differences. WM had main effects on RTs at words 3 and 4 for NSs and at word 3 for K2s. High-span NSs and K2s spent less time processing implausibility information than their low-span counterparts. On the other hand, WM had no main effect on RTs at words 3 and 4 for HSs. Instead, WM had main
effects on whole-sentence RTs, QTs, and QA rates for HSs. High-span HSs took less time to read sentences and to respond to the paraphrase judgment task than low-span HSs. High-span HSs were likely to show lower QA rates in the SOR-biased plausibility condition than their low-span counterparts. This implies that high-span HSs were more dependent on plausibility information in their final interpretation than their low-span counterparts. At the end of the sentence in the Neutral and SR-biased conditions, high-span HSs displayed longer reading times than their low-span counterparts. In sum, with some group differences in WM capacity, WM was more likely to influence RTs, QTs, and QAs for HSs than for NSs and K2s, whereas WM showed a significant relationship with implausibility detection for NSs and K2s.

In general, HSs were slightly more likely to show native-like processing of plausibility and syntactic preference information than K2s, and WM was more likely to influence total RTs, QTs, and QAs for HSs than for K2s. This was due in part to the overall difference in the level of their Korean proficiency, which was not controlled for in this study because of logistical issues related to recruiting participants. However, the difference in processing performance between HSs and K2s can be attributed to multiple factors such as the difference in the proficiency level and some exogenous factors such as the timing of exposure to Korean, immersed/non-immersed environments, and formal/informal instruction, among others.

K2s had diverse L1 backgrounds. Due to a smaller and unequal number of participants per each L1, however, no additional analysis of K2s was conducted to test the transfer effect of their L1s or the effect of formal/ informal Korean instruction. The two major L1s for K2s were Chinese (27 out of 42) and Japanese (13 of 42). Korean and Japanese are all head-final languages while Chinese and English are considered head-initial languages. For instance, there are differences in terms of the word order (S-O-V for Korean and Japanese and S-V-O for Chinese
and English). However, the three languages, Korean, Japanese, and Chinese share some similarities in the RC construction, such as pronominal RCs with no RC pronoun. RCs generally precede the modifying noun. Nevertheless, there are still some differences in terms of the specific construction of RCs among three languages. All these factors may lead to some differences processing Korean RCs. In this regard, future research should consider a larger sample of Chinese and Japanese K2s to provide a better understanding of L2 sentence processing in Korean.

Research on HSs’ and K2s’ language processing should provide language teachers and policymakers with valuable insights into bilingual education, particularly for identifying their educational needs and developing better language programs (Montrul, 2008). In this regard, future research should consider NSs, K2s, and HSs of comparable levels of Korean proficiency to provide a better understanding of bilingual processing in Korean.
REFERENCES


APPENDIX A
KOREAN PROFICIENCY TEST

**Instructions:** In the following text, some of the words have been replaced by spaces which are numbered from 1 to 20. First, read the complete text in order to understand it. Then reread it and choose, from the list of words on the answer sheet, not on the text.

나무

나무는 둘을 지냈다. 나무는 주어진 분수에 만족할 줄을 알았다. 나무로 태어난 것을 탐지하지 아니하고, 왜 여기 놓이고 저기 놓이지 않았는가___(1) 말하지 아니한다. 동성이에 서면 햇살이 따사로울가, 골짜기에 내려서면 물이 종이를 하여, 새로운 자리를 엽보는 일도 없다. 물과 흙과 태양의 아들로, 물과 흙과 태양이 주는 대로 받고, 적게 얻은과 불만족을 말하지 아니한다. 이웃 친구의 처지___(2) 눈에 보는 일도 없다. 소나무는 소나무대로 스스로 죽고, 진달래는 진달래대로 스스로 죽다.

나무는 고통하다. 나무는 모든 고독을 ____ (3). 안개에 잠긴 아침의 고독을 알고, 구름에 ____ (4) 저녁의 고독을 알다. 부슬비 내리는 가을 저녁의 고독도 알고, 햇살에 빗 떨리는 겨울 아침의 고독도 알다. 나무는 파리도 옷을 뜨고, 불만족을 말하지 아니한다. 이웃 친구의 처지___(2) 눈에 보는 일도 없다. 소나무는 소나무대로 스스로 죽고, 진달래는 진달래대로 스스로 죽다.

나무에 아주 친구가 없는 것은 _____(5). 달이 있고, 바람이 있고, 새가 있다. 달은 때를 아기지 아니하고 찾고, 고독한 여름밤을 같이 지내고 가는, 의리있고 다정한 친구다. 웃을 뿐만이 ______(6), 이심전심 의사가 소통되고 아주 비위에 맞는 친구다. 바람은 달과 ____ (7) 아주 변덕 많고 수다스럽고 믿지 못할 친구다. 그야말로 바람과 바람장이 친구다. 자기 ___(8) 내키는 때 찾아올 뿐 아니라, 어떤 때에는 쓸쓸수서 알랑거리고, 어떤 때에는 난데없이 휘돌러 우악스럽게 남의 팔다리에 ____ (9)를 내놓고 달아난다.


같은 나무, 이웃 나무____(13) 가장 좋은 친구가 되는 것은 두 말할 것도 없다. 나무는 서로____(14) 이해하고 진심으로 동정하고 공감한다. 서로 마주____(15) 기뻐하고, 일생을 아끼고 살아도 심증나지 않는 참다운 친구다.

그리나, 나무는 친구끼리 서로 즐긴다느니보다는, 제각기 하늘이 준 힘을 다하여 넓리 가지를 펴고, 아름다운 꽃을 _____ (16), 열매를 ____ (17) 네 힘을 쓴다. 그리고, 하늘을 ____ (18) 향상 감사하고 찬송하고 목도하는 것으로 일암한다. 그리고 나무는 항상 하늘을 향하여 손을 쳤다 쳐있다. 온갖 나뭇잎이 우거진 숲은 찾는 사람이, 거룩한 전당에 들어선 것처럼, ____ (19)하고 경건한 마음으로 절로 웃겼을 여미고, 우쭐한 찬가에 귀를 ______(20) 되는 이유도 여기에 있다.
Answer Choices
(1) a. 에  b. 가  c. 를
(2) a. 에  b. 가  c. 로
(3) a. 알다  b. 안다  c. 알았다
(4) a. 간  b. 있는  c. 덧인
(5) a. 아니다  b. 모른다  c. 일지도 모른다
(6) a. 없고  b. 없으나  c. 없었지만
(7) a. 말랐지만  b. 다르지  c. 달라서
(8) a. 마음  b. 계획  c. 느낌
(9) a. 모양  b. 상처  c. 바람
(10) a. 달아난다  b. 쫓겨난다  c. 깨어난다
(11) a. 가  b. 를  c. 로
(12) a. 가리지  b. 가릴 수  c. 가릴 줄
(13) a. 는  b. 에  c. 가
(14) a. 속속들이  b. 길이길이  c. 나날이
(15) a. 보기 때문에  b. 보기만 해도  c. 보기로
(16) a. 피고  b. 피하고  c. 피우고
(17) a. 봉이는  b. 봉는  c. 봉히는
(18) a. 우려하다  b. 우려르는  c. 우려르게
(19) a. 현란  b. 산만  c. 엄숙
(20) a. 열개  b. 귀울이게  c. 뜨이게
APPENDIX B

ENGLISH PROFICIENCY TEST

Subject #:_________

Cloze Test

Please fill in the blanks in the following passage. Each blank must have one and only one word.

Joe came home from work on Friday. It was payday, but he wasn’t __________ excited about it. He knew that _________ he sat down and paid his _________ and set aside money for groceries, _________ for the car and a small __________ in his savings account, there wasn’t _________ much left over for a good __________.

He thought about going out for __________ at his favorite restaurant, but he _________ wasn’t in the mood. He wandered __________ his apartment and ate a sandwich. __________ a while, he couldn’t stop himself _________ worrying about the money situation. Finally, __________ got into his car and started __________.

He didn’t have a destination in __________, but he knew that he wanted ________ be far away from the city __________ he lived.

He drove into a quiet country __________. The country sights made him feel __________. His mind wandered as he drove _________ small farms and he began to __________ living on his own piece of __________ and becoming self-sufficient. It had always _________ a dream of his, but he _________ never done anything to make it __________ reality. Even as he was thinking, _________ logical side was scoffing at his __________ imaginings. He debated the advantages and __________ of living in the country and __________ his own food. He imagined his __________ equipped with a solar energy panel _________ the roof to heat the house _________ winter and power a water heater. _________ envisioned fields of vegetables for canning _________ preserving to last through the winter. _________ the crops had a good yield, _________ he could sell the surplus and _________ some farming equipment with the extra _________.

Suddenly, Joe stopped thinking and laughed _________ loud, “Am I really going to _________ through with this?”
APPENDIX C
EXPERIMENTAL SENTENCES

SRSR1 나는 선배를 사랑하는 여자친구에게 이별을 고했다.
I asked (my lover) to break up because she loved my senior classmate.

SORSR1 나는 선배를 사랑하는 여자친구에게 후회한다고 말했다.
I told my girlfriend who (I) loved that I felt regretful about my senior classmate.

SRSR2 연수는 음악을 작곡하는 정화에게 피아노를 사주었다.
Yeon-Su bought a piano for Jeong-Hwa, who composed music.

SORSR2 연수는 음악을 작곡하는 정화에게 써달라고 부탁했다.
Yeon-Su asked Jeong-Hwa who composed (music) to write a piece of music.

SRSR3 아저씨는 길을 걷고있는 사람에게 길안내를 부탁했다.
A man asked a person who was walking on the road for help with the direction.

SORSR3 아저씨는 길을 걷고있는 사람에게 도와주리 다가갔다.
A man approached to a person who was walking on the road to help him.

SRSR4 나는 꽃을 아끼는 애인에게 한다발을 선물했다.
I presented a bunch of flowers to my lover who cherished flower.

SORSR4 나는 꽃을 아끼는 애인에게 자랑하려고 했다.
I tried to show off flowers to my lover who I loved.

SRSR5 과장님은 행사를 안내하는 도우미에게 심부름을 시켰다.
The section chief asked the helper who guided the event to run an errand.

SORSR5 과장님은 행사를 안내하는 도우미에게 대기하라 지시했다.
The section chief directed the helper who guided to be ready for the event.
정아는 숲을 지나가는 아저씨에게 도움을 요청했다.

Jung-A asked a man who was passing by the forest for help.

정아는 숲을 지나가는 아저씨에게 안내하려고 따라갔다.

Jung-A followed a man who was passing by to guide the forest.

영수는 감기약을 조제하는 약사에게 약값을 물어보았다.

Young-Su asked the price of the medicine to the pharmacist who was compounding the cold medicine.

영수는 감기약을 조제하는 약사에게 서두르라 부탁했다.

Young-Su asked the pharmacist who was compounding to hurry up with the cold medicine.

웨이터가 디저트를 먹고있는 우리에게 계산서를 갖고왔다.

The waiter brought a bill to us who were eating to entertain desserts.

웨이터가 디저트를 먹고있는 우리에게 접대하려고 다가왔다.

The waiter came to us who were eating to entertain desserts.

아버지의 동생을 가르치는 선생님께 전화를 하셨다.

My father gave a phone call to the teacher who taught my younger brother.

아버지의 동생을 가르치는 선생님께 데려다달라고 부탁했다.

My father told the teacher who taught to bring my younger brother back.

의사는 피를 수혈받은 환자에게 응급수술을 시켰다.

The doctor performed an emergency operation to the patient who received a blood transfusion.

의사는 피를 수혈받은 환자에게 금지하도록 지시했다.

The doctor ordered the patient who received a (blood) transfusion to prevent blood.
시경이는 클래식을 좋아하는 여자친구에게 피아노를 연주해주었다.
Si-Kyung played the piano for his girl friend that liked classic music.

시경이는 클래식을 좋아하는 여자친구에게 만나리 가자고했다.
Si-Kyung asked his girlfriend whom he liked to go to meet classic music.

영이는 살림을 잘하는 아주머니께 부엌일을 맡겼다.
Young-I left kitchen work to the woman who was good at housekeeping.

영이는 살림을 잘하는 아주머니께 감사하다고 말했다.
Young-I told her thanks to the woman who was good for her housekeeping.

사장은 유학을 준비하던 아들에게 회사를 물려주었다.
The president turned over his company to his son who was preparing for studying aboard.

사장은 유학을 준비하던 아들에게 배워보라고 권했다.
The president encouraged his son who was preparing to learn about study aboard.

경찰은 차를 불법주차하는 운전자에게 벌금을 부었다.
The policeman imposed a fine to a driver who parked his car illegally.

경찰은 차를 불법주차하는 운전자에게 알려주라고 세웠다.
The policeman stopped a car to notify the driver who parked illegally.

감독은 공을 잘차는 선수에게 패널티킥을 지었다.
The coach commaned a penalty kick to the player who was good at kicking a ball.

감독은 공을 잘차는 선수에게 쉬지말라고 지시했다.
The coach directed the player who was good at kicking not to stop the ball.

엄마가 우유병을 들고있는 아기에게 과자를 사주었다.
Mother bought cookies for a baby who was holding his milk bottle.

엄마가 우유병을 들고있는 아기에게 물려주려고 다가갔다.
Mother approached to a baby whom she was holding to give a milk bottle.
My bother bought lunch for his junior who spared money.
My brother decided to lend money to his junior who he trusted.
The wholesaler discounted the price to the person who purchased products in bulk.
The wholesaler decided to give the thing to the person with big orders.
The department head gave a bonus to the employee who left his job.
The department head promised to find a job for the employee who left.
The charity provided meals to old people who were skipping lunch.
The charity decided to provide lunch to old people who were starved.
원장님은 아이를 입양하러온 부부에게 여자애를 소개했다.
The director introduced a girl to the couple who came to adopt a child.

원장님은 아이를 입양하러온 부부에게 소개하기로 했다.
The director decided to introduce a child to the course who were for adoption.

아버지지는 신문을 배달하는 아이에게 수고비를 주었다.
Father gave a tip to the boy who delivered newspaper.

아버지지는 신문을 배달하는 아이에게 읽어달라고 했다.
Father asked the delivery boy to read the newspaper.

Ki-Tae sold a car to the staff who did business with used cars.
Ki-Tae sold the used car to the staff who he did business with.

Soon-I drew a rough map to the foreigner who did not know Korean.
Soon-I tried to teach Korean to the foreigner whom she didn’t know.

The loan shark pressed the debtor who owed money for the repayment.
The load shark pressed the debtor who owed to pay the money back.

The staff imposed a late fee on the member who delayed paying the charge.
The staff notified the member whose payment was late to pay the charge.
Father dispatched a telegram to his family who was waiting for his contact.

Father tried to contact his family who was waiting.

The fund manager recommended stocks to the customer who wanted to start the installment savings.

The fund manager recommended the installment savings to the customer who wanted to become a member.

Dad scolded a child who was fretting about money.

Dad decided to give money to the child who was fretting.

Joon-A brought a drink to her friend who was working on the photos.

Joon-A promised her friend who rested to show the photo.

He hurled criticism to the singer who did not sing about an egg very well.

He approached the signer who did not sing well to throw an egg.

Young-Mi bought a swim suit to her friend who competed in swimming with her.
Young-Mi told her friend who competed with her to go swimming together.

The president decided to give a bonus to the section head, Mr. Kim who settled an account for his son.

The president decided to send his son to the section head, Mr. Kim who settled an account.

Mom asked the fortuneteller who predicted others’ troubles about her worries

Mom went out to ask her worries to the fortuneteller.

Hyun-Soo made the person who decorated doghouses in charge of the interior.

Hyun-Soo decided to leave the doghouse to the person who was an interior designer.

The conductor asked the member who was tuning the music score to take a rest.

The conductor signaled the member who was tuning to play the music score.

The rich man decided to return thanks to the neighbors who took care of his property.

The rich man decided to distribute his property to the neighbors who he took care of.
The gangster inflicted a lynching to the victim who shouted the violence.
The gangster stated that he assaulted the victim who shouted.
The son asked his parents who helped him with his undutiful behavior for their forgiveness.
The son cried while he was apologizing for his undutiful behavior to his parents who helped him.
The husband attached a brooch to his wife who loved clothing.
The husband liked meddling clothing of his wife whom he loved.
The dweller told the landlord who was clearing a key about the inconvenience.
The dweller went to the landlord who was clearing to return the key.
The coach gave a warning to the player who came for a game.
The coach announced to ban the game to the player to came (to him).
The prosecutor sentenced an imprisonment to the criminal who was looking on about his crime.
The prosecutor made the looking-on criminal acknowledge his crime.