WORKING MEMORY AND RELATIVE CLAUSE ATTACHMENT
IN FIRST AND SECOND LANGUAGE PROCESSING

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAI’I IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

IN

ENGLISH AS A SECOND LANGUAGE

AUGUST 2005

By
Akira Omaki

Thesis Committee:
Bonnie D. Schwartz, Chairperson
Robert Bley-Vroman
John Norris
Amy J. Schafer
©Copyright 2005

by

Akira Omaki
To my mother
ACKNOWLEDGMENTS

Ever since I started my MA program in Hawai‘i in 2002, I was looking forward to writing this section of my MA thesis. I have always liked to read acknowledgments in journal articles, dissertations and so on, since I could learn who has learned what from whom and how people can develop themselves through academic and personal encounters and interactions. I had thought that it would be fun, easy but nice to list up and thank the nice people around me who helped me in many ways, but now that I am faced with this task, I realize how difficult it is to mention everyone who deserves credit for my work and to express my sincere gratitude to them in a language that fully expresses how grateful I am.

First, I must sincerely thank my advisor and thesis committee chair, Bonnie D. Schwartz, for everything she has offered and taught me (most notably ‘facetiousness’). We spent countless hours discussing various academic issues in syntax, language acquisition, and sentence processing. She also taught me how to write cover letters, curriculum vitae and grant applications. Every time I interacted with her in a classroom, in a reading group, in her office, on the little lanai where she spends many hours, I always learned something new, something that was never visible to my eyes but was visible to her unbelievably detail-oriented eyes! Her insights always amazed me, and I just wish I had had more and more hours of discussion with her. Sometimes, we also chatted about nothing while she was smoking, which was always fun despite the danger of second-hand smoke. I have learned immensely from her intellect, integrity and knowledge, and
without her supervision throughout my MA program, the quality of my academic work in
general would have never reached the current level. I will really miss visiting her office
and the little lanai.

Amy J. Schafer gave me fundamental training in psycholinguistics. Her expertise
and research experience in sentence processing inspired me a number of times, and her
careful and critical comments have trained me to think more carefully about
argumentations in psycholinguistic research. Without all the training she has provided to
me, this thesis would not have existed. Robert Bley-Vroman was also a logical and
critical thinker (on top of being an excellent contra dancing caller!), and his advice and
suggestions on various aspects of the thesis inspired many thoughts and ideas. John
Norris kindly offered help with statistical analyses, and the extensive discussions I had
with him about the use of ANCOVA and correlational analyses helped me understand
inferential statistics far better than I used to. With all of the assistance from my
committee members, I can only feel grateful that I made a right decision to come to the
University of Hawai'i to receive such wonderful education for my MA degree.

Beyond my thesis committee members, I owe thanks to many, many people in the
University of Hawai'i community who helped me grow up as a researcher. I enjoyed
learning tremendously about child language acquisition and minimalist syntax from
Kamil Ud Deen and Yuko Otsuka in the Linguistics Department. Being around Ben
Bergen and William O'Grady was always intellectually stimulating, although I regret that
I had no chance to take their courses to get to know them better. I owe thanks to Cathy
Doughty, who used to be involved in advising of my thesis, for her comments and
encouragement. I thank my friends and colleagues who shared and exchanged lots of ideas and discussions: Gerald Bullock, Wei Chu, Yumiko Enyo, Valerie Guerin, Tomomi Hasegawa, Hyekyung Hwang, Tanja Ilic, Tomoko Iwasaki, Cathy Kawahata, Rino Kawase, Younhee Kim, Soo-Ok Kweon, Sunyoung Lee (Linguistics and Second Language Studies), Michiko Nakamura, Jun Nomura, Taka Okazaki, Tomomi Sasaki, Manami Sato, Kyoung-sook Shin, Jennie Tran, Annie Tremblay, and many, many other people who made my life in Moore Hall and the LAE Labs exciting and fun throughout these three years. I owe special thanks to Matt Prior, who kindly spent hours and hours with me to improve the instruments I used in the thesis research. He also helped me at various points during my MA program with many things, and he always impressed me with his wisdom and numerous sarcastic jokes. I am indebted to Mie Hiramoto, who kindly introduced me to many of her friends when I was in desperate need of subjects for my research. I also thank Yasuko Ito, as it was always fun and informative to have discussions, exchange information and learn together with her.

My thesis research was also supported by many people outside of the University of Hawaii. I received many helpful comments at the 18th CUNY Conference on Sentence Processing at the University of Arizona, and I am particularly grateful to Sachiko Aoshima, Evan Chen, Ian Cunnings, John Hale, Scott Jackson, Charles Lin, Colin Phillips, Hiromu Sakai, Clare Stroud, Jun-ichi Tanaka, and Masaya Yoshida. Also, being one of Bonnie’s students brought me in contact with her students in Europe too (nice and smart ones, I should add!), and among those people, I thank Sharon Unsworth for her comments on the reading span test and my preliminary ideas for my thesis research as
well as for being such a nice office mate in Moore 477. I also owe thanks to Holger Hopp, one of the best L2 processing researchers, who kindly helped me with the use of E-prime in self-paced reading experiments.

My thesis research was supported in part by an Elizabeth Holmes-Carr Scholarship and a Graduate Student Organization grant, received with much appreciation. I am also grateful to the 18th CUNY Conference for providing me with student travel support.

I am deeply indebted to Ken Ariji, who really initiated my interest and experience in psycholinguistics. He gave me essential training when I was still an undergraduate student in Japan. He was a teaching assistant for the neurolinguistics course at Sophia University, and he asked me if I wanted to collaborate with him on his psycholinguistics project. We worked together a lot, or to be more precise, he trained and taught me a lot so I was able to contribute something to his project. He introduced to my life all the new and exciting things in the academic world. If I had not met him before coming to Hawaii, my academic life here would have been completely different. We stopped communicating after the project was over, but I hope we can work on a project together again.

I owe special thanks to my dear partner, Barbara Schulz. Her warm encouragement and support helped me immensely when I was under stress with my research and teaching, despite the fact that she was also constantly going through the battle with stress and pressure by being a doctorate student in our program. She was not only an emotional support for me, but also an important colleague who assisted me with various aspects of my academic work. She read earlier drafts and offered many, many
valuable comments, which contributed immensely to the improvement of my thesis. She also made lots of efforts to make sure that I can live a more healthy life, and most importantly, she taught me the importance of enjoying life.

Finally, I sincerely thank my parents, Yasuko and Chuichi Omaki, for their support and encouragement. I am particularly indebted to my mother, who has always supported me in my pursuit of academic success, ever since I was a little boy. I really owe her the biggest thanks of my life forever.
ABSTRACT

This thesis investigates resolution of Japanese and English relative clause (RC) attachment ambiguity to shed light on the role of working memory in first and second language processing. English natives and advanced Japanese learners of English participated in off-line and on-line experiments. The results show that (a) working memory capacity influenced English natives' off-line but not on-line RC attachment preferences, (b) Japanese and English RC attachment preferences of the Japanese learners were not associated with working memory capacity, and that (c) when examined individually, some learners showed target-like RC attachment preferences while others seemingly transferred their Japanese RC attachment preferences.
# TABLE OF CONTENTS

Acknowledgments ........................................................................... v
Abstract ......................................................................................... x
List of tables .................................................................................... xiv
List of figures ................................................................................... xv
List of abbreviations ........................................................................ xvi

## Chapter 1: Introduction ............................................................... 1
1.1. Transfer, learning, and working memory in second language processing ............................... 1
1.2. Overview .................................................................................. 5

## Chapter 2: Working memory in sentence processing .................. 7
2.1. Working memory and syntactic processing ................................................. 7
2.2. Reading span test ........................................................................ 19
   2.2.1. Daneman and Carpenter's reading span test and its limitations ................. 20
   2.2.2. The design of the reading span test used in the present study ............... 25
2.3. Conclusion ................................................................................. 30

## Chapter 3: Relative clause attachment in L1 and L2 processing ... 32
3.1. Cross-linguistic investigations of relative clause attachment ................................. 32
3.2. Relative clause attachment in L2 sentence processing ..................................... 39
3.3. Relative clause attachment and working memory capacity ............................ 43
3.4. Conclusion ................................................................................. 46

## Chapter 4: Attachment preference data from English native speakers ............ 48
4.1. Experiment 1 ........................................................................... 48
   4.1.1. Method ............................................................................ 53
      4.1.1.1. Participants ......................................................... 53
      4.1.1.2. Materials .............................................................. 53
      4.1.1.3. Procedure ............................................................ 55
      4.1.1.4. Data analysis ....................................................... 57
   4.1.2. Results ........................................................................... 58
      4.1.2.1. Reading span ....................................................... 58
      4.1.2.2. Comprehension data ........................................... 59
   4.1.3. Discussion ....................................................................... 63
4.2. Experiment 2 ........................................................................... 65
   4.2.1. Method ........................................................................... 66
      4.2.1.1. Participants ......................................................... 66
      4.2.1.2. Materials .............................................................. 66
      4.2.1.3. Procedure ............................................................ 68
      4.2.1.4. Data analysis ....................................................... 69
   4.2.2. Results ........................................................................... 70
      4.2.2.1. Reading span ....................................................... 70
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>RST sentence types, regions, and duration</td>
<td>28</td>
</tr>
<tr>
<td>2.</td>
<td>Individual data in Experiment 1</td>
<td>60</td>
</tr>
<tr>
<td>3.</td>
<td>Regions of interest in the target sentences in Experiment 2</td>
<td>74</td>
</tr>
<tr>
<td>4.</td>
<td>Profiles of the Japanese participants</td>
<td>93</td>
</tr>
<tr>
<td>5.</td>
<td>Individual data from Experiment 3</td>
<td>102</td>
</tr>
<tr>
<td>6.</td>
<td>Individual data from Experiments 1 and 3</td>
<td>107</td>
</tr>
<tr>
<td>7.</td>
<td>L2ers’ individual data from Experiments 4 and 3</td>
<td>115</td>
</tr>
<tr>
<td>8.</td>
<td>Regions of interest in the target sentences (same as Table 3)</td>
<td>129</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Mean non-local attachment response by span group in Experiment 1...........</td>
<td>61</td>
</tr>
<tr>
<td>2.</td>
<td>Mean comprehension accuracy in the target conditions by span groups in Experiment 2</td>
<td>72</td>
</tr>
<tr>
<td>3.</td>
<td>Mean reading times in crucial regions in Experiment 2</td>
<td>75</td>
</tr>
<tr>
<td>4.</td>
<td>Mean reading times by span group in crucial regions in Experiment 2........</td>
<td>78</td>
</tr>
<tr>
<td>5.</td>
<td>Mean reading times (in ms) by span group in Region 12 in Experiment 2....</td>
<td>80</td>
</tr>
<tr>
<td>6.</td>
<td>Mean reading times (in ms) by span group in Region 9 in Experiment 2.....</td>
<td>81</td>
</tr>
<tr>
<td>7.</td>
<td>Mean reading times in crucial regions in Experiment 5</td>
<td>130</td>
</tr>
<tr>
<td>8.</td>
<td>Mean reading times (in ms) in Region 9 by low-span English native speakers and overall Japanese L2ers</td>
<td>132</td>
</tr>
</tbody>
</table>
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc</td>
<td>Accusative</td>
</tr>
<tr>
<td>Comp</td>
<td>Complementizer</td>
</tr>
<tr>
<td>Dat</td>
<td>Dative</td>
</tr>
<tr>
<td>Gen</td>
<td>Genitive</td>
</tr>
<tr>
<td>Nom</td>
<td>Nominative</td>
</tr>
<tr>
<td>Pass</td>
<td>Passive</td>
</tr>
<tr>
<td>Prog</td>
<td>Progressive</td>
</tr>
<tr>
<td>Pst</td>
<td>Past</td>
</tr>
<tr>
<td>Top</td>
<td>Topic</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1. Transfer, learning, and working memory in second language processing

A goal of sentence processing research is to figure out the properties of the parsing mechanism, a mechanism that rapidly analyzes linguistic input to achieve sentence comprehension. Over a decade of first language (L1) processing research has revealed that various constraints, such as linguistic information (syntactic, semantic, and prosodic), discourse information, and working memory capacity, play important roles in sentence comprehension (for a review, see Clifton & Duffy, 2001; Gibson, 1998; Gibson & Pearlmutter, 1998). However, little is known about how sentence processing in a second language (L2) is constrained. One particularly important factor that has not been fully addressed in L2 processing is working memory. Working memory is a component of cognition that allows humans to simultaneously store certain information while processing or computing other information, and this plays an important role in language processing as well.¹ It has been pointed out in the L1 psycholinguistic literature that individual differences in working memory capacity may influence parsing behavior (e.g., Just & Carpenter, 1992). Assuming that parsing in L2 places a higher burden on the processing mechanism since retrieval of lexical and grammatical knowledge is more difficult, it is possible that the overall picture of L2 processing may be complicated by individual

¹ For example, filler-gap dependencies as observed in wh-movement (e.g., Fiebach, Schlesewsky, & Friederici, 2002; Gibson & Warren, 2004) or scrambling (e.g., Aoshima, Phillips, & Weinberg, 2004) require storage of the filler (i.e., the dislocated phrase) up to the gap position while processing other intervening constituents.
differences in working memory capacity. In line with Just and Carpenter's (1992) proposal that the parser adopts different parsing strategies depending on the size of working memory capacity, L2 learners (L2ers) in general may behave like those native speakers who have smaller working memory capacity since L2ers may have to consume a great amount of resources in L2 processing. Alternatively, L2ers may behave 'target-like' only if their working memory capacity is sufficiently large, since otherwise their computational resources may run out and the parse may fail before it is completed. For example, if there is a construction that native speakers of one language process differently from native speakers of another language, and if speakers of the former language learn the other language as their L2, then it is possible that these L2ers' acquisition of target-like processing strategies may be modulated by the learners' working memory capacity, in that only those learners who have large working memory capacity will be able to learn to utilize target-like processing strategies and not be overwhelmed by L2 processing difficulties. In this sense, working memory capacity may in fact play important roles in explaining L2ers' acquisition of target-like sentence processing strategies (or lack thereof).

This thesis presents a series of off-line and on-line experiments\(^2\) that examine the nature of L2 sentence processing by advanced Japanese-speaking learners of English in comparison to that of English native speakers. The target phenomenon examined in all of the experiments is the resolution of relative clause attachment ambiguity, as exemplified in (1).

\(^2\) Off-line experiments employ a dependent variable that is measured after the whole sentence is processed (e.g., acceptability judgment, comprehension question, etc.), while on-line experiments employ a dependent variable that is measured during sentence processing (e.g., reading time, eye movement, etc.).
(1) Someone shot the servant of the actress who was on the balcony.

This construction is globally ambiguous since the relative clause “who was on the balcony” can modify either the non-local NP “the servant (of the actress)” or the local NP, “the actress.” The resolution of relative clause attachment ambiguity provides an ideal test case for investigating L1 transfer and working memory capacity, the two factors that this thesis is concerned with. Firstly, it has been shown that cross-linguistic differences exist in the way speakers resolve this type of structural ambiguity (e.g., Cuetos & Mitchell, 1988). Kamide and Mitchell (1997) found that Japanese speakers show a different attachment preference from that of English speakers in off-line processing but exhibit a similar preference to that of English speakers in on-line processing. In this context, testing how Japanese L2ers of English resolve English relative clause attachment ambiguity off-line and on-line can address the question of whether L1 processing behavior influences L2 processing behavior. Secondly, Mendelsohn and Pearlmutter (1999) and Swets, Desmet, Hambrick, and Ferreira (2004) found in their L1 off-line processing studies that relative clause attachment preferences are influenced by a subject’s working memory capacity, in that subjects with larger capacities tended to have a different attachment preference from those with smaller capacities. In this context, testing the interaction of working memory capacity and L2ers’ relative clause attachment preferences can reveal to what extent working memory constrains L2 sentence processing. For these reasons, investigations of (a) how Japanese L2ers process relative clause attachment ambiguity in English and (b) whether working memory capacity affects these L2ers’ processing of relative clause
attachment ambiguity will shed light on the role of L1 transfer and working memory capacity in L2 sentence processing.

Furthermore, the thesis examines the interaction of sentence complexity and relative clause attachment preferences as an additional test of the relation between working memory and ambiguity resolution. Using a different kind of syntactic ambiguity than the one investigated here, Eastwick and Phillips (1999) showed that increased sentence complexity can make (native) subjects with larger working memory capacities behave like those with smaller capacities. The present study extended the same logic to relative clause attachment ambiguity in order to test whether working memory influences attachment preferences.

The combined use of off-line and on-line methods pursued in the present study can also inform the debate in the L1 psycholinguistic literature concerning theories of working memory resources in sentence processing. This debate centers on the question of whether there is a single working memory resource for all kinds of cognitive/linguistic processes (e.g., Just & Carpenter, 1992) or whether there is a dedicated resource for sentence processing (e.g., Caplan & Waters, 1999). Caplan and Waters (1999) argued that the verbal working memory system is divided into two subsystems, namely, interpretive processing and post-interpretive processing, and they further argued that no individual differences exist in interpretive processing, which roughly corresponds to on-line syntactic processing, but that differences are observed in post-interpretive, off-line processing of the meaning extracted from the on-line processing. In this sense, the off-line studies that have found a relation between attachment preferences and memory capacity may have only captured
post-interpretive processing rather than interpretive processing. The thesis attempts to shed light on this issue by comparing the results from off-line and on-line methods targeting relative clause attachment.

1.2. Overview

This thesis is structured as follows. Chapter 2 summarizes and discusses the L1 and L2 psycholinguistic literature with respect to the role of working memory in sentence processing, highlighting why it might be an important factor in L2 sentence processing. In this context, the chapter also discusses potential problems with the standard reading span test by Daneman and Carpenter (1980), and illustrates the design of the reading span test devised in Ariji, Omaki, and Tatsuta (2003) and Omaki and Ariji (2003), which is adopted in the current study. Chapter 3 then reviews past studies that have examined relative clause attachment ambiguity in L1 and L2 processing. This chapter also introduces various explanations and hypotheses regarding relative clause attachment preferences in L1 and L2 processing, which establish the motivation for the experiments reported in the subsequent chapters.

Chapters 4 and 5 report the methods and results of the experiments that were designed to investigate the relations between off-line and on-line relative clause attachment preferences and working memory capacity. Chapter 4 presents the results of the off-line and on-line experiments with English native speakers, which establishes a baseline for comparison with the Japanese L2ers of English. In Chapter 5, the results of the same English experiments with the Japanese L2ers of English are presented and examined in
comparison to those of the native speakers. This chapter also presents the results from a
Japanese off-line experiment, which was designed to directly compare the L2 subjects’ L1
and L2 off-line relative clause attachment preferences.

Finally, Chapter 6 provides the general discussion of the results presented in
previous chapters, examining what implications the current study has for the investigation
of the role of working memory in L1 and L2 sentence processing.
CHAPTER 2
WORKING MEMORY IN SENTENCE PROCESSING

This chapter reviews the role of working memory in sentence processing in order to motivate the investigation of this variable in the present study. The first part of this chapter presents a review of the role of working memory for theories of sentence processing. Particular attention will be given to studies that have examined the effects of individual differences in working memory on sentence processing behavior. It is argued that this can be a major complicating factor in investigations of L2 processing, since L2ers in general are likely to suffer from resource limitations. In the second half of the chapter, issues concerning the measurement of verbal working memory capacity are discussed. The standard test used in psycholinguistics to assess working memory capacity is the reading span test, a test originally developed by Daneman and Carpenter (1980). It has been pointed out, however, that their original test has serious flaws (Ariji et al., 2003; Waters & Caplan, 1996b). Solutions to these problems that have been suggested are discussed in turn, and this chapter will conclude by illustrating the design of the reading span test adopted in the current study, namely, the one developed in Ariji et al. (2003) and Omaki and Ariji (2003).

2.1. Working memory and syntactic processing

Sentence processing is a cognitively complex and demanding task. The parser has to process the lexical items that are linearly and rapidly presented to it, while it also needs
to consider multiple sources of linguistic and discourse information to achieve accurate comprehension. In dealing with such a complex task, the parser is also constrained by the capacity of a short-term memory system: The parser can hold only a limited amount of information while processing the continuously incoming lexical items.

Based on such observations, it has been posited from the early days of sentence processing research that memory constraints play an important role in sentence processing. For example, a very influential processing principle called *Minimal Attachment* (Frazier, 1978) states that the parser always prefers to build as simple a structure as possible, and this principle was partly motivated by the idea that simpler structures would be easier to retain in working memory (Frazier & Fodor, 1978). Another processing principle called *Late Closure* (Frazier, 1978), *Recency* (Gibson, Pearlmutter, Canseco-Gonzalez, & Hickok, 1996), or *Right Association* (Kimball, 1973) states that when there is a structural attachment ambiguity where its resolution patterns do not differ in terms of the structural complexity of the resulting parse, the parser will favor local integration, that is, attaching the incoming material to the phrase currently being processed. This is based on the assumption that non-local integration (i.e., attaching incoming material to a constituent that had been processed earlier) would require reactivation of this early constituent. Therefore, non-local integration is more demanding in terms of memory resources than local integration, which consequently requires no reactivation. These processing principles have received robust empirical support (for Minimal Attachment, e.g., Clifton, Traxler, Mohamed, Williams, Morris, & Rayner, 2003; Ferreira & Clifton, 1986; Frazier & Rayner, 1982; Rayner, Carlson, & Frazier, 1983; but cf. Trueswell, Tanenhaus, & Garnsay, 1994;

Memory constraints also play an important role in explaining perceived syntactic complexity in the following types of sentences:

(2) a. Nested sentence

The reporter [who the senator [who John met] attacked] disliked the editor.

b. Right-branching sentence

John met the senator [who attacked the reporter [who disliked the editor]].

(Examples from Gibson, 2000)

The words used and ideas expressed in (2a) and (2b) are identical, but for some reason (2a) is perceived as much more complex than (2b). Gibson (1998, 2000) explained their differences in complexity by defining complexity based on memory resources. Since Gibson’s model becomes important in the current study, the details of the model are reviewed here.

Gibson posits that working memory resources are necessary for two aspects of language comprehension: (a) storage of the structure built thus far, and (b) integration of an incoming word into the structure built up to that point by connecting it to a syntactic head processed earlier. These two aspects of comprehension require memory resources. Gibson (1998) defines storage costs in terms of the number of syntactic heads that are
predicted so as to form a complete grammatical sentence. Integration costs are defined based on the number of intervening discourse referents between the newly integrated word and the syntactic head with which the new word is integrated (Gibson, 2000). In the example of a nested sentence such as (2a), both storage costs and integration costs are higher than in a right-branching sentence such as (2b). First, in (2a), the storage costs become largest at the point of “John” since five syntactic heads are required to form a grammatical sentence, namely, a verb for “the reporter,” an empty category position for the first “who,” a verb for “the senator,” another empty category position for the second “who,” and a verb for “John.” In (2b), on the other hand, the largest storage cost incurred in this sentence is at either the first or second “who,” since two syntactic heads are required to form a grammatical sentence, namely, a verb and an empty category position for the relative pronoun. Next, assuming that one energy unit is required for each discourse referent, the integration costs in the nested sentence (2a) become largest at the point of “attacked” with seven energy units: One energy unit is required for the event referent indicated by the verb “attacked”; two energy units for the integration of the verb “attacked” to its subject NP “the senator” as it crosses two intervening discourse referents, namely “John” and the event referent indicated by “met”; four energy units for co-indexing the object of “attacked” with the preceding relative pronoun “who” which is separated by four discourse referents, namely “the senator,” “John,” the event referent indicated by “met,” and the event referent indicated by “attacked.” On the other hand, the maximal integration cost in (2b) is one, since all the integrations are local and do not span over any intervening

---

1 The definition of storage costs slightly differs between Gibson (1998) and Gibson (2000), but it does not make a difference in predictions about the complexity in nested sentences in (2), so the definition from only Gibson (1998) is presented here.
discourse referents. In this way, Gibson's (1998, 2000) explanation of linguistic complexity based on memory resources can provide an elegant account of the observed complexity difference between (2a) and (2b), and his model has received robust empirical support in various reading time and neurolinguistic experiments on English (e.g., Chen, Gibson, & Wolf, 2005; Kaan, Harris, Gibson, & Holcomb, 2000; Warren & Gibson, 2002), German (Felser, Clahsen, & Münte, 2003) and Japanese (Babyonyshev & Gibson, 1999; Nakatani & Gibson, 2003).

In these ways, memory constraints have played an important part in understanding the nature of sentence processing mechanisms. These processing principles and theories were mainly concerned with working memory constraints as a factor that is universally shared by all individuals. Starting with King and Just's (1991) research on processing of subject and object relative clauses, however, psycholinguists have also asked whether there are individual differences in parsing behavior as a function of individual differences in working memory capacity. King and Just (1991) conducted a word-by-word self-paced reading task, in which subjects had to read each word of a sentence by button-pressing, and additionally remember the final word of sentences in each sentence set where the set size ranged from one to three sentences. They also measured the subjects' reading span size by using Daneman and Carpenter's (1980) reading span test and compared the performance of subjects who had a high span (i.e., high-spans) to those who had a low span (i.e., low-spans). They found that the low-spans had longer reading times than the high-spans in the complex object relative clause regions, which led them to claim that low-spans experience more difficulties than high-spans when processing syntactically complex sentences.
Following King and Just's research on the impact of differences in working memory capacity, Just and Carpenter (1992) investigated the resolution of Main Verb/Reduced Relative (MV/RR) ambiguity in sentences such as (3):

(3)  
a. The defendant examined by the lawyer shocked the jury.  
b. The defendant that was examined by the lawyer shocked the jury.  
c. The evidence examined by the lawyer shocked the jury.  
d. The evidence that was examined by the lawyer shocked the jury.

(3a) and (3c) both involve temporary MV/RR ambiguity, since, at the point of processing the verb “examined,” these sentences are ambiguous between a main verb analysis (as in “the defendant examined the microphone”) and a reduced relative clause analysis, the latter of which turns out to be correct. On the other hand, (3b) and (3d) are unambiguous, unreduced relative clauses, and these conditions served as control sentences for (3a) and (3c). In this context, it is possible that the semantic cue provided by the inanimate subject “the evidence” in (3c) signals that it is unlikely to be assigned the thematic role of agent (as compared to theme), and if such information is available from the beginning, then for (3c) the parser may opt for a reduced relative clause analysis or compute both analyses simultaneously. On the other hand, in (3a), where the subject “the defendant” is an animate NP which is likely to be assigned an agent thematic role, the parser is simply guided by the structural principle of Minimal Attachment and is likely to choose a simpler structure, that is, the main verb analysis, in turn engendering a slow-down in reading time in the
disambiguating “by the lawyer” region. Ferreira and Clifton (1986) investigated whether there is such a semantic effect in the early stages of processing, but they found that in both sentences like (3a) and (3c) subjects were garden-pathed in the disambiguating region when compared to unambiguous conditions like (3b) and (3d). Hence, they reasoned that semantic information does not play a role in early stages of processing. However, Just and Carpenter (1992) examined the processing of similar sentences by high-spans and low-spans and found that the low-spans were garden-pathed to a similar extent in both (3a) and (3c), as observed in Ferreira and Clifton (1986), but that the high-spans processed the inanimate condition (3c) much faster than the animate condition (3a). Based on this finding, Just and Carpenter claimed that high-spans and low-spans adopt different parsing strategies. High-spans had sufficient resources to use the animacy information to compute both the main verb analysis and the reduced relative analysis simultaneously. On the other hand, the low-spans did not have sufficient resources to do so and therefore could only adopt the simpler main verb analysis, which consequently leads to a garden-path effect (cf. Eastwick & Phillips, 1999).

A study by MacDonald, Just, and Carpenter (1992) lent further support to the view that individual differences exist in parsing behavior. MacDonald et al. examined the processing of sentences like (4), which also contain MV/RR ambiguity.

(4) a. The experienced soldiers warned about the dangers before the midnight raid.
   b. The experienced soldiers warned about the dangers conducted the midnight raid.
In both (4a) and (4b), “warned about the dangers” is ambiguous between (a) a main verb plus a prepositional phrase and (b) a reduced relative clause without a relative pronoun. MacDonald et al. (1992) found, somewhat counter-intuitively, that high-spans had longer reading times than low-spans in the final region of both sentences like (4a) and (4b). They interpreted this to indicate that high-spans were computing both analyses throughout the ambiguous region (i.e., “warned about the dangers”) and hence showed longer reading times, whereas the low-spans simply maintained a more frequent, simpler analysis (i.e., main verb analysis, as in (4a)) out of the two possible analyses.

In summary, these three studies have been claimed to show that high-spans and low-spans behave differently, with high-spans computing multiple analyses simultaneously and low-spans being committed to a single analysis. Since these researchers all view low-spans as suffering from lower processing resources, Caplan and Waters (1999) termed this perspective the Single Resource model, which assumes that there is but one resource of working memory which is shared across all cognitive domains; low-spans will have low processing resources for sentence processing and hence behave differently from high-spans (Just & Carpenter, 1992; MacDonald et al., 1992). Also, this model predicts an interaction of syntactic complexity and working memory capacity such that low-spans will show substantially more processing difficulties than high-spans in processing syntactically complex sentences (King & Just, 1991).

However, in Caplan and Waters’ (1999) critical reanalysis of these studies, they argued that none provides compelling support for the view that low-spans and high-spans behave differently. Regarding the King and Just (1991) relative clause study, for example,
Caplan and Waters point out that no statistical analysis was reported on the reading time data of the low-spans and high-spans for the region in which the processing demand was expected to be largest, and hence where low-spans were supposed to show substantially slower reading times. Since the numerical differences between low-spans and high-spans may not be statistically significant, King and Just’s conclusion may not be warranted. As for the Just and Carpenter (1992) study, Waters and Caplan (1996a) point out that Just and Carpenter did not make a crucial comparison in their data analysis. Just and Carpenter only compared reading times between the animate and inanimate reduced relative conditions (see, e.g., (3a) and (3c) above), but this comparison does not show the presence or absence of a garden-path effect, which is the crucial evidence needed to support their view. The right comparison is between the temporarily ambiguous condition and the unambiguous control condition, that is, between (3a) and (3b) or between (3c) and (3d). The garden-path effect holds only when the ambiguous conditions are substantially longer than the unambiguous control conditions: Faster reaction times in (3c) in comparison to (3a) may be meaningless if inanimate conditions overall are faster than the animate counterparts. In fact, Waters and Caplan (1996a) reanalyzed Just and Carpenter’s data and demonstrated that garden-path effects do exist in both the animate and inanimate conditions for both low-spans and high-spans, and the high-spans’ reading times are faster for sentences with an inanimate subject in both the ambiguous and unambiguous conditions. Thus, even though the fact that only high-spans processed inanimate ambiguous and unambiguous conditions faster may indicate that high-spans were somewhat sensitive to the animacy information, this does not demonstrate that high-spans and low-spans are different in terms
of syntactic processing behavior. Finally, turning to MacDonald et al. (1992), Caplan and Waters point out that the reading time differences MacDonald et al. observed are only in the final region that follows the critical, ambiguous region, and that the reading time data within the ambiguous region, for example, "warned about the dangers" in (4), show no significant differences between high-spans and low-spans. If, as MacDonald et al. claimed, the high-spans alone are computing multiple analyses, then the reading time differences should be observed in the ambiguous region, and the lack of such effects seems to constitute counter-evidence to their claim. In addition to these criticisms, Caplan and Waters report on their lab's replications of such studies that showed no such differences as these previous studies claim to have found.

Based on these results, Caplan and Waters (1999) proposed the *Separate-Sentence-Interpretation-Resource* model of verbal working memory, in which there are two subsystems dedicated to distinct language tasks. One is concerned with *interpretive processing*, which refers to syntactic processing that consists of word recognition, construction of a syntactic, prosodic and semantic representation, and assignment of thematic roles. Such interpretive processing leads to the extraction of meaning from the sentence currently being processed. The other subsystem of verbal working memory is concerned with *post-interpretive processing*, which refers to processes in which the meaning extracted through interpretive processing is used to accomplish other verbal tasks, such as storing information in long-term semantic memory, reasoning based on the sentence meaning, planning an action based on the sentence meaning, and so forth. Furthermore, Caplan and Waters argued, based on their critical review of a number of
psycholinguistic studies that looked at relations between various types of verbal tasks and working memory in normal, aging, and aphasic populations, that there is no individual difference in the resources of the interpretive processing component, whereas such differences exist in the post-interpretive processing component. This proposal of a finer distinction within verbal working memory challenges the view represented by the Single Resource model, which contends that individual differences exist in sentence processing behavior. However, the debate still continues (e.g., Fiebach, Schlesewsky, & Friederici, 2002; Vos, Gunter, Schriefers, & Friederici, 2001).  

Now let us turn to the implications of these studies for L2 sentence processing research. According to the Single Resource model, it is possible that limited processing resources influence sentence processing behavior. Given that L2ers, be they high-spans or low-spans, require more working memory resources in processing L2 input as a result of their unstable L2 knowledge, one might expect them to behave differently from native speakers who have stable knowledge of their native language. Alternatively, high-span L2ers, who might have sufficient processing resources, may be able to behave similarly to native speakers, while low-span L2ers would suffer from the limitation of memory resources and fail to process L2 input efficiently in a target-like manner. If this is true, then averaging data across L2ers without taking into consideration their working memory capacity may potentially mis-represent the nature of L2 sentence processing. Even if the overall performance of one L2 group appears different from that of a native speaker group, a comparison between high-span L2ers and native speakers may show no difference. In

---

4 See MacDonald & Christiansen (2002) for yet a different view on the role of working memory in language processing (see also Caplan & Waters, 2002, or Roberts & Gibson, 2002, for a critical review of MacDonald & Christiansen's proposal).
other words, working memory capacity may act as a complicating factor in interpreting L2 processing data in general.

So far, however, few studies have investigated whether working memory capacity influences L2ers' sentence processing behavior. Juffs (2004) examined the following: (a) whether there is a correlation between the reading time in a critical region of garden-path sentences and working memory capacity based on reading span or word span size, and (b) whether working memory capacity interacts with the two sentence types he tested, that is, garden-path versus non-garden-path sentences. The results show two things. First, there was no significant correlation between reading times in garden-path sentences and working memory measures. Second, low-spans showed significantly slower reading times in the experiment than high-spans, but the effect of memory capacity did not interact with sentence complexity, since this difference was obtained in both the garden-path and the non-garden-path conditions. Juffs (2005) also tested whether there is a correlation between the reading time in a critical region of a long-distance subject extraction sentence and working memory measures, but again he found no significant correlations. These results appear to indicate that working memory capacity does not influence L2 sentence processing behavior, but it is clear that more empirical studies are needed to confirm this possibility.

5 It is questionable whether examining correlations with raw reading times in one condition and working memory measures is a valid procedure, given that individual differences in reading times may be dependent on factors other than working memory, such as word recognition speed. Correlational analyses may become useful if one were to investigate the interaction of working memory and different sentence types, for example, examining the correlation between span size and reading time in one condition subtracted from the reading time in the other condition. This so-called residual effect would only reflect a difference between two conditions, and hence this would be a better reflection of an effect of sentence type than raw reading time data are.
One concern in investigating the relation between working memory and sentence processing is how to measure working memory capacity precisely. Most psycholinguistic studies dealing with working memory as a variable in sentence processing, including the L2 studies by Juffs (2004, 2005), have used variants of Daneman and Carpenter’s (1980) reading span test. However, this test has been criticized as potentially invalid as a measure of working memory capacity (Ariji et al., 2003; Waters & Caplan, 1996b). As the present thesis also measures working memory capacity by a reading span test in English (Experiments 1, 2, 3 and 5) and in Japanese (Experiment 4), the next section reviews the literature on the assessment of working memory capacity and discusses the reading span test that the current study adopted.

2.2. Reading span test

The very first reading span test was designed by Daneman and Carpenter (1980) as a measure of verbal working memory capacity. Various versions of reading span tests have since been proposed, but Daneman and Carpenter’s (1980) reading span test still remains popular among sentence processing studies that examine the effect of span size on processing behavior (e.g., Juffs, 2004, 2005; Just & Carpenter, 1992; MacDonald et al., 1992). However, as has been pointed out by Waters and Caplan (1996b), Roberts and Gibson (2002), Ariji et al. (2003), *inter alia*, this original reading span test (henceforth D&C reading span test) is not without limitations, and these limitations could potentially obscure the relationship between span size and sentence processing. This section therefore first discusses problems with the D&C reading span test, and then illustrates a variant of
Waters and Caplan's (1996b) reading span test that was developed in Ariji et al. (2003) and Omaki and Ariji (2003).

2.2.1. Daneman and Carpenter's reading span test and its limitations

Since Baddeley and Hitch (1974), the concept of short-term memory has shifted from one in which it is only seen as a short-term buffer (e.g., in the dual storage model of Atkinson & Shiffrin, 1968) to one of working memory capacity in which the memory system consists of both storage and processing components. According to this change, the traditional measurement of working memory capacity had to be adjusted, since older methods such as word span tests or digit span tests seem to reflect only the short-term storage component of the working memory system.

Motivated by this lack of a proper measurement of verbal working memory capacity, Daneman and Carpenter (1980) designed a reading span test, which was intended to tax both storage and processing resources. The D&C reading span test aims to achieve this by using the following procedure. The researcher shows subjects a set of sentences written on a sheet of paper, which they have to read aloud. After reading the sentences in each set, they are asked to recall the final word (i.e., target word) of each sentence in the set. The number of sentences in each set steadily increases (usually from a two-sentence condition to a six-sentence condition), and there are five trials in each sentence condition. The reading span size is then defined as the maximum number of sentences the subjects are able to read aloud while recalling all the target words in the majority of the trials, that is, in minimally three out of five trials. Using this reading span test, Daneman and Carpenter
found that performance on other reading measurements (e.g., verbal SAT) significantly correlated with reading span size, unlike the traditional span tasks such as digit span or word span (for a review of such correlational studies, see Daneman & Merikle, 1996). Another finding of Daneman and Carpenter (1980) was that such tasks as reading span tests that tax both processing and storage functions tend to show clear individual differences, which indicates that, unlike traditional span tasks, reading span tests may be able to reveal more clearly the working memory constraints on complex cognitive tasks, such as language comprehension.

However, some problems pointed out in later studies may call into question its validity as a measure of working memory capacity. Five major problems with the D&C reading span test, based on Waters and Caplan (1996b), Roberts and Gibson (2002) and Ariji et al. (2003), are summarized below.

First, despite Daneman and Carpenter's intention to tax both the processing component and the storage component of working memory capacity, their reading span test has no way of ensuring that the processing component is in fact taxed (Ariji et al., 2003). In their procedure, the processing component was supposed to be burdened as a result of the read-aloud portion of the task. Notice, however, that reading aloud does not necessarily mean that the subjects are processing the sentences for meaning. In other words, although subjects may be attending to the sound of each lexical item to ensure correct pronunciation, they may not be processing syntactic and semantic information of the sentences for the purpose of comprehension. In this sense, the D&C reading span test does not ensure that the sentences are being processed, and hence does not ensure that the processing
component of working memory is taxed in the manner that it normally is in language comprehension.

Second, the fact that the test does not have a measure of processing performance allows subjects to focus only on memorizing the target word rather than reading the sentences (Roberts & Gibson, 2002; Waters & Caplan, 1996b). This would indeed defeat the purpose of the task, as performance on this test may only reflect how many target words subjects can store in short-term memory, that is, the storage component of working memory, which is exactly what the traditional span tests such as word span or digit span were designed to measure. Furthermore, since processing performance is not measured, reading span size based solely on recall performance may not be consistent across subjects, as subjects can choose to selectively attend to either one of the two portions of the task, that is, the read-aloud part or the recall part. For example, one subject may attend more to reading aloud and hence fail to perform well on the recall part, while a different subject, whose memory capacity itself is equal to the other subject’s, may score very well by selectively attending to the recall part. Based on the fact that the D&C reading span test allows subjects to pay attention to one of the two portions of the task, Waters and Caplan (1996b) noted that a high score on the D&C reading span test may only reflect how skillfully subjects shift their attention away from the read-aloud part to the recall part, rather than reflect the capacity of the working memory system itself. This problem of trade-off between storage and processing tasks could be solved if the D&C reading span test took into consideration both processing and storage performance, but it lacks a measure of the processing performance and hence the validity of the reading span test.
scores is called into question. A valid assessment of working memory capacity needs to have an objective measure of both storage and processing performance, if we are to interpret reading span as a reflection of working memory capacity.

Third, D&C's design, in which the subjects have to memorize the final word of the sentences, allows them to store the target word after processing the sentence. Hence this design does not tax the processing and storage component simultaneously within each trial (Ariji et al., 2003). However, as the primary role of working memory in language comprehension is to store certain information temporarily while processing new information, using the final word as the target word to be memorized seems completely different from what happens in language comprehension. Consider the following sentence with a long-distance filler-gap dependency:

(5) \[ NP \text{What kind of books}, \text{do you think} t_i \text{that Barbara likes} t_i ? \]

In this sentence, the moved NP "what kind of books" needs to be retained in working memory initially up to the intermediate trace position and then to the sentence final position where the thematic role of theme can be assigned by the verb "likes," while the new information, that is, the intervening lexical items "do you think" and "that Barbara likes" is being processed. In this way, the role of working memory in language comprehension is best characterized as the storage of information while other information is processed, but the D&C reading span test fails to capture this point since memorization of the final word occurs after processing of the sentence.
Fourth, the sentences used in the D&C reading span test were randomly taken from magazines and hence not well controlled (Ariji et al., 2003). The length and complexity of the sentences varied widely across items, and such variation may inadvertently affect subjects' performance. For example, if the 3-sentence condition happened to contain many complex sentences with center-embeddings such as subject-modifying object relative clauses (e.g., King & Just, 1991), while the 4-sentence condition happened to contain relatively long but non-complex sentences with a conjunction, then it is expected that the 4-sentence condition will be easier than the 3-sentence condition, despite the fact that the former is supposed to be more difficult due to the increased number of sentences to process and target words to recall. In this way, the reliability of the D&C reading span test may be unpredictably lowered due to the lack of control of the sentence types. Therefore, the sentence types need to be controlled and evenly distributed across the sentence conditions so as to prevent the reading span data from being skewed by the length or complexity of the sentences.

Finally, the read-aloud task can be intentionally performed slowly in such a way that it is easy for subjects to rehearse the target words during the reading task. Since this is an uncontrollable variable, that is, some subjects may read slowly, rehearse more and hence score better, while others may not, this again calls the validity of the test into question. The presentation of the sentence stimuli needs to be controlled in one way or another so that the duration of the target stimuli will be consistent across subjects.

The D&C reading span test suffers from these problems and hence does not seem to be a valid measure of working memory capacity. With the aim of providing a better
assessment of working memory capacity that could replace the D&C reading span test, Ariji et al. (2003) and Omaki and Ariji (2003) created a reading span test, based to some extent on Waters and Caplan (1996b), for Japanese and English. As the present study uses this reading span test, the next section illustrates its design and rationale.

2.2.2. The design of the reading span test used in the present study

The overall structure of the reading span test in Ariji et al. is similar to that of the D&C reading span test. Subjects start with a two-sentence condition, and the number of sentences increases up to five, subsequently. Each sentence condition consists of five trials, and there are 70 sentences in total (2x5 + 3x5 + 4x5 + 5x5 = 70). However, Ariji et al.'s reading span test differs from the D&C reading span test in essential ways in order to address the shortcomings that were discussed above.

Ariji et al. (2003) identified the absence of a measurement of processing performance as the most serious problem of the D&C reading span test. In order to ensure that the subjects are processing the sentences, they adopted from Waters and Caplan (1996b) an acceptability judgment task instead of a read-aloud task, and subjects are asked to judge the acceptability of each sentence after reading it. Making acceptability judgments, unlike reading aloud, requires subjects to process the sentences for meaning, since the unacceptable sentences were created by using various types of syntactic and semantic violations. In the Japanese version, the following types of unacceptable sentences were created: (a) selectional restriction errors, (b) subcategorization errors, (c) mis-assignment of thematic roles, and (d) un-licensed negative polarity items. In the English
version, following Waters and Caplan (1996b), the unacceptable sentences were created based on mis-assignment of thematic roles. The test sentences used in the Japanese and English reading span test are listed in Appendix A and Appendix B, respectively.

Moreover, although Waters and Caplan (1996b) did not take accuracy on the acceptability judgment task into consideration when determining the reading span size, Ariji et al. argued that this is crucial since some subjects could still selectively focus on the recall task while others might not, in which case the recall task score may not be an accurate representation of working memory capacity. Therefore, Ariji et al. assigned one point per sentence only when subjects performed accurately on both acceptability and recall.6 This would ensure that the scores reflect processing and storage performance, excluding cases where one of the two tasks was selectively attended to.

The third problem with the D&C reading span test was that the storage component was taxed only after the processing of a sentence was completed, which does not reflect the role of working memory in normal language comprehension. To solve this problem, Ariji et al. chose the first or second noun of the sentence as the target word so that the storage component is taxed while subjects process the rest of the sentence.

The fourth problem was concerned with the fact that uncontrolled sentence stimuli were distributed randomly across trials, and this was solved by adopting Waters and Caplan's (1996b) idea of using four types of sentences distributed evenly across trials and sentence conditions. The sentences vary along two dimensions: (a) number of propositions...

---

6 In calculating the reading span size, Ariji et al. took points from all sentences across conditions (hence the total being 70 points), unlike Daneman and Carpenter (1980) who did not calculate scores from trials that subjects did not pass. See Friedman and Miyake (2004) for a claim that the use of total scores increases the reliability.
(i.e., one or two propositions) and (b) syntactic complexity as a function of non-canonical directionality of theta role assignment and embedding of a clause inside an NP (e.g., Gibson, 1998). In the Japanese version (Ariji et al., 2003), the four types of sentences used were: (i) simple active sentences, (ii) simple sentences with the object scrambled out of the VP, (iii) active sentences with a subordinate clause, and (iv) active sentences with a subordinate clause in which the object in the subordinate clause is scrambled long-distance into the matrix clause. In the English version (Omaki & Ariji, 2003), the target sentences were: (i) subject cleft sentences, (ii) object cleft sentences, (iii) object-modifying subject relative clauses and (iv) object-modifying object relative clauses. In the Japanese and the English versions, both (i) and (ii) consist of one proposition, while (iii) and (iv) contain two propositions, and (ii) and (iv) are more complex and harder to process than (i) and (iii), respectively, due to the non-canonical directionality of theta role assignment (Caplan, Hildebrandt, & Waters, 1994). Examples of Japanese and English sentences are listed in Table 1.

Finally, the fifth problem of allowing subjects to read slowly for the purpose of rehearsing the target words was solved by timing the presentation of the stimuli on a computer. All the sentences are presented in a non-cumulative phrase-by-phrase moving-window fashion on a computer screen, and in this way all the subjects are given exactly the same amount of time to read the target sentences. Moreover, the judgment time after reading each sentence is also limited to five seconds, so that subjects cannot spend extra time rehearsing the target words during the judgment period. Controlling the presentation of stimuli in these ways can decrease the chances of extra rehearsals and thus increase the
Table 1. RST sentence types, regions and duration

**English version**

<table>
<thead>
<tr>
<th>Duration for cleft sentences</th>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Subject Cleft</td>
<td>NS*-800ms</td>
<td>NS-800ms</td>
<td>NS-1300ms</td>
</tr>
<tr>
<td>2. Object Cleft</td>
<td>L2*-1000ms</td>
<td>L2-1000ms</td>
<td>L2-2000ms</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Duration for relative clauses</th>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
<th>Region 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Subject relatives</td>
<td>NS-800ms</td>
<td>NS-800ms</td>
<td>NS-800ms</td>
<td>NS-1300ms</td>
</tr>
<tr>
<td>4. Object relatives</td>
<td>L2-1000ms</td>
<td>L2-1000ms</td>
<td>L2-1000ms</td>
<td>L2-2000ms</td>
</tr>
</tbody>
</table>

### Region 1
- **Subject Cleft**: It was the price** that pleased the customer
- **Object Cleft**: It was the scientist that the experiment excited

### Region 2
- **Region 2**: The painter praised the architect that designed the museum
- **Region 3**: The land excited the sailors that the journey exhausted

**Japanese version**

<table>
<thead>
<tr>
<th>Duration for mono-clausal sentences</th>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
<th>Region 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Simple active</td>
<td>1000ms</td>
<td>1000ms</td>
<td>1000ms</td>
<td>1000ms</td>
</tr>
<tr>
<td></td>
<td>koubou-de factory-at</td>
<td>shokunin-ga** worker-Nom</td>
<td>shatu-o shirt-Acc</td>
<td>tenuishite-iru hand sew-Prog</td>
</tr>
<tr>
<td></td>
<td>“A worker is hand sewing a shirt at the factory”</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Duration for biclausal sentences</th>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
<th>Region 4</th>
<th>Region 5</th>
<th>Region 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Simple biclausal</td>
<td>1000ms</td>
<td>1000ms</td>
<td>1000ms</td>
<td>1000ms</td>
<td>1000ms</td>
<td>1000ms</td>
</tr>
<tr>
<td></td>
<td>totuzen suddenly</td>
<td>basugaido-wa tour guide-Nom</td>
<td>untensyu-ga driver-Nom</td>
<td>mati-0 to</td>
<td>matigaeta-to mistake-Pst</td>
<td>sidasi-ta explain-Pst</td>
</tr>
<tr>
<td></td>
<td>“Suddenly, the tour guide started to explain that the driver went the wrong way.”</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Duration for biclausal sentences</th>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
<th>Region 4</th>
<th>Region 5</th>
<th>Region 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Biclausal with scrambling</td>
<td>1000ms</td>
<td>1000ms</td>
<td>1000ms</td>
<td>1000ms</td>
<td>1000ms</td>
<td>1000ms</td>
</tr>
<tr>
<td></td>
<td>kaigi-de suddenly</td>
<td>syogakukin scholarship meeting-at</td>
<td>gaikyo- wa -Acc</td>
<td>daigakuinseira -ga daigakuinseira</td>
<td>-to happyosi-ta</td>
<td>-to announce-Pst</td>
</tr>
<tr>
<td></td>
<td>“At the meeting, the dean announced that the graduate students received the scholarship.”</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*‘NS’ indicates the duration used for English native speakers and ‘L2’ indicates the duration used for L2ers.

**Although English determiners “the” as well as Japanese case markers “ga,” “o,” and “wa” were underlined, subjects were instructed not to remember them.
reliability and validity of the reading span scores as a measure of storage and processing capacity.

Furthermore, the duration of the stimuli was modulated across the English version and the Japanese version as well as the English version for native speakers and the English version for Japanese L2ers of English (see Table 1). It was reasoned that it is important to adjust the duration of stimuli to the reading time that is considered minimally necessary for a given population, since otherwise the performance on the test may merely reflect subjects not having enough time to recognize and process words presented in the various regions. The duration of stimuli was set to be slightly longer in the Japanese version than in the English version, because it was reasoned that Japanese speakers would take longer to read the Japanese orthography in which both Chinese characters and Japanese orthography are used. As for the native speaker version and the L2 version of the English reading span test, it was expected, in accordance with previous L2 sentence processing studies (e.g., Juffs, 1998), that L2ers would in general need more time than English native speakers to process the lexical items, and for this reason the duration was slightly lengthened in the L2 version. The duration of the stimuli for each sentence type in the Japanese and the English versions is summarized in Table 1 above.

In these ways, the revised reading span test in Ariji et al. (2003) and Omaki and Ariji (2003) sought to overcome the problems with the D&C reading span test. Importantly, it combines measures of the processing component and the storage component of working memory unlike previous reading span tests. It remains an empirical question whether the revised reading span test indeed serves as a more accurate measure of
working memory capacity, but the rationale provided for the revisions suggests that this reading span test developed in Ariji et al. (2003) and Omaki and Ariji (2003) seems to be a better measure of working memory capacity than the D&C reading span test.

2.3. Conclusion

This chapter reviewed the L1 and L2 psycholinguistic literature on the relation between working memory capacity and sentence processing as well as methodological issues pertaining to reading span measurements. The debate on whether working memory capacity actually influences sentence processing behavior is not resolved in the L1 psycholinguistics literature. However, the possibility that speakers who may have fewer resources might behave differently from those with larger resources in their sentence processing behavior is worth investigating in the context of L2 processing, since L2ers need to use more resources to access their (unstable) L2 knowledge (and hence are left with fewer resources) than native speakers do when accessing their (stable) L1 knowledge. The experiments in Juffs (2004, 2005) show that L2ers' sentence processing behavior does not interact with working memory capacity, but clearly more empirical studies are needed to investigate this claim by using a different construction, such as resolution of relative clause attachment ambiguity, to which we turn in Chapter 3.

This chapter also discussed the valid assessment of working memory capacity. The traditional D&C reading span test may not reflect the role of working memory in sentence comprehension. The present study will adopt the versions of the reading span test designed in Ariji et al. (2003) and Omaki and Ariji (2003), since their revisions are likely to
overcome the problems with the D&C reading span test and therefore the versions of their test seem to be a better measurement of working memory capacity.
CHAPTER 3

RELATIVE CLAUSE ATTACHMENT IN L1 AND L2 PROCESSING

3.1. Cross-linguistic investigations of relative clause attachment

Psycholinguists have extensively investigated how the parser deals with a structurally ambiguous sentence, such as in (6), where the relative clause (RC) “who was on the balcony” can modify either one of the two NPs in the complex NP (i.e., “the servant of the actress”): the non-local noun phrase (NP1) “the servant” or the local noun phrase (NP2) “the actress.”

(6) Someone shot [NP [NP1 the servant] [PP of [NP2 the actress]]] [RC who was on the balcony].

At first, this type of ambiguity was thought to be resolved by a widely attested locality principle such as Late Closure (Frazier, 1978), Recency (Gibson et al., 1996), or Right Association (Kimball, 1973), favoring local NP2 attachment in which “the actress” is interpreted to be the one on the balcony, and not “the servant of the actress.” However, it was Cuetos and Mitchell (1988) who first reported that the locality principle does not apply to the resolution of RC attachment ambiguities in on-line and off-line processing of

---

7 Technically, it is the complex NP “the servant of the actress” that is modified in non-local attachment. However, for expository purposes, we will omit the PP modifier of “the servant” (i.e., “of the actress”) and refer to non-local attachment as a modification of NP1, rather than the complex NP.
Spanish, as their Spanish subjects preferred non-local attachment. Subsequent studies on RC attachment preferences in on-line and off-line processing showed that there are considerable cross-linguistic differences. For example, a local attachment preference has been found in English (e.g., Carreiras & Clifton, 1999; Cuetos & Mitchell, 1988; but cf. Traxler, Pickering, & Clifton, 1998), Norwegian, Romanian, and Swedish (Ehrlich, Fernández, Fodor, Stenshoel, & Vinereau, 1999), whereas a non-local attachment preference has been attested in other languages like Spanish (e.g., Carreiras & Clifton, 1999; Cuetos & Mitchell, 1988), Dutch (e.g., Brysbaert & Mitchell, 1996), Greek (Papadopoulou & Clahsen, 2003), Russian (Sekerina, 1997), German (e.g., Hemforth, Konieczny, & Scheepers, 1997, 2000), and others (for a review of cross-linguistic differences in RC attachment preferences, see Fodor, 2002; Mitchell & Brysbaert, 1998).

Various accounts for these cross-linguistic differences have been proposed. Firstly, the Construal hypothesis (Frazier & Clifton, 1996) claims that their locality principle Late Closure only applies to processing of primary relations (i.e., arguments), and adjuncts such as RCs are considered non-primary relations, the processing of which is construed based on non-structural principles such as the Referentiality Principle. This principle favors attachment of an adjunct to a host that is more prominent in the discourse model, which corresponds to the argument of the matrix verb. This predicts that non-local attachment should be preferred in all languages, which contradicts the local attachment preference shown by English speakers. Frazier and Clifton (1996) base their explanation of the local

---

8 Fernández and Sainz (2004) found that Spanish speakers show a local attachment preference depending on the disambiguation cues employed in on-line experiments (e.g., number agreement vs. plausibility), but further studies are needed to identify what type of cues biases local or non-local attachment and why such biasing should occur.
attachment preference in English on the presence of two types of genitive constructions, i.e., the Saxon genitive and the Norman genitive. Contrary to the Norman genitive construction as in (6) which allows an attachment ambiguity, the Saxon genitive does not allow such ambiguity, in that an RC following a Saxon genitive construction (e.g., “the actress’s servant”) can only modify the local NP (e.g., “servant”). Given this contrast, Frazier and Clifton argued that the Gricean maxim of “Avoid Ambiguity” will cause subjects to infer that the only reason for a speaker to use an ambiguous Norman genitive construction is because s/he intended to modify the local NP in (6), since the non-local NP in (6) can be modified unambiguously by using a Saxon genitive.

This account predicts that speakers of all languages in which a Saxon and Norman genitive alternation exists should prefer local attachment. Mitchell, Brysbaert, Grondeelaers, and Swanepoel (2000) tested this prediction in Dutch and Afrikaans since these two languages have equivalents of the Norman and Saxon genitives of English. They found a non-local attachment preference in both Dutch and Afrikaans using off-line and on-line methods, and thus called into question the validity of Frazier and Clifton’s account of the local attachment preference in English. However, Heydel and Murray (2005) argued that the “functional availability” of the two genitive constructions varies across individuals, and this affects to what extent individuals prefer local attachment. They examined individual differences in their English native speakers’ functional availability (productivity, in other words) of the Saxon genitive by using a sentence completion task. The subjects were asked to read a short passage describing either a non-local or local attachment scenario, and then to paraphrase the passage in a sentence with a blank, which they would most likely fill
with either a Norman genitive or a Saxon genitive. The functional availability of the Saxon genitive can be assessed by how frequently individuals fill the blank with a Saxon genitive in the non-local attachment condition. They predicted that if the local attachment preference is driven by the functional availability of the Saxon genitive, then individuals to whom the Saxon genitive is more functionally available will be more inclined to prefer local attachment. They found in fact that the functional availability of the Saxon genitive shows a significant positive correlation with local attachment responses in an off-line questionnaire, but not in the on-line eye-tracking experiment. These results lend partial support to Frazier and Clifton's account, but a revision may be necessary since the hypothesis does not state that primary and non-primary distinctions exist only in off-line processing.

A second proposal about RC attachment preferences is the Tuning hypothesis by Mitchell and colleagues (Cuetos, Mitchell, & Corley, 1996; Mitchell, Cuetos, Corley, & Brysbaert, 1995; Desmet, Brysbaert, & De Baecke, 2002). This hypothesis states that the cross-linguistic differences in RC attachment preferences are learned purely from being exposed to instances of non-local and local attachment resolution. They predict, for example, that in sentences containing an NP1-of-NP2-RC sequence, Spanish speakers experience many more instances of non-local attachment resolution (either by syntactic or

---

9 Heydel and Murray found that in off-line processing, those subjects with a low functional availability of the Saxon genitive tended to produce more non-local attachment responses than those with a high functional availability, but in on-line processing, the effect of functional availability disappeared. This means that, somewhat puzzlingly, some subjects (at least slightly) changed their attachment preferences, depending on whether the experiment was off-line or on-line. Similar findings are reported by De Vincenzi and colleagues (e.g., De Vincenzi & Job, 1993, 1995). But see Frenck-Mestre and Pynte (2000) for a critical review of De Vincenzi's studies.
semantic cues), whereas English speakers are exposed to more instances of local attachment resolution, and that the relative frequencies of one of the two types of ambiguity resolution is directly reflected in the attachment preferences in reading comprehension. This claim has been supported by corpus studies which show correlations between attachment preferences in sentence comprehension and corpus frequencies of the occurrence of the same type of ambiguity resolution (e.g., Desmet et al., 2002). However, Gibson and colleagues (Gibson & Schütze, 1999; Gibson, Schütze, & Salomon, 1996) used a syntactic ambiguity involving conjoined NPs with three potential attachment sites and found that corpus frequencies did not match the attachment preferences in off-line survey and on-line reading time studies. This shows that tuning to statistical frequencies in the input may not necessarily correspond to preferences in ambiguity resolution (but cf. Desmet & Gibson, 2003).

The third account comes from Gibson et al. (1996) who conducted an on-line word-by-word grammaticality judgment task, testing the resolution of RC attachment ambiguity with three potential NP hosts, as in (7):

(7) The lamp near the painting of the house that was damaged in the flood...

|  | NP1 | NP2 | NP3 |
---|-----|-----|-----|

Gibson et al. found attachment preferences to NP1 and NP3 but not to NP2. The same findings were replicated in Spanish by the same task (Gibson et al., 1996) and also by eye-

---

10 Corpus studies on RC attachment preferences (e.g., Desmet et al., 2002) take into consideration only those cases that are unambiguously biased by syntactic or semantic cues, but it is not clear whether the cases where contextual information disambiguates the RC attachment should also be considered to influence the speakers' attachment preferences.
movement measures (Gibson, Pearlmutter, & Torrens, 1999), and the same pattern of data was found in Japanese as well (Miyamoto, Gibson, Pearlmutter, Aikawa, & Miyagawa, 1999). Based on these findings, Gibson et al. proposed that there are two processing principles involved in RC attachment resolution, namely, *Recency* and *Predicate Proximity*. Recency is a universal locality principle which preferentially attaches incoming material to the phrase that is currently being processed, while Predicate Proximity favors attachment to a phrase closer to the predicate phrase (e.g., VP), which leads to non-local attachment to the NP that is structurally higher. Gibson et al. hypothesized that the strength of Predicate Proximity is ranked differently across languages, and it is ranked higher than Recency in languages that show a non-local preference in RC attachment with two (or more) NP hosts. This account has been very effective in explaining the preferences observed in attachment with three NP hosts, but the model suffers from having no way of predicting the weight of Predicate Proximity in a given language (Mitchell & Brysbaert, 1998). Gibson and Pearlmutter (1998) suggested that the weight of Predicate Proximity may be larger in languages with relatively free word order, but this proposal is not supported by German data, in which researchers have found a non-local attachment preference when the genitive is marked with a genitive case marker (Hemforth, Konieczny, & Scheepers, 2000), but a local attachment preference is observed when the complex NP contains an equivalent of the English preposition “of” (Augurzky, Alter, & Pechmann, 2004).

Finally, Fodor (1998, 2002) proposed the *Implicit Prosody* hypothesis, which states that RC attachment preferences in sentence comprehension are in fact driven by the
prosodic representations projected upon processing the attachment ambiguity, even in silent reading. This Implicit Prosody account was first motivated by findings that short RCs tend to attach locally while long RCs tend to attach non-locally (Fernández, 2000; Fodor, 1998; Lovrić, Bradley, & Fodor, 2001), which correlated with the distribution of prosodic breaks in a production experiment, where forced non-local attachment sentences led to lengthening of NP2 and forced local attachment sentences led to lengthening of NP1 (Lovrić et al., 2001). Moreover, Quinn, Abdelghany, and Fodor (2000) showed that in English, French, and Arabic, the fundamental frequency (f0) peaks on NP1 (e.g., “the servant” in (6)) and on the RC (e.g., “who was on the balcony” in (6)) are higher than those on NP2 (e.g., “the actress” in (6)) when the attachment is non-local, whereas the f0 peak on NP2 was higher than the f0 peak on the RC when the attachment is local, again showing a correlation between the interpretation and prosodic representation. Jun (2003) further tested native speakers of English, Greek, Spanish, French, Farsi, Japanese, and Korean, and showed that in each language the default prosody assigned upon reading globally ambiguous sentences involving RC attachment exhibits properties that correlate with the reported attachment preferences in reading studies, lending further support to the Implicit Prosody hypothesis. However, these are all indirect sources of evidence for the Implicit Prosody hypothesis; none of these studies shows that the prosody is actually used during on-line processing of RC attachment. In fact, the default prosody collected in production experiments may be a result of attachment preferences driven by other non-prosodic factors, rather than a cause of an attachment preference. Further studies are needed to
examine whether prosody indeed has a causal effect on attachment preferences in silent reading.\footnote{Another popular account is the attachment-binding proposal of Hemforth et al. (2000), but it is not discussed here for two reasons: (a) this account makes no prediction for a language like Japanese (a language of interest due to potential L1 influence for the L2 English experiments) since relative pronouns are not used in Japanese RCs, and (b) their German data have already been challenged by Augurzky et al. (2004) who found a local attachment preference.}

In summary, the debate among the major accounts reviewed here, such as the Construal hypothesis, the Tuning hypothesis, Recency and Predicate Proximity, and the Implicit Prosody hypothesis, is far from being resolved yet, and further studies with carefully controlled methodologies are necessary to test the predictions of each account. To further complicate the issue, another variable, namely working memory capacity, has been found to influence RC attachment, and research on this will be reviewed in Section 3.3. Before we turn to this, the next section summarizes findings from L2 sentence processing research that has investigated resolution of RC attachment ambiguity in off-line and on-line experiments, with a focus on the role of L1 influence in L2 processing behavior.

3.2. Relative clause attachment in L2 sentence processing

Resolution of RC attachment ambiguity has been recently investigated in L2 processing as well. One potential difference regarding RC attachment preferences between L1 and L2 processing is that the latter poses an interesting question of whether there is L1 influence on L2 processing behavior. Given that there are robust findings in L2 research that properties of L1 grammar are transferred to L2 (e.g., Schwartz, 1998; Schwartz & Sprouse, 1996), Spanish-speaking L2ers of English, for example, may transfer their L1
processing strategies to the processing of L2 English and prefer non-local attachment, unlike native speakers of English. Past studies have shown mixed results on this issue.

Fernández (2000) investigated RC attachment preferences by English-Spanish and Spanish-English bilinguals. Her late bilinguals (i.e., L2ers whose first exposure to L2 was after 15) showed L1 transfer in off-line processing, such that English-Spanish bilinguals preferred local attachment while Spanish-English speakers preferred non-local attachment, although a clear attachment preference was found neither in late bilinguals’ on-line processing nor in early bilinguals’ (i.e., L2ers whose first exposure to L2 was before 15) off-line or on-line processing. Similarly, Felser, Roberts, Marinis, and Gross (2003) tested two groups of advanced L2ers of English whose L1 (Greek or German) shows a non-local attachment preference. Their L2ers as a group showed no clear attachment preference in the off-line and on-line experiments, while the native speaker controls showed a local attachment preference in on-line processing; they interpreted this to mean that there was no L1 transfer.

Papadopoulou and Clahsen (2003) examined RC attachment preferences off-line and on-line by advanced L2ers whose L1 (German, Russian, or Spanish) and L2 (Greek)

---

12 Fernández uses the term “bilinguals” and hence this term is kept here as well, but these bilinguals vary greatly with respect to their language learning background. Some are simultaneous bilinguals, whereas others can be classified as L2ers whose first exposure was in childhood or L2ers whose first exposure was after puberty. Even in the group that she terms “early bilinguals,” the criterion used was whether the first exposure was before or after age 15; so in the standard L2 literature, some of these “early bilinguals” would be considered adult L2ers. As for the proficiency of these bilinguals, she did not have an independent test of proficiency, but judging from the fact that these speakers have lived in the US at least for several years and that they were attending universities in New York, it can be inferred that these speakers are intermediate to advanced learners of their L2.

13 As noted above, Augurzky et al. (2004) found a local attachment preference in the equivalent of NP1-of-NP2 in German, which casts doubt on Felser et al.’s (2003) and Papadopoulou and Clahsen’s (2003) predictions regarding L1 transfer.

14 These results, as well as those considered next in Papadopoulou and Clahsen (2003), do not show that L1 transfer never occurred. Given that their subjects were quite advanced in their L2, they may have simply passed the L1 transfer phase.
are non-local attachment languages, and still found no attachment preference. They argued that the fact that speakers of an L1 with a non-local attachment preference failed to show a non-local attachment preference in the target L2 presents evidence against any kind of L1 influence.

Dussias (2003) tested advanced English-Spanish and Spanish-English L2ers living in the US, and she found that in off-line processing both groups showed a local attachment preference in their L2, although they showed no clear preference in on-line processing. For English-Spanish speakers, the findings can be accounted for by L1 influence, but in the Spanish-English group, their local attachment preference cannot come from their L1, since Spanish speakers generally prefer non-local attachment. This could indicate that these Spanish-English L2ers have acquired target-like processing strategies. One possible interpretation that Dussias (2003) discussed was that L2ers, who are computationally burdened in L2 processing despite their high proficiency, may simply resort to the computationally easy local attachment. Another possible interpretation discussed in Dussias (2003) is that, in accordance with the Tuning hypothesis discussed in the previous section, the local attachment preference observed in the two L2 groups may simply be due to the statistical frequencies of local attachment resolution in the English-speaking environment, an environment in which all of the subjects resided at the time of testing.

All of these on-line studies used a self-paced reading task and found no clear attachment preference in the group results. By contrast, Frenck-Mestre (1997, 2002) found a clear attachment preference using a more sensitive measure of initial and re-analysis processes, namely, eye-tracking measures. Looking at L2 French, she tested low-proficient
L1 English speakers, high-proficient L1 English speakers, and low-proficient L1 Spanish speakers, and found a clear transfer effect in processing French RC attachment: The low-proficient English speakers exhibited a trend for a local attachment preference, while the Spanish speakers clearly preferred non-local attachment. Furthermore, the high-proficient English group showed a non-local attachment preference, just like French native speakers did. Summarizing these results in comparison to the previous L2 self-paced reading studies, Frenck-Mestre (2005) argues that previous findings from the self-paced reading tasks may not necessarily mean that L2ers have no clear attachment preference. She argues that Felser et al.'s finding of no clear attachment preference may be due to (a) the learners going through a transitional phase, moving from consistent use of L1 processing strategies to more systematic use of L2 processing strategies, or (b) some learners possibly applying their L1 strategy while others possibly applying the target-like strategy, which results in a null overall effect. She also points out that Papadopoulou and Clahsen's (2003) reading time data are extremely slow (more than 2000ms) and therefore may reflect more than what happens in the initial stage of processing.

As the review of these previous studies shows, the question of whether L1 processing strategies influence L2 processing still requires further investigations. The present thesis investigates whether working memory capacity could be a further complicating factor in L2ers' RC attachment resolution. As discussed in Section 2.1, it is possible that L2ers in general may be rather sensitive to constraints of processing resources (cf. Dussias, 2003), and if sentence processing behavior itself can be influenced by individual differences in working memory capacity, working memory may even be an
important explanatory variable in L2ers’ RC attachment resolutions. In fact, recent studies in L1 sentence processing have shown that RC attachment does indeed seem to be influenced by individual differences in working memory capacity, and these studies are reviewed in the next section.

3.3. Relative clause attachment and working memory capacity

The assumption in Dussias (2003) that local attachment is computationally less costly is a widely-held notion. For example, in an activation-based parser as in Gibson (1998: see Section 2.1), the activation level of a word decays over time and this word needs to be reactivated to a target threshold of activation when a new incoming word is integrated with it. This reactivation consumes computational resources, while local attachment to a phrase that is currently being processed does not require any reactivation and hence is less costly. Thus, in the resolution of RC attachment ambiguities as in (6), attachment of the RC to the local NP2, the NP that is processed more recently, should be less computationally costly.

However, recent studies that investigated the relation between working memory capacity and RC attachment preferences have found a somewhat counter-intuitive pattern of results. Mendelsohn and Pearlmutter (1999) conducted an off-line questionnaire study and found that in resolving English RC attachment ambiguity, high-spans preferred local attachment, whereas low-spans preferred non-local attachment. Furthermore, Swets et al. (2004) conducted off-line experiments in English (a local attachment language) and Dutch (a non-local attachment language), and found a statistically significant negative correlation
between non-local attachment responses and reading span scores (i.e., the lower the reading span scores, the more non-local attachment responses were produced, and vice versa). Swets et al. still found differences in overall RC attachment preferences between English speakers and Dutch speakers, so a working memory constraint does not explain the general cross-linguistic differences in attachment preferences, as reviewed in Section 3.1. However, the correlation of working memory and RC attachment preferences seems robust in these two languages.

Mendelsohn and Pearlmutter suggested an explanation that appears to incorporate the idea of an activation-based model of the parser (Gibson, 1998). Due to limited computational resources, low-spans can only encode or keep active the NP1 (e.g., “the servant” in (6)) in working memory by allocating its limited resources, and thus do not encode or keep the NP2 (e.g., “the actress” in (6)) as a potential attachment host. On the other hand, high-spans have sufficient resources to keep both NP1 and NP2 active in working memory, which makes the option of local attachment readily available. They suggested another explanation, based on the Construal hypothesis, that only high-spans may have the ability to use the knowledge that an alternative genitive form (i.e., a Saxon genitive) is available. In other words, it could be the case that only high-spans have sufficient computational resources to consider the alternative Saxon genitive when exposed to the Norman genitive, and this consideration of the Saxon genitive will induce the Gricean maxim and hence lead to a local attachment preference. Alternatively, Swets et al. suggested that the Recency principle (Gibson et al., 1996) that derives local attachment is simply a more costly processing strategy for working memory than the Predicate Proximity
principle that derives non-local attachment, and hence low-spans preferentially adopt the less costly Predicate Proximity principle. However, this claim seems rather ad hoc as this view is not supported by any independent motivation.

None of the theories of RC attachment preferences reviewed in Section 3.1 seems to have clear explanations for the observed relation between working memory and RC attachment preferences. One possibility in line with the Construal hypothesis was, as described above, that the functional availability of the Saxon genitive may interact with working memory capacity. The Tuning hypothesis states that attachment preferences are solely determined by frequencies in the input, but there seems to be little reason to expect that high-spans and low-spans would be exposed to different kinds of input in a given language. The Recency and Predicate Proximity approach assumes an activation-based model of the parser as in Gibson (1998), so this approach can accommodate the suggestion made by Mendelsohn and Pearlmutter, discussed above. As for the Implicit Prosody hypothesis, A. J. Schafer (personal communication, March 10, 2005) points out that only high-spans may prefer to create a larger prosodic phrase, which may lead them to produce a prosodic phrase boundary before the RC, which then leads to a local attachment preference. In any case, all of these are speculative and hence in need of empirical testing.

\[15\] Mendelsohn and Pearlmutter and Swets et al. make the same predictions regarding RC attachment preferences, but they differ in terms of compatibility with other sentence processing phenomena. The activation-based account still retains the idea that local attachment in general is a less costly strategy, as it attributes the decreased preference for local attachment to the subjects’ inability to hold the second NP active in working memory. On the other hand, Swets et al.’s interpretation does not seem to be compatible with the robust findings in the literature that local attachment is preferred over non-local attachment in other types of attachment ambiguity (e.g., Frazier, 1978).
3.4. Conclusion

This chapter reviewed L1 and L2 sentence processing research on the resolution of RC attachment ambiguity. We first discussed theories of cross-linguistic differences in RC attachment preferences such as the Construal hypothesis (Frazier & Clifton, 1996), the Tuning hypothesis (Cuetos et al., 1996), the Recency and Predicate Proximity account (Gibson et al., 1996), and the Implicit Prosody hypothesis (Fodor, 2002), though it was pointed out that further studies are needed to test these hypotheses.

Next, the review of L2 studies on RC attachment resolution revealed mixed findings; some found clear attachment preferences in off-line processing but not in on-line processing (e.g., Dussias, 2003; Fernández, 2000), others found no clear attachment preferences in off-line or on-line group results (Felser et al., 2003; Papadopoulou & Clahsen, 2003), and yet Frenck-Mestre (1997, 2002) reported clear on-line attachment preferences. It is not clear why L2ers show such variations in their L2 RC attachment preferences, but in relation to the L1 findings that RC attachment preferences are modulated by working memory capacity (Mendelsohn & Pearlmutter, 1999; Swets et al., 2004), it was suggested that working memory may also play a role as a variable in L2 RC attachment preferences. The individual differences in working memory capacity may in fact explain the null group results, as such findings may have resulted from averaging across high-spans preferring local attachment and low-spans preferring non-local attachment.

The precise nature of why such relations hold between working memory and RC attachment preferences has yet to be made clear, but the present study starts out by
investigating whether the same pattern of results can be replicated in English for the native speaker group. Experiment 1 attempts to replicate the previous off-line processing studies, but in this study, the additional variable of *sentence complexity* was added to further test the relation between working memory and RC attachment. Experiment 2 is an on-line version of Experiment 1, so as to test whether working memory capacity interacts with attachment preferences in on-line processing as well. These two experiments with English native speakers are reported in the next chapter.

The same experimental methods are repeated as Experiments 3 and 5 for Japanese L2ers of English so as to test: (a) whether the advanced Japanese L2ers can behave similarly to English native speakers in Experiments 1 and 2, and (b) whether working memory capacity may function as an explanatory variable in Japanese L2ers’ English RC attachment resolution in off-line and on-line processing. Furthermore, a separate off-line Japanese experiment (Experiment 4) is conducted between the two English experiments so as to directly compare L2ers’ attachment preferences in L1 Japanese and L2 English on an individual level. The same group of advanced Japanese L2ers of English thus participated in these three experiments, so that potential individual factors (if any) would be held constant across the experiments. These experiments with advanced Japanese L2ers are reported in Chapter 5.
CHAPTER 4

ATTACHMENT PREFERENCE DATA FROM ENGLISH NATIVE SPEAKERS

This chapter reports on two native English experiments which were designed to investigate the relation between working memory capacity and RC attachment preferences off-line (Experiment 1) and on-line (Experiment 2). Both experiments operationalized two ways of examining the interaction of working memory capacity and attachment preference: (a) comparing the attachment preferences between low-spans and high-spans, whose reading span is determined on the basis of the reading span test (see Section 2.2.2); and (b) investigating whether the attachment preferences interact with increased sentence complexity (see below). The comparison between outcomes of Experiment 1 and Experiment 2 also sheds light on the nature of verbal working memory system, in that the Single Resource model (see Section 2.1) predicts that the interaction of working memory capacity will be observed in both off-line and on-line tasks, while the Separate-Sentence-Interpretation-Resource model (see Section 2.1) predicts discrepancies between the two. The results of these experiments will also be referred to as a baseline for testing L2ers’ processing behavior, reported in the next chapter.

4.1. Experiment 1

This off-line experiment was designed to replicate previous studies regarding working memory capacity and RC attachment. The research questions addressed in this experiment are stated in (8).
Research questions addressed in Experiment 1

a. Do English RC attachment preferences vary according to working memory capacity in off-line L1 processing?

b. Does sentence complexity also interact with English RC attachment preferences?

Before we illustrate the target stimuli of the present experiment, let us first examine the study by Eastwick and Phillips (1999) and explain why the present study manipulated sentence complexity to investigate the effect of working memory and RC attachment preferences.

Eastwick and Phillips (1999) investigated the effects of increased sentence complexity on the resolution of Main Verb/Reduced Relative (MV/RR) ambiguities (8) by low-spans and high-spans.

(9) a. Animate/Inanimate, Unembedded

The defendant/evidence (that was) recently examined by the lawyer was not very reliable.

b. Animate/Inanimate, Embedded

The judge remembered that the memo stating that [the defendant/evidence (that was) recently examined by the lawyer was not very reliable] had been stolen from the filing cabinet.
In (9a), there is a temporary ambiguity regarding whether the verb “examined” is a main verb as in “The defendant examined the evidence” or a past participle of a reduced RC, as in “The defendant (that was) examined by the lawyer.” As discussed in Section 2.1, Ferreira and Clifton (1986) showed that the parser gets garden-pathed in (9a) regardless of the animacy of the first noun, even though the inanimate noun “the evidence” is unlikely to be the agent of “examined” and thus the animacy information could potentially increase the probability of the reduced relative reading (but cf. Trueswell et al., 1994). However, Just and Carpenter (1992) showed that only high-spans can use the semantic information of animacy to disambiguate the MV/RR ambiguity and hence they can avoid being garden-pathed, while low-spans cannot use the semantic information and hence were garden-pathed (but cf. Clifton et al., 2003; Waters & Caplan, 1996a). Eastwick and Phillips (1999) lent further support to Just and Carpenter’s claim by showing that high-spans can behave like low-spans when their processing resources are extremely taxed. They embedded the MV/RR ambiguity (9a) within an embedded clause following “the memo stating…” as in (9b), thereby increasing the syntactic complexity, which in turn requires greater memory resources for processing (see below for the definition of syntactic complexity). Eastwick and Phillips found that even high-spans lost the ability to use the semantic information in the syntactically complex (9b) condition and hence were garden-pathed like low-spans, even though the same high-spans could still use the semantic information in the less complex (9a) condition. This shows, in line with Just and Carpenter’s original claim, that the availability of processing resources does affect the way the parser resolves syntactic ambiguities.
Eastwick and Phillips' findings show that complexity manipulation can indeed be used to test whether an ambiguity resolution interacts with working memory resources. The present experiment investigates the relation between working memory capacity and RC attachment ambiguity by adopting the same logic as in Eastwick and Phillips (1999), namely that an increase in sentence complexity should affect the behavior of high-spans. Consider the sentences in (10).

(10)  

a. Embedded Clause (EC) condition

The babysitter that [the sister of the schoolgirl who burned herself the other day adored] was very nice.

b. Sentential Complement (SC) condition

The babysitter said that [the sister of the schoolgirl who burned herself the other day was very nice].

In (10a), an NP containing an RC attachment ambiguity ("the sister of the schoolgirl who burned herself the other day") is embedded as an RC modifying the head of the matrix subject NP "the babysitter," in contrast to the baseline condition (10b), in which this same complex NP functions as the subject of the subordinate clause. Theories of sentence complexity predict that (10a) is much more costly than (10b). For example, Gibson's (1998, 2000) model introduced in Chapter 2 predicts that, for (10a), in the "who" region where the target RC is to attach to either NP1 or NP2, there would be a storage cost of four, that is, the number of required syntactic heads is four: (a) a verb for the matrix
subject; (b) an empty category position for the null relative operator projected at “that”; (c) a verb for the embedded clause subject; and (d) another empty category position for the relative pronoun “who.” A storage cost of four is so computationally costly that it is nearly unprocessable (e.g., Babyonyshev & Gibson, 1999; Gibson, 1998). On the other hand, much less complexity is expected in the (10b) condition, since in the same “who” region where attachment is expected to occur, the storage cost is predicted to be only two, that is, (a) a verb for the subordinate clause subject; and (b) an empty category position for the relative pronoun “who.”

Following Eastwick and Phillips (1999), it is assumed here that the increased sentence complexity in (10a) should tax working memory resources and hence affect attachment preferences, if the attachment preferences are truly influenced by working memory constraints. We can predict that there could be the following two patterns of results. First, as Eastwick and Phillips (1999) showed in the resolution of embedded MV/RR ambiguities, those high-spans, who normally behave differently from low-spans and prefer local attachment, would behave like low-spans when they are under such an extreme processing burden as in the (10a) condition, but not in the less complex (10b) condition. In this case, attachment preferences will substantially differ between the two conditions only within the high-span group but not within the low-span group. Alternatively, the complexity in the (10a) condition may equally affect high-spans and low-spans, causing both groups to produce substantially more local attachment responses than they do in the (10b) condition. In this case, both groups are expected to show substantial differences between the two conditions.
4.1.1. Method

4.1.1.1. Participants

Forty native speakers of American English in the University of Hawai‘i community participated in the study. They all had normal or corrected-to-normal vision, and they were paid $10 for their participation. However, 12 subjects were excluded from data analyses since their comprehension accuracy for unambiguous fillers \((n = 65)\) was lower than 80\% (see below), leaving a total of 28 subjects.

4.1.1.2. Materials

Two versions of 32 experimental sentences exemplified in (10), repeated below as (11), were constructed.

\[(11)\]

\[\begin{align*}
\text{a. Embedded Clause (EC) condition} \\
\text{The babysitter that [the sister of the schoolgirl who burned herself the other day adored] was very nice.}
\end{align*}\]

\[\begin{align*}
\text{b. Sentential Complement (SC) condition} \\
\text{The babysitter said that [the sister of the schoolgirl who burned herself the other day was very nice].}
\end{align*}\]

In both conditions, the first noun is the head of the matrix subject, which is always animate. In the SC condition, the first noun is followed by a verb that takes a sentential complement (e.g., “said”) and a complementizer “that,” while in the EC condition the first
noun is immediately followed by “that” which initiates an object RC modifying the first NP. Each condition consists of a clause that contains an RC attachment ambiguity, in which NP1 (e.g., “the sister”) and NP2 (e.g., “the schoolgirl”) are the same gender such that either could be modified by the RC that contains a reflexive pronoun (e.g., “herself”). In the EC condition, a verb (e.g., “adored”) follows the ambiguous region. This verb takes the sentence-initial NP (e.g., “the babysitter”) as its object, and is then followed by a predicate (e.g., “was very nice”) of the matrix subject. In the SC condition, the predicate “was very nice” is a predicate of the subject of the subordinate clause (e.g., “the sister of the school girl who burned herself the other day”).

Seventeen of the 32 experimental items contained a set of male NP1, NP2 and “himself,” and the remaining 15 items consisted of female NP1, NP2 and “herself.” The semantic relationship between the head of NP1 and NP2 was controlled to be that of family/kinship, since Gilboy, Sopena, Clifton, and Frazier (1995) showed that the family/kinship relationship is relatively neutral in terms of semantic biases on RC attachment. The length of the target RC region was restricted to four to six words (i.e., a verb, the reflexive pronoun “himself/herself,” and an adverbial/preposition phrase) in order to control for a length effect (e.g., Fodor, 2002; Jun, 2003). The list of all experimental items is given in Appendix C.

These 32 experimental sentences were combined with 75 fillers to form two lists. The fillers consisted of 10 ambiguous sentences with PP attachment ambiguity and 65 unambiguous sentences, most of which were similar to the experimental items in length and complexity. Comprehension accuracy for the unambiguous sentences was used to
assess whether subjects were paying sufficient attention to the task. As already noted, surprisingly many subjects \((n = 12)\) did not reach 80% comprehension accuracy. Some of these unambiguous fillers consisted of rather complex sentences such as doubly-nested sentences as in (12).

(12) The hockey player who the fans who the sports writer ridiculed cheered for scored a goal to win the game.

Since these fillers are fairly complex, this may be why the comprehension accuracy of the 12 subjects was lower than 80%. The experimental items were counter-balanced across two lists using a Latin-square design, such that each version of an experimental item appeared in exactly one list. In each list, the experimental items were interspersed between the fillers so that there was always at least one filler intervening between experimental items. Each subject saw one list in a different pseudo-random order.

4.1.1.3. Procedure

The experiment measured subjects' off-line reading comprehension on a computer screen, using E-Prime version 1.1. First, the target sentence appears in its entirety on the top of the screen,\(^\text{16}\) and the subjects were asked to press the spacebar when they finish reading the sentence. Upon the key press, a \(wh\)-question about the target sentence and two answer choices (e.g., “1. the sister 2. the schoolgirl) appear below the target sentence,

\(^{16}\)Some of the target sentences did not fit on a single line, but in such cases, it was ensured that at least the ambiguous region (i.e., up to the reflexive) was on the first line.
which remains on the computer screen. Subjects answered the question by pressing either 1 or 2 on the keyboard. The experimental trials were preceded by instructions and five practice trials. There was no time limit in answering questions, and the time the subjects needed to finish the task ranged from 30 to 50 minutes.

After finishing the comprehension task, they took the reading span test described in Section 2.2.2. This reading span test was also implemented on E-Prime version 1.1. Subjects were tested on all span sizes from 2 (i.e., sequences of two sentences) to 5 (i.e., sequences of five sentences). There were 5 trials of each span size, so that a total of 70 sentences were created. Nearly half of the sentences (32 out of 70) were unacceptable. The order of sentence presentation was kept constant across subjects. Each sentence was presented on a single line on a computer display. A phrase-by-phrase computer-paced (not self-paced) moving-window was used. On each trial, a fixation asterisk appeared on the screen for 500ms. Then, an experimental sentence was presented phrase-by-phrase (duration of the stimuli was given in Table 1, Section 2.2.2). After the presentation of the last word of the experimental sentence, subjects saw a "JUDGMENT" sign which indicated the beginning of the 5000ms in which the subjects had to make an acceptability judgment. After making the last judgment in each set, a "RECALL" sign appeared on the screen, indicating that they should write down on the answer sheet the target words of the trial. There was no time limit for the recall task, and subjects were asked to press Enter

17 There were two reasons why the target sentence was left on the screen with the question. First, preventing subjects from forgetting the answer and randomly guessing was important, given that the target sentences were fairly complex. Second, this makes the present study comparable to Mendelsohn and Pearlmutter (1999) and Swets et al. (2004) who used an off-line questionnaire task, in which subjects could also see the sentence and the question at the same time. However, this may have potentially reduced the effect of sentence complexity. See the results and discussion section below for more details.
when ready to move on to the next trial. Subjects were allowed to recall the target words in any order. They were told that spelling would not count. The reading span test took 15 to 20 minutes.

After the reading span test, they completed a cloze test, which took approximately 10 minutes. The cloze test, developed in Brown (1980), was administered in order to compare scores of the English native speakers with those of the L2ers who participated in Experiment 3. For this reason, the English native speakers' cloze test scores (based on any appropriate answers) are not discussed in this chapter (see Section 5.2 below), although the individual data are presented in Table 2 below. The cloze test and its answer keys are given in Appendix D.

4.1.1.4. Data analysis

In the reading span test, subjects' reading span was calculated by using a procedure outlined in Section 2.2.2. One point was given for each sentence for which the subject performed accurately in both acceptability judgment and target word recall. In other words, even if a subject correctly recalled a target word for a sentence, it was not counted as a correct answer if he or she incorrectly judged the acceptability of the sentence. Since there were 70 sentences in total, the possible maximum score was 70 if a subject accurately judged all the sentences and recalled all the target words.

In the comprehension task, a subject's response was coded as a non-local attachment response when the answer to an experimental item was “1” (i.e., NP1), while it was coded as a local attachment response when the answer was “2” (i.e., NP2). Each
subject had 16 instances of EC sentences (11a) and SC sentences (11b), and the mean non-local attachment response (in percentage) is used to analyze the subjects' RC attachment preferences. The mean non-local attachment response was submitted to a one sample t-test, separately for the EC condition and for the SC condition, to examine whether the subjects had a clear non-local or local attachment preference, that is, whether they show an attachment preference that is significantly different from chance-level performance (i.e., 50%). For the data of this experiment, analyses of individual data are also conducted in order to examine the size of the variation across the native English subjects.

Then, the subjects were divided into two groups according to their reading span scores. The cut-off score was defined as 49 (based on the mean reading span scores of the group; see below), and subjects who scored higher than 49 were labeled as high-spans, and those who scored 49 or below were labeled as low-spans. In order to see whether the span size affected RC attachment preferences, the comprehension data were submitted to a repeated measures ANOVA with span group as a between-group variable and with complexity (EC vs. SC) as a within-subjects variable. For all the statistical tests reported in this thesis, the alpha level was set at .05.

4.1.2. Results

4.1.2.1. Reading span

The mean reading span score for the 28 subjects in Experiment 1 was 49.00 ($SD = 9.58$), with a range of 28 to 66. Individual reading span scores are presented in Table 2 below as well as in Appendix E, which presents reading span scores for all the subjects investigated in this thesis.
4.1.2.2. Comprehension data

As noted above, 12 subjects were excluded from further data analyses since their comprehension accuracy for the unambiguous fillers \((n = 65)\) was lower than 80%, most likely due to the complexity of the fillers used in the present experiment. The mean comprehension accuracy for the unambiguous fillers for the remaining 28 subjects was 91.76% \((SD = 5.23)\), showing fairly high comprehension accuracy despite the complexity of some of the fillers.

For these remaining 28 subjects, mean non-local attachment responses were calculated for the EC and SC conditions. The individual data are presented in Table 2. For the EC condition, the mean non-local attachment response was 37.05% \((SD = 25.68)\), while for the SC condition, the mean non-local attachment response was 40.63% \((SD = 27.24)\). A one sample \(t\)-test shows that the number of non-local attachment responses was significantly below chance-level (i.e., the test value set at 50%) in the EC condition \([t(27) = -2.668, p = .013]\), but not in the SC condition \([t(27) = -1.821, p = .08]\). These data show that a reliable local attachment preference was observed at least in the EC condition, but only numerically so in the SC condition. This overall trend towards a local attachment preference replicates the results of previous studies that found general local attachment preferences for English (e.g., Carreiras & Clifton, 1999; Cuetos & Mitchell, 1988).

However, as the large standard deviations indicate, there was considerable variation among the subjects. In order to examine the size of the variation, individual data were categorized into three preference groups, based on their mean non-local attachment response: (a) Non-local preference (65%-100%), (b) null preference (35%-65%), and (c) local preference
Table 2. Individual data in Experiment 1

<table>
<thead>
<tr>
<th>Subject</th>
<th>Cloze (Max = 50)</th>
<th>Reading span (Max = 70)</th>
<th>EC preference*</th>
<th>SC preference*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low-spans</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>44</td>
<td>28</td>
<td>87.50% Null</td>
<td>100.00% Non-local</td>
</tr>
<tr>
<td>E2</td>
<td>42</td>
<td>31</td>
<td>50.00% Null</td>
<td>43.75% Non-local</td>
</tr>
<tr>
<td>E3</td>
<td>49</td>
<td>36</td>
<td>0.00% Local</td>
<td>12.50% Local</td>
</tr>
<tr>
<td>E4</td>
<td>45</td>
<td>37</td>
<td>43.75% Null</td>
<td>43.75% Null</td>
</tr>
<tr>
<td>E5</td>
<td>45</td>
<td>38</td>
<td>37.50% Null</td>
<td>62.50% Null</td>
</tr>
<tr>
<td>E6</td>
<td>46</td>
<td>38</td>
<td>50.00% Null</td>
<td>50.00% Null</td>
</tr>
<tr>
<td>E7</td>
<td>48</td>
<td>42</td>
<td>87.50% Non-local</td>
<td>81.25% Non-local</td>
</tr>
<tr>
<td>E8</td>
<td>40</td>
<td>43</td>
<td>62.50% Null</td>
<td>56.25% Null</td>
</tr>
<tr>
<td>E9</td>
<td>46</td>
<td>46</td>
<td>50.00% Null</td>
<td>37.50% Null</td>
</tr>
<tr>
<td>E10</td>
<td>45</td>
<td>47</td>
<td>62.50% Null</td>
<td>56.25% Null</td>
</tr>
<tr>
<td><strong>E11</strong></td>
<td><strong>49</strong></td>
<td><strong>47</strong></td>
<td>18.75% Local</td>
<td><strong>43.75% Null</strong></td>
</tr>
<tr>
<td><strong>E12</strong></td>
<td><strong>44</strong></td>
<td><strong>47</strong></td>
<td>62.50% Null</td>
<td><strong>81.25% Non-local</strong></td>
</tr>
<tr>
<td>E13</td>
<td>46</td>
<td>49</td>
<td>18.75% Local</td>
<td>25.00% Local</td>
</tr>
<tr>
<td><strong>High-spans</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E14</td>
<td>48</td>
<td>50</td>
<td>68.75% Non-local</td>
<td>81.25% Non-local</td>
</tr>
<tr>
<td>E15</td>
<td>40</td>
<td>51</td>
<td>25.00% Local</td>
<td>25.00% Local</td>
</tr>
<tr>
<td>E16</td>
<td>47</td>
<td>51</td>
<td>31.25% Local</td>
<td>18.75% Local</td>
</tr>
<tr>
<td>E17</td>
<td>48</td>
<td>53</td>
<td>6.25% Local</td>
<td>12.50% Local</td>
</tr>
<tr>
<td>E18</td>
<td>45</td>
<td>54</td>
<td>25.00% Local</td>
<td>25.00% Local</td>
</tr>
<tr>
<td>E19</td>
<td>43</td>
<td>55</td>
<td>62.50% Null</td>
<td>56.25% Null</td>
</tr>
<tr>
<td>E20</td>
<td>45</td>
<td>56</td>
<td>12.50% Local</td>
<td>12.50% Local</td>
</tr>
<tr>
<td><strong>E21</strong></td>
<td><strong>45</strong></td>
<td><strong>56</strong></td>
<td><strong>50.00% Null</strong></td>
<td><strong>81.25% Non-local</strong></td>
</tr>
<tr>
<td>E22</td>
<td>45</td>
<td>56</td>
<td>0.00% Local</td>
<td>18.75% Local</td>
</tr>
<tr>
<td>E23</td>
<td>46</td>
<td>57</td>
<td>50.00% Null</td>
<td>43.75% Null</td>
</tr>
<tr>
<td>E24</td>
<td>39</td>
<td>58</td>
<td>18.75% Local</td>
<td>6.25% Local</td>
</tr>
<tr>
<td>E25</td>
<td>46</td>
<td>58</td>
<td>0.00% Local</td>
<td>6.25% Local</td>
</tr>
<tr>
<td><strong>E26</strong></td>
<td><strong>40</strong></td>
<td><strong>58</strong></td>
<td>31.25% Local</td>
<td><strong>37.50% Null</strong></td>
</tr>
<tr>
<td>E27</td>
<td>46</td>
<td>64</td>
<td>12.50% Local</td>
<td>18.75% Local</td>
</tr>
<tr>
<td>E28</td>
<td>44</td>
<td>66</td>
<td>12.50% Local</td>
<td>0.00% Local</td>
</tr>
</tbody>
</table>

Total number of ‘non-local’ preference | 3 (2 low, 1 high) | 5 (3 low, 2 high) |
Total number of ‘local’ preference     | 14 (2 low, 12 high) | 12 (2 low, 10 high) |
Total number of ‘null’ preference      | 11 (8 low, 3 high) | 11 (8 low, 3 high) |

* Subjects were classified into three categories based on the mean non-local responses: Non-local preference (65%-100%), null preference (35%-65%), and local preference (0%-35%).
** Shading highlights the subjects whose attachment preference was different in the EC condition and the SC condition.

60
As Table 2 shows, 11 subjects fell into the null preference category for the EC condition and the SC condition, while 14 showed a local attachment preference for the EC condition, and 12 for the SC condition. This individual analysis shows that, despite the apparent local attachment preference in the group results, approximately one third of the subjects did not have a clear attachment preference.

Next, the effect of reading span on RC attachment preferences was examined. As noted above, the subjects were divided into two span groups (low vs. high) based on the cut-off score of 49, and their data are presented graphically in Figure 1, and the ANOVA table for this statistical analysis is presented in Appendix F. The mean non-local attachment response for the low-span group was 48.56% ($SD = 25.66$) for the EC condition.

![Figure 1. Mean non-local attachment response by span group in Experiment 1](image)
and 53.37% ($SD = 23.88$) for the SC condition. As for the high-span group, the mean non-local attachment response was 27.08% ($SD = 21.86$) for the EC condition and 29.58% ($SD = 25.71$) for the SC condition. A repeated measures ANOVA with span size as a between-subjects variable and complexity (EC vs. SC) as a within-subjects variable shows that there is a main effect of span size [$F (1, 26) = 6.469, MS = 7131.93, p = .017$], indicating that low-spans produced significantly more non-local attachment responses. However, this does not mean that the low-spans as a group had a non-local attachment preference. As Figure 1 shows, the mean non-local attachment response for the low-span group was around 50% for each condition, and the individual data presented in Table 2 also show that the majority of the low-spans (8 out of 14) fall into the null preference category for the EC condition and the SC condition, rather than the non-local preference category. On the other hand, Figure 1 shows that high-spans as a group exhibit a fairly clear local attachment preference, and this is also confirmed in the individual data in Table 2, in that there are 12 high-spans with a local attachment preference for the EC condition and 10 high-spans for the SC condition.

As for the effect of complexity manipulation, neither a main effect of complexity nor a significant interaction of span size and complexity was observed [$F (1, 26) = 2.382, MS = 185.95, p = .135; F (1, 26) = .238, MS = 18.54, p = .630$, respectively]. These results show that, although there was a reliable difference in mean non-local attachment response between the high-span group and the low-span group, the complexity manipulation did not affect their RC attachment responses (see Figure 1). The categorical analysis of the individual data in Table 2 also shows that only four subjects (i.e., E11, E12, E21, E26, as
highlighted by shading) fell into a different category for the EC condition and the SC condition.

4.1.3. Discussion

In summary, the results of Experiment 1 show, on the one hand, that working memory capacity is an important factor that is associated with off-line English RC attachment preferences. The present results demonstrated that high-spans showed a local attachment preference, whereas low-spans produced more non-local attachment responses, replicating the findings from Mendelsohn and Pearlmutter (1999) and Swets et al. (2004) that working memory capacity and RC attachment preferences interact. On the other hand, unlike Eastwick and Phillips (1999), the present study found no effect of sentence complexity, which seems to contradict the main effect of span size on RC attachment preferences. Let us discuss these findings in turn.

First, even though we found that low-spans and high-spans behave differently in RC attachment, it was only the high-spans who showed a clear RC attachment preference; low-spans generally showed a null preference. This is somewhat different from Mendelsohn & Pearlmutter’s findings (1999), as their low-spans had a non-local attachment preference. However, this discrepancy may be merely due to a sampling

---

18 Note that the present findings differ from those of Mendelsohn and Pearlmutter’s in that their low-spans actually had a clear non-local attachment preference, whereas our low-spans overall had a null preference. Swets et al. did not categorize their subjects into low-spans and high-spans, so it is unknown how their low-spans behaved as a group.
19 The precise details of their data were not available, but the low-span group’s mean non-local attachment response was approximately 60% and thus appears to be substantially higher than that of the low-spans in the present study.
problem; they had 53 low-spans and 20 high-spans. The 13 low-spans in our study may not accurately represent the population of low-span English native speakers.\(^{20}\)

Second, as for the lack of complexity effect, two possible explanations are in order. One possibility is that the subjects may have adopted a type of question-answering strategy, such that they first looked at the question, checking what is asked about, and then went back to the sentence to find an answer (see also footnote 17). This was possible since the target sentence remained on the screen after the question and answer choices were presented. If such a strategy was used, then the effect of sentence complexity might disappear. In the case of the experimental sentences with RC attachment ambiguity, the question can be answered by looking only at the NP1, NP2 and the RC, and therefore subjects may not have read through the whole sentence and hence avoided taxing their working memory resources.

Another possibility is that the off-line measure of RC attachment preferences may be a reflection of processes that are independent of on-line syntactic processing. In other words, answering a comprehension question after reading a sentence can be considered what Caplan and Waters (1999) called post-interpretive processing, in that it is a task that asks subjects to reflect on the meaning extracted from interpretive processing to find an answer to a particular question. Recall that in Caplan and Waters’ subdivision of verbal working memory, there exist individual differences in post-interpretive processing but not in the interpretive processing component. Given that sentence complexity defined in terms

\(^{20}\)In this respect, Swets et al. (2004) tested over 100 subjects, and hence their data might better represent the population of English native subjects. Unfortunately, however, they only conducted correlational analyses and did not examine span size as a between-subjects variable. It is thus difficult to compare their findings to our findings to shed light on the RC attachment preferences of the low-spans.
of storage and integration costs should take effect during interpretive processing rather than post-interpretive processing, sentence complexity may be handled in the same way by all subjects during interpretive processing, and only at the post-interpretive processing phase may the differences in reading span size start to emerge.

In fact, the on-line self-paced reading task employed in Experiment 2 can address both possibilities. If the null findings of sentence complexity were merely due to subjects' question-answering strategies, then the sentence complexity effect should be observed in the self-paced reading task, in which subjects are forced to read through the first to the last word while their reading times are measured. Such findings would also provide empirical support for the Single Resource model, since the model predicts that low-span and high-span differences should be observed in all kinds of linguistic processing. If the presence of the span size effect and the absence of the complexity effect were a reflection of the distinction between interpretive processing and post-interpretive processing, then the span size effect on attachment preferences should disappear in an on-line self-paced reading task, since the reading times are direct reflections of interpretive processing.

### 4.2 Experiment 2

Experiment 2 was designed to test whether working memory and RC attachment preferences interact during on-line processing. The research questions addressed in this experiment are stated in (13).
Research questions addressed in Experiment 2

a. Do working memory capacity and RC attachment preferences interact in online L1 processing?

b. If so, do both span size and complexity manipulation interact with RC attachment preferences?

c. What implications does this experiment present for the debate between the SR and the SSIR model of verbal working memory?

4.2.1. Method

4.2.1.1. Participants

Thirty-three native speakers of American English in the University of Hawaii community, who did not participate in Experiment 1, participated in this study. They all had normal or corrected-to-normal vision, and they were paid $7 for their participation. However, one subject was excluded from data analyses since comprehension accuracy was exceptionally low (the criterion will be discussed in the results section), leaving 32 subjects.

4.2.1.2. Materials

Thirty-two sentences were created by slightly changing the experimental items used in Experiment 1 while retaining all other lexical items, so that the data from the two experiments would be comparable. The experimental sentences are shown in (14).
(14) a. Embedded Clause (EC), forced local attachment

The babysitter that [the brother of the schoolgirl who burned herself the other
day adored] was very nice.

b. Embedded Clause (EC), forced non-local attachment

The babysitter that [the brother of the schoolgirl who burned himself the other
day adored] was very nice.

c. Sentential Complement (SC), forced local attachment

The babysitter said that [the brother of the schoolgirl who burned herself the
other day was very nice].

d. Sentential Complement (SC), forced non-local attachment

The babysitter said that [the brother of the schoolgirl who burned himself the
other day was very nice].

The two independent variables manipulated in (14) are (a) sentence complexity (EC vs.
SC) as in Experiment 1, and (b) attachment site of the RC (forced local vs. forced non-
local attachment). The attachment site is manipulated by using gender mismatch\textsuperscript{21} between
either of the two NPs and the reflexive “himself/herself,” which is the critical region in
these sentences: For example, if a subject has a local attachment preference, then this
subject would attach the RC to NP2 as soon as the relative pronoun “who” is encountered.
Then the subject should show significantly longer reading times in the forced non-local
attachment conditions (14b) and (14d) than in the forced local attachment conditions (14a)
and (14c), since the initial expectation that the RC will modify the local NP2 is not met in
\textsuperscript{21} In (14), NP1 is male and NP2 is female, but the gender of these two NPs alternated across items.
the forced non-local attachment conditions. Similarly, if a subject has a non-local attachment preference, then this subject should show significantly longer reading times in the forced local attachment conditions (14a) and (14c) than in the forced non-local attachment conditions (14b) and (14d).

Using a Latin-square design, these 32 sentences were counter-balanced across four lists so that a subject never saw more than one version of each experimental sentence. Within each list, these experimental sentences were interspersed between 73 fillers, which were controlled in length and complexity. The items were pseudo-randomized for presentation with each subject. All experimental items for Experiment 2 are listed in Appendix G.

4.2.1.3. Procedure

The experiment was run on E-Prime version 1.1 with a button box. The sentences were presented in a word-by-word, self-paced, non-cumulative moving-window display (Just, Carpenter, & Woolley, 1982). Subjects pressed a button to reveal each word of the sentence, and they could see only one word at a time as the preceding word disappeared upon each button pressing. Most sentences appeared on a single line from left to right, but some of the long, experimental sentences did not fit on one line. In that case, it was ensured that at least all the lexical items within the embedded clause appear on the first line, since these are the critical regions of interest. The time it takes to read one region and press a button to move on to the next word was recorded on the computer. All sentences were followed by a yes/no comprehension question, and feedback (i.e., correct or
incorrect) was occasionally given to some of the fillers so as to keep the subjects’ attention. None of the comprehension questions for the experimental items targeted the attachment ambiguity so as to avoid drawing their attention to the ambiguity. Instructions and a brief practice session with five sentences preceded the experiment, and this task took 25 to 35 minutes in total.

Next, subjects all took part in the same reading span test as Experiment 1. The materials and procedure were identical, and this test took approximately 15 minutes.

4.2.1.4. Data analysis

Reading span scores were calculated in the same way as described in Section 4.1.1.4. In data analyses for this experiment, the reading span scores were used as between-subjects variable. The subjects were divided into a low-span group and a high-span group based on the cut-off point set at 46, which was determined based on the mean score.

In the self-paced reading experiment, both comprehension accuracy and reading times for the experimental items are examined as dependent variables. Within-subject independent variables are attachment site (local vs. non-local), sentence complexity (complex EC vs. non-complex SC); span size (low-span vs. high-span) is a between-subjects independent variable. Initially, in order to examine the overall reading time pattern of English native speakers, the reading time data are examined without taking the span size into consideration. A two-way repeated measures ANOVA was conducted for each region.
of interest.\textsuperscript{22} When a statistically significant interaction was found, further analyses were conducted to investigate the nature of the interaction.

Next, the subjects were divided into two groups according to their reading span scores. As noted above, the cut-off score was defined as 46. In order to see whether the span size affected RC attachment preferences, the comprehension data were submitted to a repeated measures ANOVA with span group as a between-subjects variable, with attachment site (local vs. non-local) and complexity (EC vs. SC) as within-subjects variables.

\textbf{4.2.2. Results}

\textbf{4.2.2.1. Reading span}

Reading span scores were calculated from the sentences in which both acceptability judgment and word recall were correct. Individual reading span scores are presented in Appendix E. The mean reading span score was 46.38 (SD = 12.16), and the scores ranged from 20 to 65. In data analyses for this experiment, the reading span scores were used as a between-subjects variable. The subjects were divided into a low-span group and a high-span group based on the cut-off point set at 46. Fifteen subjects who scored below or equal to 46 were categorized as low-span, while 17 subjects who scored over 46 were categorized high-span. The mean reading span score for the low-span group was 35.87 (SD = 8.12) and the range was 20 to 46, and the mean score for the high-span group was 55.65 (SD = 5.83), and the range was 47 to 65.

\textsuperscript{22} Running ANOVAs multiple times would increase the chances of errors, so the statistical findings need to be interpreted with caution.
4.2.2.2. Comprehension questions

The overall mean comprehension accuracy calculated from all the sentences was 74.14% (SD = 8.99), ranging from 56% to 94.29%. This low comprehension accuracy was probably due to the complexity of the test sentences (both experimental items and fillers), the latter having been controlled to be similar to experimental items in terms of length and complexity. In reading time studies, it is usually the case that subjects whose overall accuracy is lower than 70% or 80% are excluded from further analyses since those subjects may not have paid sufficient attention. In the present study, however, it appears that the low accuracy did not necessarily result from low attention span. For this reason, a somewhat looser cut-off score was used in the current study, and subjects whose mean overall accuracy was lower than 2 standard deviations from the grand mean were excluded from further analyses. One subject fell into this category and was excluded from further analyses, thus leaving 32 subjects. The mean overall comprehension accuracy for the remaining subjects was 74.7% (SD = 8.52).

Next, the mean comprehension accuracy for the four target conditions was computed: 71.88% (SD = 17.10) for the EC/forced local attachment condition (14a), 62.11% (SD = 22.78) for the EC/forced non-local attachment condition (14b), 67.97% (SD = 21.33) for the SC/forced local attachment condition (14c), and 75.78% (SD = 17.37) for the SC/forced non-local attachment condition (14d). These data were divided into high-span and low-span results, and these are graphically presented in Figure 2 below. A repeated measures ANOVA with complexity (EC vs. SC) and attachment (forced local vs. forced non-local) as within-subjects variables and reading span (low-span vs. high-span) as
a between-subjects variable was conducted on the mean comprehension accuracy. First, no main effect was observed for complexity \(F(1, 30) = 2.823, MS = 8.789, p = .103\) or attachment \(F(1, 30) = .182, MS = 3.175, p = .673\), but a main effect for reading span was observed \(F(1, 30) = 7.811, MS = .505, p = .009\) as the high-spans showed a significantly higher mean accuracy than low-spans across the four conditions.

Next, no significant interaction was observed for complexity and reading span \(F(1, 30) = 3.705, MS = .1115, p = .064\) or for attachment and reading span \(F(1, 30) = .052, MS = 9.000, p = .822\), whereas a main effect was observed for the interaction of attachment and complexity \(F(1, 30) = 13.569, MS = .278, p = .001\), showing that the effect of attachment was dependent on the complexity level. Moreover, a main effect of the interaction of complexity, attachment, and reading span was observed \(F(1, 30) = 9.631, MS = .197, p = .004\), showing that the interaction of attachment and complexity was also
dependent on the reading span size. ANOVA tables for these statistical analyses are presented in Appendix H.

Furthermore, the comprehension accuracy data were analyzed separately for the low-span group and the high-span group. Within the low-span group, there was a main effect of complexity \( [F(1, 14) = 9.701, MS = .190, p = .008] \) as well as a significant interaction of complexity and attachment \( [F(1, 14) = 22.779, MS = .444, p = .000] \). Within the EC condition for the low-span group, a main effect of attachment was observed \( [F(1, 14) = 16.588, MS = .234, p = .001] \), showing that the mean accuracy for the EC/forced local attachment condition (14a) was significantly higher than for the EC/forced non-local attachment condition (14b). However, within the SC conditions for the low-span group, a main effect of attachment was observed in the opposite direction \( [F(1, 14) = 7.71, MS = .210, p = .015] \), showing that the mean accuracy for the SC/forced local attachment condition (14c) was significantly lower than for the SC/forced non-local attachment condition (14d). In contrast to the low-span group, the high-span group showed no main effect of complexity or attachment, and showed no significant interaction of complexity and attachment.

The comprehension data from the low-span group, as shown in Figure 2, are somewhat puzzling. The EC conditions show that the local attachment condition was easier than the non-local attachment condition, but this relation is reversed in the SC conditions, where the non-local attachment condition appears to be easier than the local attachment condition. Possible ways to account for the comprehension data together with the reading time data will be examined in the discussion section.
4.2.2.3. Reading times

Reading time data are often excluded from analyses when they come from sentences for which comprehension questions were not correctly answered, since this may indicate the subjects were not reading the sentence seriously. In the present study, however, reading times for all experimental items were analyzed regardless of how the comprehension questions were answered, since the accuracy rate was fairly low across subjects, and this may simply have been due to the complexity of the experimental items and not to low attention. In order to exclude outliers, reading times beyond 3 standard deviations from the mean for a given condition and position were trimmed to the value of 3 standard deviations, affecting 1.4% of the data. Reading times for the initial non-crucial regions before “that” (e.g., “the babysitter” in the EC condition and “the babysitter said” in the SC condition) and the sentence final region after “was” (e.g., “very nice”) were not analyzed here. Also, reading times for the adverbial region, in which the number of words differed across sentences, are a sum of average reading times for each word in the relevant region. Regions of the target sentence that were statistically analyzed are shown in Table 3.

Table 3. Regions of interest in the target sentences in Experiment 2

<table>
<thead>
<tr>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
<th>Region 4</th>
<th>Region 5</th>
<th>Region 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>that</td>
<td>the</td>
<td>brother</td>
<td>of</td>
<td>the</td>
<td>schoolgirl</td>
</tr>
<tr>
<td>Region 7</td>
<td>Region 8</td>
<td>Region 9</td>
<td>Region 10</td>
<td>Region 11*</td>
<td>Region 12</td>
</tr>
<tr>
<td>who</td>
<td>burned</td>
<td>herself/himself</td>
<td>the other day</td>
<td>adored</td>
<td>was</td>
</tr>
</tbody>
</table>

*Region 11 is missing in the SC condition, but “was” in the SC condition is still referred to as Region 12 for the sake of convenience.
Figure 3. Mean reading times in crucial regions in Experiment 2
Overall reading time data (i.e., collapsing the two span groups) from the crucial regions are presented in Figure 3. Differences in reading times were expected to show up after Region 1, since the storage costs (Gibson, 1998, 2000) in the EC and the SC conditions start to diverge at this point. The reaction time data in these regions were submitted to repeated measures ANOVAs with complexity (EC vs. SC) and attachment (local vs. non-local) as within-subjects independent variables. No main effect or significant interaction was observed in Regions 1, 2, 3, and 5. The main effect of complexity was observed for Region 4 “of” \( [F (1, 31) = 5.385, MS = 158600.02, p = .027] \), Region 6 “schoolgirl” \( [F (1, 31) = 12.381, MS = 3763828.15, p = .001] \), Region 7 “who” \( [F (1, 31) = 8.866, MS = 1146017.16, p = .006] \), and up to the verb of the RC Region 8 “burned” \( [F (1, 31) = 7.654, MS = 289044.62, p = .009] \), whereas no main effect of attachment or significant complexity-by-attachment interaction was observed in any of these regions. These results show that the BC condition, which was assumed to be complex based on Gibson’s model, is indeed perceived by the subjects as more difficult and causes slower reading times than the SC condition.

Next, in the critical Region 9 “himself/herself” where the attachment preferences should be reflected, the following results obtain (the ANOVA table for the following analyses is presented in Appendix H). In contrast to the preceding regions, there was no main effect of complexity in this region \( [F (1, 31) = .104, MS = 9698.06 , p = .749] \), presumably because the storage cost differences between the two conditions become smaller as the expected syntactic heads are integrated into the sentence currently being processed. There was also no significant complexity-by-attachment interaction \( [F (1, 31) \)
On the other hand, a main effect of attachment was observed in this region \([F(1, 31) = 4.403, MS = 179793.50, p = .044]\), showing that the reflexive region of the local attachment condition was processed significantly faster than that of the non-local attachment condition when the EC and the SC conditions are collapsed. However, this main effect was lost when the analysis was conducted within the EC and SC conditions separately [for the EC condition, \(F(1, 31) = .411, MS = 24307.90, p = .526\); for the SC condition, \(F(1, 31) = 2.417, MS = 196910.352, p = .130\)]. No main effect or significant interaction was observed in Region 10 “the other day” and Region 11 “adored,” showing that there was no spill-over effect from Region 9.

In Region 12 “was,” integration cost differences were expected to emerge between the two conditions (Gibson, 1998, 2000), since the integration of “was” to the matrix subject involves more intervening discourse referents than the integration of “was” to the embedded clause subject in the SC condition. As predicted, a main effect of complexity was observed in Region 12 \([F(1, 31) = 18.401, MS = 1206197.70, p = .000]\), with the EC condition producing significantly longer reading times than the SC condition (the ANOVA table for this statistical analysis is given in Appendix H). Neither a main effect of attachment nor a significant complexity-by-attachment interaction was observed in this region. Taken together with the storage cost effects observed in Regions 4, 6, 7, and 8, it is clear that the EC condition caused more processing difficulties than the SC condition.

Next, in order to see if span size difference influenced the reading time data, the reading time data were divided into the low-span and high-span groups and submitted to a separate repeated measures ANOVA. See Figure 4 for the collapsed reading time data for
Figure 4. Mean reading times by span group in crucial regions in Experiment 2
the low-span group and high-span group in the regions of interest. The effect of reading span as a between-subjects variable was observed only in two regions: Region 1 and Region 12. In Region 1, a significant interaction of complexity, attachment and reading span was observed \( F(1, 30) = 4.958, MS = 87465.50, p = .034 \), while there was no other main effect in this region. One could make sense of the effect of complexity itself, since the word "that" in Region 1 is quite different in the EC and SC conditions. In the EC condition, it is a relative pronoun with which an empty category in the subject gap position is potentially processed together, whereas in the SC condition, it is only a complementizer; these differences may lead to a main effect of complexity. Also, following King and Just (1991), one could hypothesize that the complexity of Region 1 may burden low-spans more than high-spans, and for this reason, complexity-by-span interaction is also plausible. However, none of these factors (i.e., complexity effect and complexity-by-span interaction) showed an effect, and strangely, the attachment manipulation interacted with complexity and span size. As the only difference between the two attachment conditions is the choice of the reflexive, the local attachment and the non-local attachment conditions exhibited no difference up to this point in the sentence. Moreover, there seems to be no way for subjects to predict the presence of such manipulation at Region 1. It is unclear why this significant interaction was found at all in this region, but it cannot be due to the manipulation of the variables of interest in the present study, so it is going to be ignored in the rest of the thesis.

More importantly, a main effect of complexity (i.e., an effect of integration costs) as well as a significant complexity-by-reading span interaction was found in Region 12 \( F \)
(1, 30) = 21.219, \( MS = 1110114.37, p = .000; F (1, 30) = 8.840, \( MS = 462476.464, p = .006, \) respectively]. Next, the reading time data were analyzed within each span group for this region, and the by-group distribution of reading time data in Region 12 is presented in Figure 5. A main effect of complexity was observed in the low-span group \( [F (1, 14) = 6.365, \( MS = 65670.417, p = .024] \) as well as the high-span group \( [F (1, 16) = 17.997, \( MS = 1603003.66, p = .001] \). Taken together, the interaction of complexity and reading span reported above indicates that the complexity effect was significantly larger in the high-span group than in the low-span group, despite the fact that the effect of complexity is significant in each group. However, given that high-spans are generally slower than low-spans even in the SC condition, it may be the case that the slow-down in the EC condition was consequently also proportionally larger for the high-span group. In other words, the apparent larger effect in the high-spans may not necessarily indicate that their parsers are qualitatively different from those of the low-spans.

![](image)

Figure 5. Mean reading times (in ms) by span group in Region 12 in Experiment 2
In regions other than Regions 1 and 12, no effect of span size was observed. However, it is worth noting that high-spans are generally slower than low-spans (see Regions 5, 7, 8, 9, 10, 11, and 12, in Figure 4). This is somewhat puzzling, given that low-spans are more likely to experience larger processing difficulties in such complex sentences as the ones in the present experiment (cf. King & Just, 1991). We will return to this point in the discussion section.

Finally, reading time data in Region 9 ("himself/herself") by span group are presented in Figure 6. Although this region showed no statistically significant findings, the null findings may be due to a rather small sample size of the experiment, and the numerical differences may be informative. At least numerically, low-spans appear to have a local

![Figure 6. Mean reading times (in ms) by span group in Region 9 in Experiment 2](image)
attachment preference in the EC condition, whereas no clear difference is observable in the SC condition. On the other hand, high-spans show no clear difference in the EC condition but appear to prefer local attachment in the SC condition.

4.2.3. Discussion

Let us first summarize the findings of the experiment. First, the comprehension data for the four target conditions showed no significant differences among the high-spans, but low-spans showed a rather complicated pattern of results when examined within the EC and SC conditions. For the low-spans, the mean accuracy of the EC/local attachment condition was substantially higher than that of the EC/non-local attachment condition, showing that the local attachment condition was processed more easily; however, within the SC condition, the pattern was reversed, and the SC/non-local attachment condition had a substantially higher mean accuracy than the SC/local attachment condition, showing the advantage of non-local attachment. Second, the reading time data showed three things: (a) the reading times were faster across several regions in the SC condition than in the EC condition, due to the effects of storage costs (Regions 4, 6, 7, and 8) and integration costs (Region 12), supporting the assumption that the EC condition caused more processing difficulties; (b) Region 9 ("himself/herself") showed an overall advantage (i.e., faster reading times) of local attachment when the EC and the SC conditions are collapsed, although the statistical significance was lost when examined separately within the EC and SC conditions; and (c) reading span as a between-subjects variable interacted with complexity in Region 12, where the integration cost effect was observed, and this
interaction appeared to suggest that the magnitude of the integration costs was larger in the high-span group than in the low-span group. However, the main effect of complexity observed within each group, and given the fact that high-spans had generally longer reading times than low-spans did, it was suggested that the observed interaction was due not to a difference in the parsers of the low-spans and high-spans but to a proportional reading time increase in the high-span group.

The comprehension data from the low-span group show a sharp contrast with the memory resource account suggested in Mendelsohn and Pearlmutter (1999), namely that taxation of memory resources may lead the parser to keep active only the non-local NP in working memory. The findings in the present study contradicts this account since the more complex conditions showed a local attachment advantage and the less complex conditions showed a non-local attachment advantage. In fact, this pattern of results intuitively makes more sense, given that local integration is assumed to be computationally less demanding.

Reading time data in Region 9, however, did not show a clear effect of span size. It appeared, at least numerically, that the low-spans had a local attachment advantage in the EC condition, whereas the high-spans had a local attachment advantage in the SC condition (see Figure 6). It could be the case that, regardless of the complexity manipulation, both low-spans and high-spans actually have a local attachment preference in on-line processing, but this may have been masked by the small sample size of the present study. If this were truly the case, then it would still leave unanswered why low-spans differ in their attachment preferences in off-line and on-line processing. One possible explanation comes from the suggestion made by Mendelsohn and Pearlmutter (1999),
which is that only high-spans may have the ability to consider the Saxon genitive while processing a Norman genitive construction (see Sections 3.1 and 3.3); to consider the possibility of Saxon genitives upon encountering Norman genitives requires an additional linguistic process, which may become possible only if there are sufficient memory resources. If we further assume the Separate-Sentence-Interpretation-Resources (SSIR) model (Caplan & Waters, 1999), which, recall, posits individual differences only in post-interpretive processing, then we could make sense of why we observe different attachment preferences in off-line and on-line processing. In fact, Heydel and Murray (2005) found a correlation between the functional availability of the Saxon genitive and a local attachment preference only in off-line processing, but not in on-line processing. This finding also patterns with the off-line versus on-line findings of the current study, though one needs to empirically test whether the functional availability of the Saxon genitive indeed correlates with working memory capacity.

Finally, let us turn to the data regarding span size and sentence complexity and their implications for theories of verbal working memory capacity. With respect to the reading time data, the present study found that, in general, high-spans were slower than low-spans in reading the experimental items, and this was observed in the regions where low-spans were expected to slow down more than high-spans. This seems incompatible with one of the predictions of the Single Resource (SR) model, as they predict that in

---

23 It is not clear why high-spans were slower than low-spans. One possible explanation for the high-spans slower reading times is that there is a speed-accuracy trade-off (A. J. Schafer, personal communication, June 13, 2005). The high-spans may be more concerned than low-spans about the accuracy in answering the end-of-the-sentence comprehension questions than low-spans, and this might have caused them to read more slowly and carefully. In fact, as Figure 2 showed, the high-spans were more accurate in answering the comprehension questions.
reading complex regions, the low-spans should experience substantially more processing difficulties. Note, however, that MacDonald et al. (1992) suggested that high-spans may be slower than low-spans when high-spans are computing multiple analyses while low-spans employ only one of the possible analyses. It is certainly possible that the high-spans in our study were also computing multiple analyses, but two reasons cast doubt on this possibility. First, in Region 9, the high-spans showed a numerical advantage for local attachment in the SC condition (see Figure 6), which is unexpected under the multiple analysis account: If they were really computing two analyses, then all they need to do at the point of disambiguation is to pick one of the analyses, so one should not find an advantage for one of the analyses. Second, the reading times are slower in the regions after the ambiguity is disambiguated as well (i.e., Regions 10, 11, and 12), which is unexpected under MacDonald et al.'s account, since they would predict that the multiple analyses are computed only in the ambiguous regions (cf. Waters & Caplan, 1996a). In this way, none of the predictions of the Single Resource model was borne out.

Next, recall that the SR model predicts a significantly larger processing difficulty for the low-spans in a very complex region (King & Just, 1991). There are mixed results for this prediction. First, span size showed no main effect in the regions where storage cost effect was expected (Regions 1 through 10), and the fact that low-spans and high-spans did not show differences seem to support the SSIR model. However, reading time in Region 12 ("was"), where integration cost was expected, showed a significant complexity-by-span interaction, suggesting that in this region the high-spans had a larger integration cost effect than the low-spans did. If this was the case, then this finding lends support to the SR model
and casts doubt on the SSIR model, since the former predicts a substantial difference in processing behavior between low-spans and high-spans whereas the latter model does not predict such differences. However, as noted above, the high-spans were generally slower than low-spans, and this trend suggests that the significant interaction merely reflects a proportional reading time increase in the high-spans. In fact, a main effect of complexity (i.e., integration cost effect) was observed in separate ANOVAs for the low-spans and high-spans, which suggests that their parsers similarly experienced an integration cost effect.

In contrast to the reading time data, recall now that the low-spans and the high-spans did show differences in the comprehension question data (see Figure 2), in that high-spans exhibiting no difference across conditions while low-spans favored local attachment in the EC condition but favored non-local attachment in the SC condition. This apparent difference between the two groups may seem to support the SR model. However, notice that answering comprehension questions is a type of post-interpretive processing, in that one needs to reflect on the meaning extracted from the interpretive processing and judge if their interpretation of the sentence matches the content of the question. In this sense, the observed difference in the comprehension question data between low-spans and high-spans is compatible with the SSIR model as well, since it predicts that the span difference exists only in post-interpretive processing but not in interpretive processing.

Taken together, it is likely that the observed significant interaction does not constitute counter-evidence against the SSIR model. In summary, though the significant

---

24 See De Vincenzi & Job (1993, 1995) for similar findings in Italian that reading time data and comprehension question accuracy point to a difference attachment preference.
interaction observed in the reading time for Region 12 ("was") must be interpreted with caution, the overall findings lend support to the SSIR model.

4.3 Summary

Off-line data in Experiment 1 replicated the previous findings that high-spans produce more local attachment responses than low-spans, but the effect of complexity on attachment preferences was not observed. Based on recent findings by Heydel and Murray (2005), it was suggested that in off-line processing, high-spans have sufficient resources to entertain a higher functional availability of the Saxon genitive and hence prefer local attachment, while low-spans do not have sufficient resources to consider the alternative genitive form and thus only prefer non-local attachment due to the Referentiality Principle (i.e., attach to a host that is more prominent in the discourse model, i.e., non-local NP).

In Experiment 2, it was found that neither span size nor the complexity seemed to have an effect on on-line RC attachment, although the overall data suggested that local attachment had an advantage. It was also shown that none of the predictions of the SR model was borne out by the data, thus lending indirect support for the SSIR model.

We have seen in this chapter that, at least in off-line processing, RC attachment can be used to investigate the role of working memory capacity in sentence processing. We turn now to non-native speakers, who are hypothesized to be subject to memory resource constraints. The next chapter presents experiments with advanced Japanese L2ers of English, who participated in the same tasks as in Experiments 1 and 2 reported in this
chapter. The native speakers’ results obtained in Experiments 1 and 2 will be used as a baseline for comparison with the L2ers.
CHAPTER 5
ATTACHMENT PREFERENCE DATA FROM JAPANESE SPEAKERS

This chapter presents three experiments with Japanese L2ers of English to investigate the role of working memory capacity and L1 influence in L2 sentence processing. Experiments 1 and 2 reported in the previous chapter are replicated with the Japanese L2ers as Experiments 3 and 5 so as to compare processing behavior between the two groups. As a further test of L1 transfer, a Japanese version of Experiment 1/3 (Experiment 4) was also conducted, and the comparison of results from these two experiments at an individual level is intended to shed light on the nature of L1 influence on L2 processing.

This chapter starts out by reviewing L1 Japanese RC attachment studies, as these will be referred to as the bases for predictions of L1 transfer. The next section presents the language proficiency background of the L2ers, who participated in all three experiments in the present chapter. The rest of the chapter presents and discusses results from the three experiments in turn.

5.1. Japanese relative clause attachment

As was seen in the review of the previous work on RC attachment studies presented in Chapter 3, most of the existing data come from head-initial languages, in which the complex NP (e.g. ‘NP1 of NP2’ in English) precedes the RC; hence in these languages, the two potential attachment sites have already been processed before the point of attachment.
ambiguity. However, not many studies have been conducted in a head-final language like Japanese, in which the RC precedes the complex NP; hence the potential attachment sites become available only after the RC is processed (15).

(15)  Taro-ga [[RC barukoni-ni ita] [NP [NP2 joyuu-no ] [NP1 mesitukai-o]]] ut-ta
Taro-Nom balcony-at was actress-Gen servant-Acc shoot-Pst
“Taro shot the servant of the actress who was on the balcony.”

The study by Kamide and Mitchell (1997) was the first to investigate RC attachment preferences in Japanese. They translated into Japanese the materials used in Cuetos and Mitchell (1988) and conducted an off-line questionnaire study. The results show that there was a significant non-local attachment preference (i.e., attachment to “mesitukai” (‘servant’) in (15)). However, in their on-line self-paced reading task, the forced local attachment condition produced significantly faster reading times than the forced non-local attachment condition, which shows that Japanese speakers have a local attachment preference in on-line processing, although this initial attachment seems to be reversed in off-line processing.

The discrepancy between the off-line and on-line results seems somewhat puzzling and it is not entirely clear why this is the case, but one possibility is that the local attachment preference in the on-line task is due to the head-final property of Japanese. As seen in (15), at the point where the local NP2 “joyuu-no” (‘actress-Gen’) is processed, NP1 is not processed yet, so the parser either attaches the RC immediately to NP2 or waits until NP1 becomes available. In fact, this ‘prediction’ that the parser waits because NP1 is yet to

25 Note that in Japanese, the local NP (NP2), marked by the genitive case marker “-no,” precedes the non-local NP (NP1).
be processed is made possible by the genitive case marker, since it unambiguously signals
that there must be an NP following it. In order to test whether this latter possibility holds
true, Miyamoto, Nakamura, and Takahashi (2004) conducted an on-line study and
presented the whole complex NP (i.e., NP2-Gen NP1) together in one region, so that the
preview of the genitive case marker as well as NPI may allow them to attach non-locally
even in the early stages of processing. However, the local attachment preference in their
on-line task was still found with this segmentation, lending support to Kamide and
Mitchell’s claim that the on-line and off-line preferences do diverge in Japanese.26

Given these findings, it can be predicted that, if Japanese L2ers of English transfer
their L1 processing strategies, their transfer may be observed differently in off-line and on-
line studies. B. D. Schwartz (personal communication, February 15, 2003) points out,
however, that the Japanese and English RC attachment constructions may not be truly
equivalent due to the difference in the linear order of the complex NP and the RC, and this
may prevent Japanese L2ers from transferring their L1 attachment preferences. This could
be particularly true in on-line processing, since the attachment hosts in Japanese are not
processed before the RC, as reviewed above. In this context, upon encountering an English
RC attachment construction in which the complex NP precedes an RC, Japanese L2ers
may simply consider this to be a new construction that does not exist in their L1. However,
at least in off-line processing in which multiple possibilities can be evaluated, English and
Japanese RC attachment ambiguity presents a similar context in which two attachment
sites are available before the final attachment decision is made. For this reason, the present

26 See Lee and Kweon (2004) for similar findings in Korean, of which the structure of RCs is very similar to
that of Japanese.
study addresses the L2ers’ RC attachment preferences in both English and Japanese off-line experiments, so that their L1 and L2 off-line processing behavior can be directly compared (Experiments 3 and 4, respectively).

With respect to working memory and RC attachment, no previous studies have investigated this in the context of a pre-nominal RC language like Japanese. If the relation between working memory and RC attachment is universal, then all other things being equal, the span size effect found in Experiment 1 should be replicated in Japanese as well, such that Japanese high-spans produce more local attachment responses than low-spans, at least in off-line processing when the two NP hosts are available before the final attachment decision is made. However, if such relations are dependent on some specific language factors, then what was observed in English (and Dutch) may not necessarily be replicated in Japanese. In fact, Chapter 4 discussed the possibility that the functional availability of the Saxon genitive construction may coincide with working memory resources. If this is the case, then it is predicted that working memory capacity would not interact with Japanese RC attachment preferences, since Japanese has only one genitive construction, employing a genitive case marker. Thus, Japanese data from Experiment 4 can shed further light on the nature of working memory capacity and RC attachment preferences in native speakers.

5.2. Profiles of Japanese participants

Twenty-four Japanese L2ers of English studying at the University of Hawaii were recruited to participate in the present study. Twenty-one were graduate students and the
remaining three were undergraduate students, and they were all studying in English at the university. None of them was taking any ESL courses at the time of the experiment, as they were either exempted from ESL requirements for their high TOEFL scores (over 600) or had gone through such ESL courses a long time before the present study was conducted. Table 4 below summarizes the mean, standard deviation and range of their age at the time of testing, age of first exposure to English, length of residence in English-speaking environments (e.g., US, UK, Canada, Australia) as well as their scores in an English cloze test developed by J.D. Brown at the University of Hawaii (Brown, 1980; see Appendix D). All the information was gathered as a part of Experiment 3 to be reported below, and the background questionnaire is given in Appendix I, and more detailed individual data are presented in Appendix J.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at the time of the experiment</td>
<td>30.00</td>
<td>6.01</td>
<td>20 to 44</td>
</tr>
<tr>
<td>Age of first exposure to English</td>
<td>11.17</td>
<td>1.81</td>
<td>6 to 13</td>
</tr>
<tr>
<td>Length of residence in English-speaking environment</td>
<td>5.56</td>
<td>4.30</td>
<td>1 to 15</td>
</tr>
<tr>
<td>English cloze test (max = 50)</td>
<td>40.08</td>
<td>4.25</td>
<td>34 to 47</td>
</tr>
</tbody>
</table>

As Table 4 shows, the mean score of the English cloze test was 40.08 (SD = 4.25), which is over 80% of the total score of 50. Their cloze test scores were compared to that of 40 native speakers (including those 12 whose comprehension data were not analyzed) collected in Experiment 1. The native speakers’ mean cloze score was 43.65 (SD = 3.75),
and a one-way ANOVA with L1 as a between-subjects variable shows that the native speakers’ cloze test scores were significantly higher than that of the L2ers $[F(1, 62) = 12.261, MS = 190.82, p = .001]$. However, as will become clear later, the L2ers’ comprehension accuracy in Experiment 3 was quite high, and the group mean of the L2ers’ comprehension accuracy was above 90% (see below), whereas 12 out of 40 native speakers did not achieve 80% comprehension accuracy in Experiment 1. This indicates that these Japanese learners are indeed quite advanced, despite the significant difference observed in the comparison of the cloze test scores. For this reason, all the L2 subjects are going to be considered for data analyses.

All of the L2 subjects participated in the three experiments, so that the results from each experiment would remain comparable. Note that testing the same group of subjects on similar target items is usually avoided in L1 sentence processing research, since having been exposed to similar items in a prior experiment may cause priming effects, i.e., the sentences that the subjects read may remain in their short-term memory and facilitate the processing of those sentences (e.g., Frazier, Taft, Roeper, & Clifton, 1984; Pickering & Traxler, 2005), and Experiments 1 and 2 reported in this thesis tested a different group of native subjects exactly for this reason. For native speakers, testing a different group of subjects and comparing the results across the two experiments is in fact possible on the assumption that native speakers of a particular language are (in large part) homogeneous, at least with respect to their linguistic knowledge. However, it is rather questionable that this assumption holds for L2ers, whose L2 acquisition is influenced by various factors, such as length of exposure, type of exposure, length of residence, proficiency, individual
differences in grammatical development, and so on. If we test different L2ers across experiments, a more serious concern is that the individual differences observed in L2 acquisition may invalidate cross-experiment comparisons. For this reason, all participants were requested to take part in three experiments, but in order to get around the potential priming effect, Experiment 3 was separated from Experiments 4 and 5 at least by an interval of 2.5 months, while there was an interval of five to seven days between Experiments 4 and 5. Experiment 4 was in Japanese while Experiment 5 was English, so it was reasoned that not having an interval between these two experiments would not lead to priming effects.

5.3. Experiment 3

Experiment 3 adopted exactly the same method as Experiment 1 and tested Japanese L2ers of English, in order to address the following research questions.

(16) Research questions addressed in Experiment 3

a. Is there L1 influence on L2ers’ RC attachment preferences, or do L2ers show target-like attachment preferences?

b. Do high-span L2ers and low-span L2ers show different attachment preferences?

c. Does structural complexity affect the L2ers’ attachment preferences?
The subjects in the present study speak Japanese as their L1, which is known to show a non-local RC attachment preference in off-line processing (Kamide & Mitchell, 1997). This contrasts with the RC attachment preference attested in English, which is generally low. If there is L1 influence on the L2ers' RC attachment preferences in their L2, it is expected that they will produce at least more non-local attachment responses than the English native speakers did in Experiment 1. Next, the present study examines whether working memory capacity influences L2ers' RC attachment preferences. If working memory has an influence on L2ers' processing behavior, then it is expected that the L2ers in the present study would also show differences in RC attachment preferences according to their working memory capacity, as observed for native speakers in Experiment 1. The structural complexity manipulation (viz. EC vs. SC) may also shed light on the relation between memory resources and attachment preferences, but as the previous chapter pointed out, this effect may not be observable due to the experimental design that allowed a question-answering strategy which eliminates the effect of complexity.

5.3.1. Method

5.3.1.1. Participants

Twenty-four advanced Japanese L2ers of English described in Section 5.2 participated in this study. They were paid $10 in compensation.
5.3.1.2. Materials

The materials used in Experiment 3 were identical to those used in Experiment 1. An example of the 32 experimental sentences was given in (10), which is repeated below as (17). These 32 sentences were counter-balanced across two lists and interspersed between fillers, which were also of the same type. The experimental sentences were pseudo-randomized for each subject.

(17) a. Embedded Clause (EC) condition

The babysitter that [the sister of the schoolgirl who burned herself the other day adored] was very nice.

b. Sentential Complement (SC) condition

The babysitter said that [the sister of the schoolgirl who burned herself the other day was very nice].

5.3.1.3. Procedure

The procedure was also identical to that of Experiment 1, except for the minor changes in the reading span test. As was described in Chapter 2, the reading span test for L2ers had the same design as the one for English native speakers except for the differences in the duration of phrase presentations (see Table 1 for the summary of the design). It turned out in a pilot study that the duration used for native speakers (i.e., an 800ms-800ms-1300ms time frame for cleft sentences; and an 800ms-800ms-800ms-1300ms time frame for RC sentences) was too short for L2ers to fully process the sentences, so the duration
was slightly lengthened for the L2ers. Cleft sentences had a 1000ms-1000ms-2000ms time frame, and RC sentences had a 1000ms-1000ms-1000ms-2000ms time frame.

The whole experiment session took 40 to 75 minutes, and the L2ers generally took longer than the English native speakers who participated in Experiment 1.

5.3.1.4. Data analysis

The data analysis method for Experiment 3 was identical to that of Experiment 1, described in Section 4.1.1.4.

5.3.2. Results

5.3.2.1. English reading span test data from the L2ers

The mean reading span score for the L2ers was 40.96 (SD = 10.56) out of the total score of 70, and the scores ranged from 23 to 66. These scores were lower than those of the 28 native subjects in Experiment 1, whose mean reading span score was 49.00 (SD = 9.58) and ranged from 28 to 66. In fact, this raises the possibility that reading span may simply be another indication of English proficiency. In order to test if reading span is a function of language proficiency, a correlational analysis of the L2ers' reading span scores and their cloze test scores was conducted. The results show that there is no significant correlation between these two factors \( r = .224, p = .293 \), indicating that L2ers' reading span score measures a variable that is independent of language proficiency. The L2ers' English reading span scores are presented in Table 5 below as well as in Appendix E.
5.3.2.2. Comprehension data

The mean comprehension accuracy for unambiguous fillers was 91.17% (SD = 5.27), which shows that the L2ers’ comprehension was quite accurate, despite the fact that the fillers included such complex sentences as doubly nested sentences.

The mean non-local attachment response in the EC condition was 50.67% (SD = 31.75) and 46.75% (SD = 32.43) in the SC condition, which in both cases was slightly higher than the native speakers’ results (37.05% (SD = 25.68) for the EC condition, 40.63% (SD = 27.24) for the SC condition). A one sample t-test was used for the EC condition and SC condition separately to examine whether these means differed significantly from 50% in each condition. No significant difference was found for the results from either the EC or SC condition [t (1, 23) = 103, p = 0.919; t (1, 23) = -0.491, p = 0.628, respectively], showing that the L2ers (as a group) did not have a clear attachment preference. However, as we saw in Experiment 1, RC attachment preferences may interact with working memory capacity, so apparent chance-level performance in the group mean may have resulted from averaging data across low-spans who produce, say, more non-local attachment responses and high-spans who produce, say, more local attachment responses.

Next, the relation between working memory capacity and RC attachment responses was examined. The L2 subjects were divided into two groups according to the cut-off score of 41, such that those who scored below 41 was classified as low-spans and those who scored 41 or above as high-spans. A repeated measures ANOVA with complexity (EC vs. SC) as a within-subjects variable and span size as a between-subjects variable revealed that there is neither a main effect of complexity or span size [F (1, 22) = 3.525, MS = 1.558, p =
nor a significant interaction of complexity and span size \( F(1, 22) = 3.525, MS = 1.558, p = .074 \). These results show, unlike in Experiment 1, that span size had no effect on RC attachment responses. The ANOVA table for this statistical analysis is given in Appendix K.

The lack of complexity effect was also observed in Experiment 1, but as noted above, this may simply be due to the use of a question-answering strategy. As for the lack of interaction between RC attachment responses and reading span, it may be due to the measurement of reading span. One may argue that L2ers' performance in the English reading span test may not be a proper indication of their working memory resources, since their performance could potentially be complicated by their L2 proficiency. However, this possibility is very unlikely. First, as reported in Section 5.3.2.1, there was only a weak correlation between their cloze test scores and reading span scores, which shows that at least reading span scores are not entirely dependent on proficiency. Second, if it is true that the lack of span size effect on RC attachment preferences is because of the English reading span test being complicated by their proficiency, then their reading span measured in the L1 should show a substantial effect on the English RC attachment responses, since these L2ers are all native speakers of Japanese and there should be no significant variation in their native language proficiency. Experiment 4, to be reported below, measured Japanese reading span scores from the same set of subjects, so we examined whether the Japanese reading span scores correlate with English reading span. The results show, first, that there was a significant correlation between Japanese and English reading span scores \( r = .677, p = .000 \), signifying that there is a high correlation between the construct being measured in
the English reading span test and the Japanese reading span test. Next, a repeated measures ANOVA was conducted again, this time based on the span group defined by Japanese reading span scores (the cut-off score was 47, based on the mean Japanese reading span score). Again, the same pattern of findings as with the English off-line data were obtained; there was no main effect for complexity or span size, and there was no significant interaction of complexity and span (see Appendix K for the ANOVA table for this analysis). This shows that the lack of interaction between English reading span and English attachment responses was not due to the assessment method of working memory resources.

Thus, it appears that the absence of clear attachment preferences in the group results is not a function of individual differences due to working memory capacity. With the lack of a reliable effect of individual factors such as working memory capacity, do the group means that were around 50% actually indicate that they have no attachment preference? As Frenck-Mestre (2005) pointed out (see Section 3.2), it may still be the case that, for some unknown reason, some learners have a clear non-local attachment preference and others have a clear local attachment preference, and averaging across these subjects produces means close to 50%. Individual data from Experiment 3 are thus analyzed below to examine whether all learners actually have, more or less, no clear attachment preference.

Table 5 presents the L2ers’ cloze test scores, reading span scores, and mean non-local attachment responses in the EC and SC conditions. In Table 5, the means of L2ers’ individual non-local responses were classified into three categories: (a) Non-local preference (65%-100%), (b) null preference (35%-65%), and (c) local preference (0%-35%). Also, in Table 5, as we are interested in how many participants fall into the ‘null
Table 5. Individual data from Experiment 3

<table>
<thead>
<tr>
<th>Subject</th>
<th>Cloze ( \text{(Max = 50)} )</th>
<th>Reading span ( \text{(Max = 70)} )</th>
<th>Mean non-local response (in %)</th>
<th><strong>preference</strong></th>
<th><strong>preference</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>J1</td>
<td>47</td>
<td>49</td>
<td>0.00% Local</td>
<td>6.25% Local</td>
<td></td>
</tr>
<tr>
<td>J2</td>
<td>46</td>
<td>43</td>
<td>6.25% Local</td>
<td>0.00% Local</td>
<td></td>
</tr>
<tr>
<td>J3</td>
<td>45</td>
<td>26</td>
<td>56.25% Null**</td>
<td>68.75% Non-local</td>
<td></td>
</tr>
<tr>
<td>J4</td>
<td>45</td>
<td>38</td>
<td>100.00% Non-local</td>
<td>100.00% Non-local</td>
<td></td>
</tr>
<tr>
<td>J5</td>
<td>45</td>
<td>35</td>
<td>87.50% Non-local</td>
<td>81.25% Non-local</td>
<td></td>
</tr>
<tr>
<td>J6</td>
<td>44</td>
<td>37</td>
<td>6.25% Local</td>
<td>31.25% Local</td>
<td></td>
</tr>
<tr>
<td>J7</td>
<td>44</td>
<td>41</td>
<td>62.50% Null</td>
<td>62.50% Null</td>
<td></td>
</tr>
<tr>
<td>J8</td>
<td>43</td>
<td>34</td>
<td>75.00% Non-local</td>
<td>62.50% Null</td>
<td></td>
</tr>
<tr>
<td>J9</td>
<td>42</td>
<td>44</td>
<td>75.00% Non-local</td>
<td>62.50% Null</td>
<td></td>
</tr>
<tr>
<td>J10</td>
<td>42</td>
<td>54</td>
<td>56.25% Null</td>
<td>50.00% Null</td>
<td></td>
</tr>
<tr>
<td>J11</td>
<td>42</td>
<td>66</td>
<td>62.50% Null</td>
<td>43.75% Null</td>
<td></td>
</tr>
<tr>
<td>J12</td>
<td>40</td>
<td>43</td>
<td>31.25% Local</td>
<td>18.75% Local</td>
<td></td>
</tr>
<tr>
<td>J13</td>
<td>40</td>
<td>56</td>
<td>68.75% Non-local</td>
<td>68.75% Non-local</td>
<td></td>
</tr>
<tr>
<td>J14</td>
<td>39</td>
<td>32</td>
<td>12.50% Local</td>
<td>12.50% Local</td>
<td></td>
</tr>
<tr>
<td>J15</td>
<td>39</td>
<td>61</td>
<td>25.00% Local</td>
<td>18.75% Local</td>
<td></td>
</tr>
<tr>
<td>J16</td>
<td>37</td>
<td>41</td>
<td>12.50% Local</td>
<td>6.25% Local</td>
<td></td>
</tr>
<tr>
<td>J17</td>
<td>37</td>
<td>35</td>
<td>62.50% Null</td>
<td>68.75% Non-local</td>
<td></td>
</tr>
<tr>
<td>J18</td>
<td>37</td>
<td>45</td>
<td>75.00% Non-local</td>
<td>68.75% Non-local</td>
<td></td>
</tr>
<tr>
<td>J19</td>
<td>36</td>
<td>41</td>
<td>18.75% Local</td>
<td>0.00% Local</td>
<td></td>
</tr>
<tr>
<td>J20</td>
<td>35</td>
<td>41</td>
<td>62.50% Null</td>
<td>56.25% Null</td>
<td></td>
</tr>
<tr>
<td>J21</td>
<td>35</td>
<td>26</td>
<td>25.00% Local</td>
<td>12.50% Local</td>
<td></td>
</tr>
<tr>
<td>J22</td>
<td>34</td>
<td>36</td>
<td>93.75% Non-local</td>
<td>93.75% Non-local</td>
<td></td>
</tr>
<tr>
<td>J23</td>
<td>34</td>
<td>36</td>
<td>37.50% Null</td>
<td>25.00% Local</td>
<td></td>
</tr>
<tr>
<td>J24</td>
<td>34</td>
<td>23</td>
<td>100.00% Non-local</td>
<td>100.00% Non-local</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>EC</th>
<th>SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of 'non-local' preference</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Total number of 'local' preference</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Total number of 'null' preference</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>

* Subjects were classified into three categories based on the mean non-local responses: Non-local preference (65%-100%), null preference (35%-65%), and local preference (0%-35%).

** Shading highlights the individual data that fell into the null preference category.

preference' category, the null preference category is marked by shading of the cell.

As Table 5 clearly shows, there are not many L2ers who were categorized as having a null preference in both the EC and SC conditions: More than half of the subjects have either a
clear local or a clear non-local attachment preference. This shows that the mean group results reported earlier were around 50% because this was, in large part, an averaging effect of two different groups of L2ers, one with a clear local attachment preference and the other with a clear non-local attachment preference.

Finally, the L2ers’ mean non-local attachment response in the EC and SC conditions was compared to that of English native speakers. A repeated measures ANOVA with L1 as a between-subjects variable and complexity as a within-subjects variable showed that there was no significant difference between English native speakers from Experiment 1 and Japanese L2ers from Experiment 3 for either the EC condition \[ F(1, 50) = 2.828, MS = .233, p = .099 \] or the SC condition \[ F(1, 50) = .534, MS = .047, p = .468 \]. This suggests, despite the apparent differences in the mean non-local attachment response, that the native speaker group and the advanced learner group did not show a significant difference in RC attachment preferences.

5.3.3. Discussion

Let us now summarize the findings from Experiment 3. First, the advanced Japanese L2ers did not show an effect of their reading span scores (both Japanese and English) on their RC attachment responses in English, indicating that memory resources did not influence their RC attachment preferences. Supporting this finding, the complexity manipulation did not influence their attachment preferences either, as the mean non-local attachment response in the EC and SC conditions did not show a significant difference. However, this could also be due to a methodological flaw as discussed in Chapter 4, in that
the presentation of stimuli allows subjects to use a question-answering strategy. Second, L2ers’ mean non-local attachment responses were close to 50% and appeared slightly higher than that of English native speakers, but no significant difference was observed between the two groups. However, the analysis of individual data shows that the group average of chance-level performance does not mean that all L2ers had no clear attachment preference and behaved around 50%: The majority of them had either a clear local or a clear non-local attachment preference, and the seemingly chance-level performance resulted from averaging across these two groups.

With respect to the lack of span size effect, two possibilities can be considered. The first is that L1 parsers and L2 parsers are simply qualitatively different in such a way that working memory resources constrain L1 parsing behavior but they do not constrain L2 parsing behavior. This would be compatible with the results in Juffs (2004, 2005), who also found no relation between working memory measures and reading time data (see Section 2.1). The second possibility is that L2ers may be different from English native speakers in grammatical knowledge of the Saxon genitive, which was suggested as a potential source of a local attachment preference in English (see Section 3.1). Recall that it was suggested in Chapter 4 that the functional availability of the Saxon genitive may correlate with working memory capacity, in such a way that only high-spans have sufficient resources to consider the availability of a Saxon genitive when processing the Norman genitive, which leads (by way of Gricean inference) to a local attachment preference. However, in the case of L2ers, they may not have acquired target-like knowledge that the Saxon genitive only allows attachment to the local NP. In this case,
even if the L2ers have high working memory capacity, they would not consider the presence of the Saxon genitive alternative as a reason to choose a local attachment response. In fact, the surface form of the English Saxon genitive (e.g., 'the schoolgirl’s sister') is rather similar to the genitive construction in Japanese, e.g., ‘schoolgirl-no sister’ (English words are used for expository purposes), where the only difference is that the ‘-’s” is replaced by the genitive case marker “-no”; however, this Japanese genitive construction allows a preceding RC to modify either of the NPs. Given this, it may not be implausible that Japanese learners may consider the English Saxon genitive as allowing both attachment possibilities.

The results of the current study do not distinguish between these two possibilities, but there are two ways to test between them. One is to actually test whether the L2ers know that a Saxon genitive followed by an RC is unambiguous. If the lack of span size effect on attachment preferences was due to incomplete knowledge of the Saxon genitive, then L2ers should consider such constructions to be ambiguous. On the other hand, if it is shown that the L2ers do have this knowledge but still exhibit a lack of interaction between memory resources and attachment preferences, then this would lend support to the first possibility, namely that the role of working memory is simply different in L1 and L2 processing.

Another test, which is adopted in the present thesis, is to examine whether a language which does not have both the Norman genitive and Saxon genitive (i.e., a language with only one genitive construction) shows a systematic relation between working memory resources and attachment preferences. If a language without more than
one type of genitive, such as Japanese, does show such effects, then this would cast doubt on the view that attributes the relation between memory resources and attachment preferences to the capacity to consider the Saxon genitive as an alternative to the Norman genitive. In fact, the influence of working memory on RC attachment has only been tested in English (present study; Mendelsohn & Pearlmutter, 1999; Swets et al., 2004) and Dutch (Swets et al., 2004), both of which have Norman and Saxon genitives, so we do not know whether languages without the two genitive forms also show a span size effect on RC attachment. By contrast, if languages with only one genitive construction show no relation between working memory capacity and attachment preferences, then it would lend indirect support to the view that links memory resources to the functional availability of the Saxon genitive. Experiment 4, which was originally designed for the purpose of testing the extent of L1 influence, can shed light on this issue, since this experiment measures both working memory resources and attachment preferences in Japanese. We will thus return to this point in Experiment 4.

Next, the fact that there was no significant difference between the English native speakers and Japanese L2ers in the group means of non-local attachment responses suggests that these advanced L2ers as a group do behave similarly to the group of English native speakers. However, these group results must be interpreted with caution: First, it is not certain whether we could even define “target-like” RC attachment preferences, given the large variation among English native speakers, as observed in Experiment 1. Table 6 on the next page presents individual data from Experiments 1 and 3. As Table 6 shows, the
Table 6. Individual data from Experiments 1 and 3

<table>
<thead>
<tr>
<th>Subject</th>
<th>Reading span</th>
<th>Mean non-local response</th>
<th>Subject</th>
<th>Cloze span</th>
<th>Mean non-local response</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>28</td>
<td>87.50% 100.00%</td>
<td>J1</td>
<td>47</td>
<td>0.00% 6.25%</td>
</tr>
<tr>
<td>E2</td>
<td>31</td>
<td>50.00% 43.75%</td>
<td>J2</td>
<td>46</td>
<td>6.25% 0.00%</td>
</tr>
<tr>
<td>E3</td>
<td>36</td>
<td>0.00% 12.50%</td>
<td>J3</td>
<td>45</td>
<td>56.25% 68.75%</td>
</tr>
<tr>
<td>E4</td>
<td>37</td>
<td>43.75% 43.75%</td>
<td>J4</td>
<td>45</td>
<td>100.00% 100.00%</td>
</tr>
<tr>
<td>E5</td>
<td>38</td>
<td>37.50% 62.50%</td>
<td>J5</td>
<td>45</td>
<td>87.50% 81.25%</td>
</tr>
<tr>
<td>E6</td>
<td>38</td>
<td>50.00% 50.00%</td>
<td>J6</td>
<td>44</td>
<td>6.25% 31.25%</td>
</tr>
<tr>
<td>E7</td>
<td>42</td>
<td>87.50% 81.25%</td>
<td>J7</td>
<td>44</td>
<td>62.50% 62.50%</td>
</tr>
<tr>
<td>E8</td>
<td>43</td>
<td>62.50% 56.25%</td>
<td>J8</td>
<td>43</td>
<td>75.00% 62.50%</td>
</tr>
<tr>
<td>E9</td>
<td>46</td>
<td>50.00% 37.50%</td>
<td>J9</td>
<td>42</td>
<td>75.00% 62.50%</td>
</tr>
<tr>
<td>E10</td>
<td>47</td>
<td>62.50% 56.25%</td>
<td>J10</td>
<td>42</td>
<td>56.25% 50.00%</td>
</tr>
<tr>
<td>E11</td>
<td>47</td>
<td>18.75% 43.75%</td>
<td>J11</td>
<td>42</td>
<td>62.50% 43.75%</td>
</tr>
<tr>
<td>E12</td>
<td>47</td>
<td>62.50% 81.25%</td>
<td>J12</td>
<td>40</td>
<td>31.25% 18.75%</td>
</tr>
<tr>
<td>E13</td>
<td>49</td>
<td>18.75% 25.00%</td>
<td>J13</td>
<td>40</td>
<td>68.75% 68.75%</td>
</tr>
<tr>
<td>E14</td>
<td>50</td>
<td>68.75% 81.25%</td>
<td>J14</td>
<td>39</td>
<td>12.50% 12.50%</td>
</tr>
<tr>
<td>E15</td>
<td>51</td>
<td>25.00% 25.00%</td>
<td>J15</td>
<td>39</td>
<td>25.00% 18.75%</td>
</tr>
<tr>
<td>E16</td>
<td>51</td>
<td>31.25% 18.75%</td>
<td>J16</td>
<td>37</td>
<td>12.50% 6.25%</td>
</tr>
<tr>
<td>E17</td>
<td>53</td>
<td>6.25% 12.50%</td>
<td>J17</td>
<td>37</td>
<td>62.50% 68.75%</td>
</tr>
<tr>
<td>E18</td>
<td>54</td>
<td>25.00% 25.00%</td>
<td>J18</td>
<td>37</td>
<td>75.00% 68.75%</td>
</tr>
<tr>
<td>E19</td>
<td>55</td>
<td>62.50% 56.25%</td>
<td>J19</td>
<td>36</td>
<td>18.75% 0.00%</td>
</tr>
<tr>
<td>E20</td>
<td>56</td>
<td>12.50% 12.50%</td>
<td>J20</td>
<td>35</td>
<td>62.50% 56.25%</td>
</tr>
<tr>
<td>E21</td>
<td>56</td>
<td>50.00% 81.25%</td>
<td>J21</td>
<td>35</td>
<td>25.00% 12.50%</td>
</tr>
<tr>
<td>E22</td>
<td>56</td>
<td>0.00% 18.75%</td>
<td>J22</td>
<td>34</td>
<td>93.75% 93.75%</td>
</tr>
<tr>
<td>E23</td>
<td>57</td>
<td>50.00% 43.75%</td>
<td>J23</td>
<td>34</td>
<td>37.50% 25.00%</td>
</tr>
<tr>
<td>E24</td>
<td>58</td>
<td>18.75% 6.25%</td>
<td>J24</td>
<td>34</td>
<td>100.00% 100.00%</td>
</tr>
<tr>
<td>E25</td>
<td>58</td>
<td>0.00% 6.25%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E26</td>
<td>58</td>
<td>31.25% 37.50%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E27</td>
<td>64</td>
<td>12.50% 18.75%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E28</td>
<td>66</td>
<td>12.50% 0.00%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

English native speakers' range of mean non-local attachment response is 0% to 87.50% for the EC condition and 0% to 100% for the SC condition, and if we define "target-like" performance based on this range, then almost any mean value would be considered "target-like."
The second reason why group-level comparisons between the English native speakers and the Japanese L2ers is that, as the individual data analyses in Table 5 showed, there seem to be three groups of L2ers, one having a clear local attachment preference, another having a clear non-local attachment preference, and the other having a null preference.\textsuperscript{27} Then, even if we assume, based on the group mean of English natives, that a local attachment preference is the English native-like performance, then nearly half of the L2ers can be considered to have learned to behave native-like. It is not entirely clear why there are such differences in RC attachment preferences among the L2ers, but this at least shows that group results can be misleading when considering whether L2ers can learn to behave target-like in off-line L2 processing. This casts doubt on the previous studies that found no clear attachment preference in off-line processing (e.g., Felser et al., 2003), and reanalysis of their data at individual levels seems necessary to test whether each of their subjects overall really had no clear attachment preference.\textsuperscript{28}

One factor that needs to be addressed in L2 RC attachment preferences is potential L1 transfer effects. Recall that Kamide and Mitchell (1997) reported that Japanese native speakers have a non-local attachment preference in off-line processing. Based on this finding, we could argue that this L1 preference may transfer to English and affect their L2 RC attachment preferences. For example, in the individual data presented in Table 5, eight subjects showed a non-local attachment preference in the EC condition, and eight subjects showed a null attachment preference.\textsuperscript{27}

\textsuperscript{27} R. Bley-Vroman (personal communication, June 13, 2005) points out that the distribution observed in the L2 English results (Table 5) may represent a uniform distribution rather than a binomial distribution, given that there are very few subjects producing non-local attachment responses around 50% of the time. However, the small sample size of the present study prevents us from drawing any conclusions regarding the nature of the distribution of the data (J. Norris, personal communication, June 30, 2005).

\textsuperscript{28} Unfortunately, however, individual data are not reported in the previous L2 sentence processing research and thus unavailable for reanalysis.
(six of whom also preferred non-local attachment in the EC condition) had a non-local attachment preference in the SC condition. One could argue that these non-local attachment preferences in L2 English may have resulted from L1 transfer, but two caveats are in order. First, given such large variations in both native and non-native RC attachment preferences (see Table 6), the question of transfer may need to be addressed at an individual level. For example, if one speaker has a local attachment preference in both Japanese and English while a group of Japanese speakers shows a non-local attachment preference in the group average, then this speaker is misinterpreted not to show L1 transfer. Second, there have not been many studies that tested Japanese off-line RC attachment preferences, and it may not be warranted to conclude that Japanese speakers in general actually have a non-local attachment preference in off-line processing, as Kamide and Mitchell (1997) found. In fact, Kamide and Mitchell as well as Miyamoto et al. (2004) showed that Japanese speakers show a local attachment preference in on-line processing, which, for some reason, differs from their off-line RC attachment preferences. For these reasons, in order to investigate L1 transfer of RC attachment preferences, it is best to test the same individuals on their L1 and L2, and examine whether their L1 and L2 RC attachment preferences coincide. This will be investigated in Experiment 4.

In summary, Experiment 3 showed that Japanese L2ers can learn to behave like English native speakers in RC attachment resolution, although their RC attachment preferences were found not to be influenced by working memory capacity. Next, Experiment 4 investigates how the same group of subjects behaves in Japanese RC attachment in order to examine the extent of L1 transfer in their L2 results. As discussed
before, this should also be able to shed light on whether the relation between working
memory resources and attachment preferences may be related to the functional availability
of the Saxon genitive construction.

5.4. Experiment 4

This experiment investigates the Japanese RC attachment preferences by the
subjects who participated in Experiment 3. The research questions addressed in this
experiment are given in (18).

(18) Research questions addressed in Experiment 4

a. In Japanese off-line processing, do working memory and RC attachment
   preferences interact, as observed in Experiment 1?

b. Do Japanese speakers have a non-local RC attachment preference in off-line
   processing, as observed in Kamide and Mitchell (1997)?

c. To what extent do the Japanese L2ers of English show the same preference
   in their L1 and L2?

5.4.1. Method

5.4.1.1. Participants

The same group of Japanese speakers as in Experiment 3 participated in
Experiment 4, which took place approximately 2.5 months after Experiment 1. They were
paid an additional $10 for their participation.
5.4.1.2. Materials

In order to keep as constant as possible the complexity manipulation used in Experiments 1 and 3, Japanese center-embedding sentences and left-branching sentences were constructed based on Babyonyshev and Gibson (1999) and Nakatani and Gibson (2003). An example of these sentences is illustrated in (19):

(19) a. Japanese embedded-clause (EC) condition

[Hahaoya-ga [musuko-ga [bebiisittaa-ga [pro senjitu yakedo-o-sita]
mother-Nom son-Nom babysitter-Nom the other day burn-Acc-did

girl-Gen sister-Dat like-Pass-was-Comp lie-Acc tell-Pst-Comp think-Pst

b. Japanese sentential complement (SC) condition

[bebiisittaa-ga [pro senjitu yakedo-o-sita] [[syoojo-no] imooto-ni]
babysitter-Nom the other day burn-Acc-did girl-Gen sister-Dat

suka-re-teita-to] [musuko-ga t₁ uso-o tui-ta-to] [Hahaoya-ga t₂ omotte-ita]
like-Pass-was-Comp son-Nom lie-Acc-tell-Pst-Comp mother-Nom think-Pst

“The mother thought that the son lied that the babysitter was liked by the sister of the girl who got burned the other day.”

These two sentences contain the RC attachment ambiguity with respect to “who got burned the other day.” The propositions expressed by these two sentences are identical, but the complexity of these sentences is quite drastically different. If we consider the storage costs according to the model of Gibson (1998, 2000) which we previously used to calculate the storage costs for English sentences, the difference in complexity between (19a) and (19b) becomes clear. Assuming that the resolution of attachment ambiguity occurs at the point of the complex NP “syoojo-no imooto-ni” (“the sister of the girl”), the center-embedding
sentence in (19a) incurs a storage cost of three,\textsuperscript{29} since the initial three nominative marked subjects each need verbs to form a grammatical sentence. In the case of (19b), on the other hand, the storage cost at the point of RC attachment is only one, since the initial subject "bebiisittaa-ga ("the babysitter")" is the only NP that requires a verb to form a grammatical sentence.\textsuperscript{30} In this way, the Japanese versions of the EC condition and the SC condition also exhibit a difference of two storage costs; hence it is predicted that the magnitude of the complexity effect should be similar to that of the English sentences used in Experiments 1 and 3. In fact, it is possible that the Japanese EC condition may cause even more processing difficulties. Lewis (1996) proposed that storage costs may increase when the elements retained in memory are similar (cf. Gordon, Hendrick, \& Johnson, 2002). In the case of Japanese center-embedding sentences as in (19a), all three constituents that create storage costs are nominative-marked subjects, and Lewis predicts that the similarity among the three constituents would create more memory interference (Lewis \& Nakayama, 2002). For this reason, the Japanese EC condition may posit even more processing difficulties than the English EC condition.

\textsuperscript{29} Note, however, that the prediction of the storage costs would slightly differ for on-line processing. (19a) involves an RC "pro senjitu yakedo-o sita" ("pro got burned the other day") inside the clause initiated by "bebiisittaa-ga" ("the babysitter"), but the predicate "senjitu yakedo-o sita" can initially be construed as the predicate of "bebiisittaa-ga" since the null subject pro is not visible. Therefore, upon reaching the NP following the RC (i.e., "syoojo-no imooto-ni" ("the sister of the girl")), can the parser discover that (a) there was an RC and (b) the preceding predicate must be construed with this RC's head NP rather than "bebiisittaa-ga." In this context, if the predicate "senjitu yakedo-o sita" was actually processed as the predicate of "bebiisittaa-ga," then the storage cost is decreased by one at the point of "syoojo-no imooto-ni," although this would cause a different processing burden due to a subsequent need for reanalysis. However, in the present off-line study, the whole sentence is visible and hence the readers are expected to process the structure correctly in the end, in which case the original prediction of a storage cost of three can be maintained.

\textsuperscript{30} Both (19a) and (19b) contain pro which will be co-indexed with one of the NPs in the complex NP. Here, this co-indexation is not counted as a source of storage cost since it is not clear in Gibson's model whether it should lead to such. If co-indexation does increase storage cost, then a storage cost of one should be added to both conditions, in which case the difference in storage costs between the two still remains the same.
Thirty-two sentences of the type exemplified in (19) were constructed. So as to keep the Japanese and English as comparable as possible, lexical items similar to those in the English sentences were used in the Japanese sentences so that there would be as little lexical effect as possible. These 32 sentences were counter-balanced across two lists and were interspersed between 65 filler sentences of various types (53 of which were unambiguous), such that experimental items were always separated by at least one filler sentence. The order of experimental items was pseudo-randomized for each subject. The list of experimental items used in Experiment 4 is given in Appendix L.

5.4.1.3. Procedure

The procedure for the comprehension experiment implemented on E-prime version 1.1 was identical to that of Experiments 1 and 3, except that in this case, their reading time and the time it took them to answer the question was measured by the computer. This reading time data can be used to examine whether the subjects actually use the question-answering strategies mentioned before. If they indeed just look at the question first and then only look at the relevant regions of the target sentence, then it is expected that reading times between the EC and SC conditions should not differ significantly. However, if the mean reading times in the EC condition are significantly longer than the mean reading times in the SC condition, then this indicates that they do read the whole sentence and experience more processing difficulties in reading the EC condition. This off-line experiment took approximately 30 minutes. Next, the subjects took the Japanese reading span test described in Section 2.2.2., which took 15 to 20 minutes. Then, the subjects
completed a questionnaire that was irrelevant for the purpose of the present study. The whole session took approximately one hour.

5.4.1.4. Data analysis

The data analysis method was mostly identical to that of Experiment 1, described in Section 4.1.1.4. In order to address L1 transfer effects, categorical attachment preferences of Experiment 4 will also be compared to the L2 English data from Experiment 3.

5.4.2. Results

5.4.2.1. Japanese reading span

As was done in the English reading span test, Japanese reading span scores were calculated from the sentences in which both acceptability judgment and word recall were correct. The mean reading span score was 47.04 (SD = 8.00), and the scores ranged from 32 to 62. Individual reading span scores are reported in Appendix E. The subjects were divided into low-span and high-span groups according to the cut-off score of 47.

5.4.2.2. Comprehension and reading time data

The mean comprehension accuracy for the 53 unambiguous fillers for the 24 subjects was 97.58% (SD = 3.58), and all subjects were accurate more than 83% of the time, showing that a fairly high comprehension accuracy was achieved.

Next, the mean non-local attachment response was calculated for the experimental items, and the individual data are presented in Table 7, together with the individual data
Table 7. L2ers’ individual data from Experiments 4 and 3

<table>
<thead>
<tr>
<th>Subject</th>
<th>Experiment 4: Japanese off-line</th>
<th></th>
<th>Experiment 3: English off-line</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean non-local response</td>
<td></td>
<td>Mean non-local response</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EC</td>
<td>SC</td>
<td>prefer*</td>
<td></td>
</tr>
<tr>
<td>J1</td>
<td>37.50%</td>
<td>Null</td>
<td>50.00%</td>
<td>Null</td>
</tr>
<tr>
<td>J2</td>
<td>0.00%</td>
<td>Local</td>
<td>0.00%</td>
<td>Local</td>
</tr>
<tr>
<td>J3</td>
<td>31.25%</td>
<td>Local</td>
<td>31.25%</td>
<td>Local</td>
</tr>
<tr>
<td>J4</td>
<td>25.00%</td>
<td>Local</td>
<td>50.00%</td>
<td>Null</td>
</tr>
<tr>
<td>J5</td>
<td>68.75%</td>
<td>Non-local</td>
<td>68.75%</td>
<td>Non-local</td>
</tr>
<tr>
<td>J6</td>
<td>12.50%</td>
<td>Local</td>
<td>12.50%</td>
<td>Local</td>
</tr>
<tr>
<td>J7</td>
<td>43.75%</td>
<td>Null</td>
<td>37.50%</td>
<td>Null</td>
</tr>
<tr>
<td>J8</td>
<td>31.25%</td>
<td>Local</td>
<td>25.00%</td>
<td>Local</td>
</tr>
<tr>
<td>J9</td>
<td>25.00%</td>
<td>Local</td>
<td>31.25%</td>
<td>Local</td>
</tr>
<tr>
<td>J10</td>
<td>18.75%</td>
<td>Local</td>
<td>56.25%</td>
<td>Null</td>
</tr>
<tr>
<td>J11</td>
<td>12.50%</td>
<td>Local</td>
<td>37.50%</td>
<td>Null</td>
</tr>
<tr>
<td>J12</td>
<td>31.25%</td>
<td>Local</td>
<td>56.25%</td>
<td>Null</td>
</tr>
<tr>
<td>J13</td>
<td>18.75%</td>
<td>Local</td>
<td>25.00%</td>
<td>Local</td>
</tr>
<tr>
<td>J14</td>
<td>37.50%</td>
<td>Null</td>
<td>31.25%</td>
<td>Null</td>
</tr>
<tr>
<td>J15</td>
<td>6.25%</td>
<td>Local</td>
<td>37.50%</td>
<td>Null</td>
</tr>
<tr>
<td>J16</td>
<td>18.75%</td>
<td>Local</td>
<td>31.25%</td>
<td>Null</td>
</tr>
<tr>
<td>J17</td>
<td>31.25%</td>
<td>Local</td>
<td>37.50%</td>
<td>Null</td>
</tr>
<tr>
<td>J18</td>
<td>81.25%</td>
<td>Non-local</td>
<td>81.25%</td>
<td>Non-local</td>
</tr>
<tr>
<td>J19</td>
<td>31.25%</td>
<td>Local</td>
<td>43.75%</td>
<td>Null</td>
</tr>
<tr>
<td>J20</td>
<td>25.00%</td>
<td>Local</td>
<td>18.75%</td>
<td>Local</td>
</tr>
<tr>
<td>J21</td>
<td>87.50%</td>
<td>Non-local</td>
<td>81.25%</td>
<td>Non-local</td>
</tr>
<tr>
<td>J22</td>
<td>50.00%</td>
<td>Null</td>
<td>37.50%</td>
<td>Null</td>
</tr>
<tr>
<td>J23</td>
<td>56.25%</td>
<td>Null</td>
<td>62.50%</td>
<td>Null</td>
</tr>
<tr>
<td>J24</td>
<td>87.50%</td>
<td>Non-local</td>
<td>81.25%</td>
<td>Non-local</td>
</tr>
</tbody>
</table>

Number of subjects with the same preference in the EC condition: 9 (6 local, 3 non-local)
Number of subjects with the same preference in the SC condition: 7 (4 local, 3 non-local)
Number of subjects with the same preference in both conditions: 6 (3 local, 3 non-local)

* Subjects were classified into three categories based on the mean non-local responses: Non-local preference (65%-100%), null preference (35%-65%), and local preference (0%-35%).
** Shading highlights the subjects who had the same RC attachment preference in their L1 and L2 for each condition.

from Experiment 3 for comparison. For the EC condition, the mean non-local attachment response was 36.20% (SD = 24.45), while for the SC condition, the mean non-local attachment response
attachment response was 42.71% ($SD = 21.47$). Two one sample $t$-test showed that the number of non-local attachment responses was significantly below chance-level (i.e., the test value set at 50%) in the EC condition [$t(23) = -2.766, p = .011$], but not in the SC condition [$t(23) = -1.664, p = .110$]. These data show that a reliable local attachment preference was observed in the EC condition, while the SC condition showed only a weak trend toward a local attachment preference. The pattern of these results appears similar to the results obtained in Experiment 1 for English native speakers (see Table 2). On average, their mean non-local attachment response was significantly lower than 50% in the EC condition but only numerically lower in the SC condition. This overall trend towards a local attachment preference in Japanese contrasts remarkably with the previous off-line results from Kamide and Mitchell (1997), who found a non-local attachment preference (i.e., mean non-local attachment response from 26 subjects was 66%).

The difference in the extent of preferences in the two conditions suggests that the complexity manipulation may have had an effect on RC attachment preferences, so the responses from the two conditions were compared against each other. First, in this experiment, reading times were measured for reading the target sentence and answering a question after reading the sentence, so let us begin by examining whether the reading times differ between the EC condition and the SC condition. The combined reading times (to read the sentence and to answer the question) were submitted to a repeated measures ANOVA with complexity (EC vs. SC) as a within-subject variable. No significant difference was found in the reading time between the two conditions [$F(1, 23) = 2.806, MS = 12899567.52, p = .107$], although the mean reading times for the EC condition was
numerically larger than that of the SC condition (18594.35ms vs. 17557.54ms, respectively). Next, the mean non-local attachment response in the EC and the SC conditions was also submitted to a repeated measures ANOVA. This time, a main effect of complexity was observed \([F(1, 23) = 5.497, MS = 5.086, p = .028]\), showing that the non-local attachment response in the EC condition was significantly lower than in the SC condition.

In order to investigate the relation between span size, complexity and RC attachment responses, we conducted a repeated measures ANOVA with span size as a between-subjects variable and complexity (EC vs. SC) as a within-subjects variable. The ANOVA table for this statistical analysis is given in Appendix M. This analysis revealed that there was a main effect of complexity \([F(1, 22) = 5.309, MS = 5.086, p = .031]\), but neither a main effect of span size nor a significant effect of complexity-by-span interaction was found. These results show, unlike the native English data in Experiment 1, that span size had no effect on RC attachment responses while complexity had a main effect, such that the EC condition exhibited significantly less non-local attachment responses.

We now turn to the extent of transfer in the data of Experiment 3. In order to investigate this, a correlation analysis was conducted on the data from Experiments 3 and 4. However, the correlational analysis for the 24 subjects shows that there was no significant correlation between English and Japanese for the mean non-local attachment response in the EC condition or the SC condition \([r = .354, p = .089, r = .217, p = .308, \text{ respectively}]\). The lack of significant correlations between the two experiments shows that, as a group, there was no systematic relation between their performance in L1 and L2.
However, if we were to analyze the data on the individual level, it might turn out that some subjects show transfer whereas for the rest of them, transfer does not occur. In order to further investigate whether any of the subjects show evidence of L1 transfer, individual data from Experiments 3 and 4 were categorized again into three categories: (a) Non-local preference (65%-100%), (b) null preference (35%-65%), and (c) local preference (0%-35%). These categories were then compared across the two experiments. It was reasoned that, if a subject falls into the same category of attachment preference under the same sentence type (i.e., the EC or SC conditions), this could be taken as evidence for L1 transfer.\(^{31}\) The cells that point to L1 transfer are marked by shading in Table 7. This individual analysis revealed that six subjects fell into the same category of preference (either local or non-local) in both the EC condition and the SC condition. There were three subjects who showed L1 transfer only in the EC condition, and there was also one subject who showed L1 transfer only in the SC condition.

5.4.3. Discussion

The major findings of the experiment can be summarized as follows: First, the Japanese speakers as a group showed a local attachment preference in the EC condition and a trend for a local attachment preference in the SC condition, which was quite different from previous findings by Kamide and Mitchell (1997). Second, the mean non-local attachment response in the EC condition was significantly lower than that of the SC condition, showing that the increased complexity in the EC condition led to increased local

\(^{31}\) When a subject showed a null preference in both experiments, this was not taken as an instance of transfer, as it was reasoned that null preference may have been caused by unpredicted extraneous factors (e.g., lexical biases in the stimuli) rather than the subjects' linguistic systems.
attachment responses. Third, no main effect of reading span size was observed for the EC condition or for the SC condition, showing that reading span did not affect the Japanese RC attachment preferences. Fourth, Japanese RC attachment responses did not strongly correlate with their L2 English attachment responses, either in the EC condition or in the SC condition. However, the analysis of the individual data indicates that there are at least 9 of 24 subjects behaving similarly across the Japanese and English experiments, showing that there are individual differences in L1 transfer. Let us examine the implications of these findings.

It is not entirely clear why the Japanese attachment preferences in this study differ markedly from the previous findings. One possible explanation is that the general complexity of the sentences used in the present study had an effect of preventing the Japanese speakers from choosing the non-local attachment response in off-line processing. Kamide and Mitchell (1997) as well as Miyamoto et al. (2004) showed that Japanese speakers prefer local attachment in on-line processing, while Kamide and Mitchell's off-line study nevertheless found a non-local attachment preference. This suggests that in off-line processing, for some unclear reason, Japanese speakers choose to change their initial analysis and prefer a non-local attachment resolution. Thus, the present findings can be interpreted as indicating that, in the current experiment, this initial local attachment, which happened in on-line processing, was not modified after reading the whole sentence. Given the increased complexity for the EC condition and the increased number of propositions for the EC and SC conditions, the parser may have run out of resources to carry out the reanalysis in off-line processing. In fact, this interpretation goes along with the fact that
more local attachment responses were found in the EC condition than in the SC condition. This can be interpreted as evidence that the complexity of the EC condition consumed so many resources that the off-line reanalysis was less likely to happen than in the SC condition. Another possible explanation for the observed local attachment preference is that the Japanese learners were influenced by the statistical frequencies of the English RC attachment (cf. Dussias, 2003). Since all of the subjects lived in the US at the time of the experiment, this possibility cannot be ruled out a priori. However, an on-going study by Miyao and Omaki (2005) used Kamide and Mitchell’s experimental items and tested 11 Japanese speakers in Hawaii, and they found that these Japanese speakers had a clear non-local RC attachment preference in off-line processing ($M = 93.94, SD = 11.79$). This lends support to the view that the discrepancy between the present findings and Kamide and Mitchell’s was most likely due to the differences in the experimental items. However, in order to exclude environmental influence as a factor more completely, the items used in the present study should be tested on Japanese speakers living in Japan. If they behave similarly to the participants of the current study, then we can safely conclude that the environmental factor was not the cause of the observed local attachment preference.

Next, the present study found no effect of reading span size on RC attachment responses in either the EC or SC conditions, unlike with the English speakers in Experiment 1. This finding is compatible with the suggestion made in the previous chapter that individual differences in working memory may in fact be the cause of individual differences in the functional availability of the Saxon genitive. Since Japanese has no such
alternative form of the genitive construction that allows unambiguous attachment,\textsuperscript{32} individual differences in working memory, which may define the ability of the parser to consider such alternative forms, do not interact with attachment preferences in Japanese. Moreover, this may explain why L2ers did not show an effect of reading span on their English RC attachment preferences (Experiment 3). It may be the case that they have not acquired the relevant grammatical properties of the English Saxon genitive, such as disallowing non-local attachment of a following RC. If this property is not acquired, then the L2ers would not consider the Saxon genitive as a reason to choose local attachment responses with the Norman genitive. This may also explain why L2ers’ mean non-local attachment responses in English were relatively higher than those of English native speakers. However, in order to establish the lack of working memory effects as resulting from not having the functional availability of the Saxon genitive, further studies are needed to test (a) whether L2ers have acquired the knowledge that a Saxon genitive followed by an RC is unambiguous and (b) whether the presence or absence of this knowledge has an effect on RC attachment preferences.\textsuperscript{33}

On the other hand, the lack of a systematic difference between high-spans and low-spans raises another question, that is, it is not clear why the complexity manipulation in the

\textsuperscript{32} B. D. Schwartz (personal communication, June 13, 2005) points out that, in Japanese, in addition to the ambiguous “RC NP2-no NP1” construction tested here, there is also an “NP2-no RC NP1” construction, in which NP1 is unambiguously modified. Note, however, that this is a very infrequent construction, and for this reason, this construction may simply not occur to subjects when processing “RC NP2-no NP1.” If this was the case, then the presence of this unambiguous RC attachment construction would not influence Japanese speakers’ RC attachment preferences in the same way as the Saxon/Norman genitive alternation does in English.

\textsuperscript{33} Another possibility is that the L2ers are unable to use the pragmatic inference based on the Gricean maxim, regardless of their grammatical knowledge of the Saxon genitive. In order to address this, one needs to use an additional test of the availability of the Gricean maxim. How this could be independently tested, however, needs to be investigated.
EC and SC conditions led to a significant difference in attachment responses, while the working memory capacity difference between low-spans and high-spans did not lead to different attachment preferences. If, as we discussed before, the reason why the EC condition produces more local attachment is that the parser is left with fewer resources in this condition, then it should naturally be the case that low-spans prefer local attachment even more than high-spans do. One speculation is that in off-line processing experiments like Experiment 4, the difficulties with reanalysis due to a lack of resources may be overridden by low-spans, whereas the effect of complexity remains constant, no matter how long subjects spend reading the sentence. Thus, based on the current results, it seems most reasonable to conclude that fewer resources lead to more local attachment responses, although further studies are clearly needed to examine if the same relation holds when the n size is larger and also when tested in on-line processing.

Finally, the lack of a strong correlation between Japanese and English RC attachment responses appears to suggest that the Japanese L2ers in the present study may not be influenced very much by their L1 attachment preferences. However, as the individual analyses showed, six subjects had the same attachment preference in L1 and L2 in both the EC and SC conditions, and two subjects showed the same preference in the EC condition for L1 and L2, and one subject had the same preference in the SC condition. These data could be considered as evidence for L1 transfer.\textsuperscript{34} It is not entirely clear why

\textsuperscript{34} Note, however, that some of the learners who showed the same preference in L1 and L2 had a local attachment preference, and it could be argued that showing a local attachment preference in English is a result of learning, rather than L1 transfer. However, given that English native speakers in Experiment 1 considerably vary in their attachment preferences, as discussed in Section 5.3.3, it is extremely difficult to define what it is that is to be learned. For this reason, we remain agnostic about whether showing a local attachment preference in L1 and L2 can be considered a result of learning.
there are such individual differences in L1 transfer, but it is nevertheless the case that some of the advanced Japanese L2ers of English still show (apparent) L1 influence in their L2 processing. This, again, raises questions about the validity of the claim made in Felser et al. (2003) or Papadopoulou and Clahsen (2003) that L1 processing strategies do not transfer; if their data were reanalyzed on an individual level, we might find some of the subjects showing patterns consistent with L1 transfer, which may not have been visible on the group level analysis.

5.5. Experiment 5

This section presents results from the on-line experiment with the advanced Japanese L2ers. It had the same design as Experiment 2, which tested on-line processing of RC attachment ambiguity by English native speakers. This experiment mainly addresses the following two issues: working memory effects on on-line processing by L2ers, and L2ers' RC attachment preferences in on-line experiments. The research questions of this experiment are stated as follows.

(20) Research questions addressed in Experiment 5

a. In Japanese L2ers' English on-line processing, do working memory and RC attachment preferences interact?

b. Do Japanese speakers show a clear RC attachment preference in on-line processing in L2 English?
Comparison of the off-line processing behavior in Experiments 1 and 3 showed that the advanced Japanese L2ers did not exhibit an effect of working memory resources on RC attachment preferences, unlike the English native speakers. However, as suggested in the previous section, the lack of span size effect on RC attachment preferences by L2ers in Experiment 3 may be due to the lack of grammatical knowledge regarding properties of the Saxon genitive which only allows unambiguous attachment of a following RC. As for English native speakers’ off-line processing, it was suggested that only high-spans may have sufficient resources to consider this alternative and hence show a local attachment preference, whereas low-spans might not have such resources available and hence prefer non-local attachment. Recall now, however, that the on-line results in Experiment 2 presented a slightly different picture: Even the native low-spans who showed a non-local attachment preference in off-line processing showed a local attachment advantage (although not significant) in the EC condition, in which sentence complexity was increased. It was suggested that attachment preferences observed in off-line processing may be irrelevant for on-line processing. Given these findings, it would be interesting to investigate whether L2ers would also show such effects. Even advanced L2ers, high-span or low-span, would still suffer from having to spend more processing resources to use their L2 knowledge, and it may be the case that L2ers as a group may behave like the low-span group of English native speakers.

Another question that this experiment addresses is whether L2ers can show a clear attachment preference at all in on-line L2 processing. The review of the L2 psycholinguistics literature in Chapter 3 indicated that previous studies investigating on-
line RC attachment preferences have seen mixed results, but this may also reflect that their overall reading time data were potentially complicated by individual differences in working memory capacity, and so this possibility is also addressed in Experiment 5. Furthermore, the effect of working memory may become even more visible in an on-line task in which the processing burden is expected to be bigger than in off-line experiments where subjects can take as much time as they need to comprehend the sentence. This possibility will be explored by comparisons between the off-line results from Experiment 3 and on-line results from Experiment 5.

5.5.1. Method

5.5.1.1. Participants

The same group of 24 advanced Japanese speakers as in Experiments 1 and 2 participated in Experiment 5, which took place approximately one week after they took part in Experiment 4. They were paid an additional $10 for their participation.

5.5.1.2. Materials

The materials used in Experiment 5 were identical to those used in Experiment 2. An example of the 32 experimental sentences was given in (14), which is repeated below as (21). These 32 sentences were counter-balanced across two lists and interspersed between 73 fillers such that the same type of sentence was never presented consecutively. The experimental sentences were pseudo-randomized for each subject.
To review, these sentences can tell us about RC attachment preferences by potential reading time slow-downs caused by a gender mismatch. For example, if someone has a local attachment preference, then this person would show a processing time slow-down in (21b) and (21d) since the local attachment resolution would cause a gender mismatch between "the schoolgirl" and "himself," which should lead to reanalysis of the initial parse and hence produce a slower reading time. Similarly, a subject with a non-local attachment preference should show a processing time slow-down in the reflexive region of (21a) and (21c), since the non-local resolution would cause a gender mismatch and subsequent revision of the initial analysis.

(21)  

a. Embedded Clause (EC), forced local attachment

The babysitter that [the brother of the schoolgirl who burned herself the other day adored] was very nice.

b. Embedded Clause (EC), forced non-local attachment

The babysitter that [the brother of the schoolgirl who burned himself the other day adored] was very nice.

c. Sentential Complement (SC), forced local attachment

The babysitter said that [the brother of the schoolgirl who burned herself the other day was very nice].

d. Sentential Complement (SC), forced non-local attachment

The babysitter said that [the brother of the schoolgirl who burned himself the other day was very nice].
5.5.1.3. Procedure

This experiment adopted exactly the same procedure as that of Experiment 2. The experiment is a word-by-word, non-cumulative, moving-window self-paced reading task (Just, Carpenter, & Woolley, 1982), and subjects' button presses, triggering elimination of the current word and presentation of the next word, were timed as reading time per region. After the experiment, they completed a background questionnaire. They did not take a reading span test this time, since the reading span data from Experiment 1 were available. The whole session took approximately 45 to 50 minutes.

5.5.1.3. Data analysis

The data analysis method for Experiment 5 was mostly identical to that of Experiment 2, described in Section 4.2.1.4.

5.5.2. Results

5.5.2.1. Comprehension questions

The advanced Japanese L2ers' overall mean comprehension accuracy calculated from all the sentences was 75.48 (SD = 8.28) and the scores ranged from 59.05% to 87.62%, which seems quite similar to native speakers' performance, with a mean accuracy of 74.14% (SD = 8.99), ranging from 56% to 94.29%. As mentioned in Chapter 4, this low comprehension accuracy was probably due to the complexity of the experimental items as well as the fillers, the latter having been controlled to be similar to experimental items in terms of length and complexity. This time, as no L2 subjects' mean accuracy was below
two standard deviations from the mean, no subject was excluded from further data analyses.

Next, the mean comprehension accuracy for the four target conditions was calculated: 74.48% (SD = 21.64) for the EC/local attachment condition (21a), 72.40% (SD = 24.45) for the EC/non-local attachment condition (21b), 68.23% (SD = 16.48) for the SC/local attachment condition (21c), and 73.96% (SD = 19.48) for the SC/non-local attachment condition (21d). A repeated measures ANOVA with complexity (EC vs. SC) and attachment (local vs. non-local) as within-subject variables and reading span measured in Experiment 1 (low-span vs. high-span) as a between-subjects variable was conducted on the mean comprehension accuracy. This analysis showed no main effect of complexity [$F(1, 22) < 1, MS = 1.318, p = .594$], attachment [$F(1, 22) < 1, MS = 7.975, p = .534$], or reading span [$F(1, 22) = 4.297, MS = .274, p = .050$]. No significant effect was found for complexity-by-attachment interaction [$F(1, 22) = 1.117, MS = 3.662, p = .302$], complexity-by-reading span interaction [$F(1, 22) = 2.639, MS = .119, p = .119$], attachment-by-reading span interaction [$F(1, 22) < 1, MS = 1.465, p = .789$], or complexity-by-attachment-by-reading span interaction [$F(1, 22) < 1, MS = 4.069, p = .728$]. In sum, none of the factors showed an effect in the mean comprehension accuracy data.

5.5.2.2. Reading times

The L2ers’ reading time data were analyzed in the same way as in Experiment 2. Reading times in all experimental items are analyzed regardless of how the comprehension
questions were answered. In order to exclude outliers, reading times beyond three standard deviations from the mean for a given condition and position were trimmed to the value of three standard deviations, affecting 1.41% of the data. Regions of the target sentence that are statistically analyzed are shown in Table 8, repeated from Table 3.

Table 8. Regions of interest in the target sentences (same as Table 3)

<table>
<thead>
<tr>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
<th>Region 4</th>
<th>Region 5</th>
<th>Region 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>that</td>
<td>the</td>
<td>brother</td>
<td>of</td>
<td>the</td>
<td>schoolgirl</td>
</tr>
<tr>
<td>Region 7</td>
<td>Region 8</td>
<td>Region 9</td>
<td>Region 10</td>
<td>Region 11*</td>
<td>Region 12</td>
</tr>
<tr>
<td>who</td>
<td>burned</td>
<td>herself/himself</td>
<td>the other day</td>
<td>adored</td>
<td>was</td>
</tr>
</tbody>
</table>

*Region 11 is missing in the SC condition, but “was” in the SC condition is still referred to as Region 12 for the sake of convenience.

Overall reading time data from the crucial regions are presented in Figure 7. The reaction time data in these regions were submitted to a repeated measures ANOVA with complexity (EC vs. SC), attachment (local vs. non-local) as within-subjects independent variables. The main effect of complexity was observed for Region 5 “the” \( [F (1, 31) = 5.864, \text{MS} = 33253.32, p < .05] \) and Region 12 “was” \( [F (1, 31) = 7.654, \text{MS} = 289044.62, p < .01] \), showing that the reading times in these regions were significantly slower in the EC condition than in the SC condition. The ANOVA tables for these two regions (Regions 9 and 12) are given in Appendix N. However, there was no main effect of complexity or attachment, nor a complexity-by-attachment interaction in any of the other regions, including the critical “himself/herself” region (Region 9) which was designed to tap into their RC attachment preferences. Compared to the native speaker results reported for Experiment 2, this shows that the effect of storage costs was not as visible in the L2ers'
Figure 7. Mean reading times in crucial regions in Experiment 5
results. However, although storage cost effects were not observed throughout the regions
in which they were expected (i.e., Regions 2 through 8), a main effect of complexity was
observed in one of the regions, Region 5, which is suggestive of the fact that the L2ers do
experience some sort of storage cost during these regions. Moreover, the integration cost
effect observed in Region 12 shows that at least in this respect, L2ers did perceive the EC
condition to be more complex in the same way English native speakers did.

Next, the reading times were further examined to see if reading span size interacts
with any of the factors in the relevant regions. However, no effects were found for any of
the factors.

Finally, although the critical Region 9 “himself/herself” showed no effect of any of
the factors, the overall mean reading time data were analyzed here numerically, since the
lack of an effect may simply have been due to the small sample size, and the pattern of
reading time distribution may become significant when more data are collected. The
reading times for this region are summarized in Figure 8, presented together with the
reading time data from the low-span English native speakers. For the L2ers (right hand
side of Figure 8), the mean reading time was 776.52ms (SD = 307.39) for the EC/local
attachment condition (21a), 832.63ms (SD = 411.77) for the EC/non-local attachment
condition (21b), 786.93ms (SD = 339.14) for the SC/local attachment condition (21c), and
775.29ms (SD = 324.42) for the SC/non-local attachment condition (21d). In fact, this
distribution of reading time data looks quite similar to that of the low-span group of
English native speakers.
5.5.3. Discussion

Let us start by summarizing the present findings in relation to the data from English native speakers tested in Experiment 2. First, the effect of the complexity manipulation was visible only in two regions for L2ers: one from Region 5 which seemingly results from storage cost effects in Region 5 (in contrast to English native speakers who showed significant storage cost effects in multiple regions, i.e., Regions 4, 6, 7 and 8), and another from Region 12 which reflects the integration cost effect, which was also observed in the English native speakers’ results. Second, no clear attachment preference and no significant reading span effect was observed in any of the regions, which is quite different from the English native speakers’ results in that they showed a significant local attachment advantage (at least when the EC/SC conditions are collapsed) as well as an effect of span size-by-complexity interaction in Region 12. Third, despite these differences and despite
the findings of no significance from the L2 data, an inspection of their overall reading times in the critical region (Region 9) suggests that their behavior is quite similar to that of low-span English native speakers.

Even though the number of regions in which the complexity effect was observed was smaller for L2ers, apparent effects of storage cost and integration cost were still observed in some of the regions for the L2ers. For this reason, we will assume here that the complexity manipulation was effective for the L2ers as well. Nevertheless, the attachment variable did not show a main effect in Region 9, so we cannot draw a firm conclusion about L2ers’ attachment preferences based on the null findings.

The general lack of significant findings in the L2 self-paced reading data may be due to the nature of the dependent measure in the self-paced reading task. Frenck-Mestre (2005) points out that reading time data may, oftentimes, be complicated, since the reading time alone may not distinguish initial analysis from reanalysis that may be happening between button presses. She furthermore suggests that eye movement measures, by contrast, can be more sensitive in that it provides multiple measures of sentence processing (e.g., first-pass, second-pass, regression, etc.) and thus more precisely reflect the various stages of processing.

This needs to be addressed by conducting more studies using various techniques, including measurement of eye movements or measurement of event-related potentials.

Nevertheless, the overall similarity we observed in Figure 8 suggests the possibility that low-span English native speakers and advanced L2ers may behave similarly. In Experiment 2, although the general trend for the English native speakers seems to be that
both low-spans and high-spans prefer local attachment, it was reported that low-spans and high-spans display a slightly different reading time distribution, in that the condition in which numeric differences appear in reading times is different for low-spans and high-spans (see Figure 6). As for the L2ers, regardless of their span size, they may generally be experiencing a huge processing burden, which may have led them to behave like low-span English native speakers. However, there was no significant difference between the local and non-local attachment conditions, and the potential effect also seems numerically smaller than in the low-span native speaker data. Further studies with a larger n size and a more sensitive measure of initial stages of processing are thus necessary before drawing any firm conclusions regarding the relation between working memory resources and online attachment preferences for L2ers.35

5.6. Summary

This chapter presented results from three experiments to investigate how advanced Japanese L2ers of English behave in English and Japanese with respect to RC attachment ambiguities. Experiment 3 replicated Experiment 1 to test these L2ers’ off-line English attachment preferences as well as test whether their preferences are associated with their working memory capacity. The results show that their reading span size did not exhibit a significant effect on their attachment preferences, unlike the case of the English native

35 Given that we found robust individual variation in Experiment 3, the ideal situation, if possible, would be to investigate the effect of individual differences in reading times. However, this is not conducted here as the author is not aware of a way to analyze the reading times on an individual level. In Experiment 3, the dependent variable was an average of non-local responses, and one could intuitively assign some value to its average. However, when it comes to reading times, raw reading time data cannot be interpreted by themselves, since it is only by using inferential statistics that we can judge whether the observed difference is meaningful or not. For this reason, only group results are reported for the on-line experiments.
speakers in Experiment 1. However, despite the lack of working memory influence, the individual analysis showed that some of the L2ers behaved target-like, exhibiting a local attachment preference.

Experiment 4 tested their attachment preferences in their L1, Japanese, to see if the findings in Experiment 3 may be complicated by the effect of transfer. On the group level, their attachment preferences in English and Japanese did not show a significant correlation, but the individual analysis showed that some of the L2ers behaved very similarly in their L1 and L2 RC attachment resolution, suggesting that L1 transfer does exist in L2 processing by these advanced Japanese L2ers of English. Moreover, the Japanese RC attachment preferences were not associated with their Japanese reading span size, which lends indirect support to the view that the reason why in English (and Dutch; Swets et al., 2004) reading span size has an effect on attachment preferences may be due to the presence of the Saxon genitive in these languages (Frazier & Clifton, 1996). Upon encountering a Norman genitive, only high-spans have sufficient resources to consider a Saxon genitive as an alternative genitive form that allows unambiguous modification of the non-local noun, and this induces the Gricean maxim of “Avoid Ambiguity” and hence they calculate that the Norman genitive was (probably) used to modify the local NP. It was thus suggested that the lack of effect of reading span on L2 English attachment responses could possibly be due to their incomplete acquisition of the relevant grammatical properties of the Saxon genitive in English.

Experiment 5 tested the L2ers’ on-line RC attachment preferences in English, adopting the method used in Experiment 2 for English native speakers. Although the effect
of complexity manipulation in the materials was less visible here than in the native speakers' data, the complex EC condition still showed a significantly slower reading time in two regions, reflecting the effect of storage and integration costs. However, no effect was found for any of the factors examined in the present study, namely, complexity, attachment or reading span. On the other hand, a numerical analysis of the reading time data in the critical region (Region 9) showed that the general pattern of the L2ers' reading time data regarding RC attachment preferences appears similar to that of low-span English native speakers, suggesting that these advanced L2ers as a group may behave like low-span native speakers. It was suggested that this happens due to L2ers' general processing difficulties.
This thesis investigated various aspects of L1 and L2 processing of RC attachment with a particular focus on the potential effect of working memory capacity. The major findings from the previous chapters are summarized here, based on which we will discuss implications for research on the role of working memory in sentence processing and general L2 processing research.

6.1. Summary of the major findings

Experiments 1 and 2 tested English native speakers' off-line and on-line RC attachment preferences and their relations to working memory capacity. In both experiments, the working memory factor was investigated by relying on two sources of information, i.e., reading span as measured in a reading span test, and a potential effect of complexity manipulation of the target sentences. It was expected that, if low-spans and high-spans show different behavior due to the difference in their memory resources, the attachment responses in the complex condition and the non-complex condition should also differ since the former condition consumes more resources, which may lead high-spans to behave as low-spans do.

Results of Experiment 1 replicated the previous findings that greater working memory capacity leads to increased local attachment responses, indicating that more high-spans than low-spans prefer local attachment (Mendelsohn & Pearlmutter, 1999; Swets et
al., 2004). However, complexity manipulation did not show a significant effect. One possible cause considered for the lack of the expected effect was the design that allowed subjects to use a question-answering strategy. However, Experiment 4, which tested the effect of complexity in Japanese and adopted the same design as Experiment 1, did show a main effect of complexity on attachment responses. So the lack of a main effect in English may have been due to there not being sufficient complexity to cause changes in off-line processing behavior. Returning to the first result, it is not entirely clear why greater working memory leads to an increase of local attachment responses, but drawing on Heydel and Murray (2005), it was suggested that upon processing the Norman genitive construction, high-spans can consider the alternative Saxon genitive in a parallel fashion, which then leads to a local attachment response due to the Gricean maxim of “Avoid Ambiguity.” On the other hand, low-spans may not have sufficient resources to perform such parallel computations, and without this information, their attachment is guided by the Referentiality Principle (Frazier & Clifton, 1996). This principle favors modification of the head of the argument of the matrix verb, which leads to non-local attachment. Moreover, this approach is indirectly supported by the findings in Experiment 4 that Japanese, which has only one form for the genitive construction, does not show an effect of reading span size on attachment responses (see below).

Experiment 2 tested similar materials in native English on-line processing, which would presumably cause more of a processing burden on the subjects. In this experiment, it was found that the attachment preferences were generally low and (again) did not interact with the complexity manipulation, despite the fact that in this experiment the increased
complexity clearly led to a slow-down of reading times in many regions. A significant interaction between the complexity factor and reading span was observed in the region where integration costs were expected to show up in the reading time data (Gibson, 1998, 2000), such that high-spans exhibited a larger integration cost effect than low-spans did. However, considering the fact that high-spans were slower than low-spans in general, this interaction was interpreted as reflecting a merely proportional increase in reading time, rather than a difference in the parsers of low-spans and high-spans. Data from other regions where low-spans and high-spans were expected to show differences in reading time data indicated that they did not differ in their behavior.

Next, using these data as a baseline for comparison, Experiments 1 and 2 were recycled for L2ers as Experiments 3 and 5 in order to address L1 transfer and working memory effects on L2 processing. Experiment 4 was also conducted to examine the L2ers’ attachment preferences in their native Japanese with materials that closely follow the design of the materials used in Experiment 1.

The results of Experiment 3 indicated that the advanced Japanese L2ers of English do not show an effect of working memory capacity on RC attachment preferences, nor did they show an effect of complexity manipulation. It was suggested that the lack of effect may stem from incomplete acquisition of the grammatical properties of the Saxon genitive construction, since knowledge of the unambiguity of a Saxon genitive followed by an RC is a prerequisite to its functioning as a trigger for the Gricean principle that leads to local attachment responses. Further, Experiment 4 tested the same L2ers’ off-line Japanese RC attachment preferences, so as to examine whether the data in Experiment 3 may have been
complicated by a potential effect of L1 transfer. First, the Japanese group data showed that, unlike the non-local preference result of the previous off-line study by Kamide and Mitchell (1997), they had a local attachment preference. This difference arose presumably due to the difference in complexity of the items between the previous study and the current study. Next, no meaningful correlation was found between the L2ers' off-line English and Japanese RC attachment responses. However, aside from these group mean results, analyses of individual data indicate that (a) some L2ers do learn to behave target-like in L2 RC attachment resolution, and (b) some of the L2ers resolve RC attachment ambiguity in a very similar way in their L1 and L2, suggesting that L1 influence exists for L2 processing for some advanced L2ers.

Finally, Experiment 5 tested the L2ers' on-line resolution of English RC attachment ambiguity. No statistically significant data were obtained with respect to their RC attachment preferences, but an inspection of the distribution of their reading time data in the critical region (Figure 8) suggests that their behavior is, to some extent, similar to that of low-span English native speakers, in that there was a slight local attachment advantage. It was suggested that advanced L2ers might behave like low-span native speakers, since L2ers are likely to experience greater processing difficulties in using their unstable L2 knowledge to parse L2 input. However, no significant differences were observed for these data, so further studies are needed to explore this possibility.
6.2. **Implications for future research**

The present findings from native speakers of English shed light on the debate between the Single Resource (SR) model (e.g., Just & Carpenter, 1992) and the Separate-Sentence-Interpretation-Resource (SSIR) model (Caplan & Waters, 1999). Recall that the SR model predicts that low-spans and high-spans should exhibit a distinct behavior in sentence processing due to the differences in their memory resources (e.g., low-spans would experience substantially more processing difficulties in complex regions), whereas the SSIR model predicts that there are no such individual differences in interpretive processing. In regard to these predictions, first of all, Experiment 2 revealed a significant interaction of span size and complexity in one region only, where integration cost effects were expected to occur. This shows that the complex EC condition produced substantially larger reading times for the high-spans, and this finding appears to support the Single Resource model, but this needs to be treated with caution. High-spans were generally slower in reading times across several regions, including this particular region in question, as evidenced by their slower reading times in the non-complex SC condition. Given these general differences (possibly due to speed-accuracy trade-offs for high-spans), it is possible that the observed significant interaction may simply have been due to a proportional reading time increase for the high-spans, and not due to differences in sentence processing mechanisms of the low-spans versus the high-spans. The fact that no other region produced significant difference between high-spans and low-spans (including regions where the complexity is increased and hence low-spans were expected to experience more processing difficulties) suggests that this interpretation seems rather likely.
to be true, which then would lend support to the SSIR model. However, considering the small sample size, we must remain agnostic about the debate between the two models until further studies with a larger number of subjects are conducted.

With respect to RC attachment preferences, the present study conducted analyses of group means as well as individual data, which led to the finding that, regardless of L1 or L2, the group average may in fact be misleading in interpreting RC attachment preferences. In the past literature, researchers generally looked only at the group results and concluded as if the group of speakers under investigation (e.g., native speakers of English or Spanish, German learners of English, and so on) all had the same preference, but if these data had been analyzed on an individual level, as in the present study, not all the subjects might not have behaved in the same way. In other words, a particular tendency found in each study may be unique to the particular sample of the population that they investigated; this is indeed possible given that the previous studies did not have a sample size that may be large enough to represent the population. To further complicate the issue, as discussed in Section 3.1, RC attachment preferences are claimed to be influenced by various factors, and it is thus extremely difficult to control for all the factors and create unbiased items. In fact, this variable nature of RC attachment preferences calls into question whether RC attachment preferences can be treated as a theoretical construct. In order to address this issue, future studies need to investigate (a) specifically, which of the suggested factors (e.g., syntax and pragmatics of the genitive construction, prosody, working memory, statistical frequencies in input, on-line vs. off-line processing, etc.) truly account for RC attachment preferences as well as how these factors (potentially) interact, (b) which of these factors are subject to
individual differences and cross-linguistic differences, and (c) whether the effects of these factors hold across languages when tested on a really large sample (e.g., 1000 subjects) from each language group.

With respect to L2ers, many more questions remain unanswered. First, although it was shown in individual-level analyses that some L2ers can behave target-like in off-line L2 processing and some L2ers show evidence of L1 transfer, it is unclear where these individual differences come from. Working memory capacity was explored in the present thesis as a potential variable underlying such individual differences, but as we saw in Experiment 3, the L2ers did not show an effect of working memory capacity on RC attachment preferences. Further L2 studies are needed to examine the various factors that have been claimed to affect RC attachment preferences (e.g., prosody, frequencies in the input, the functional availability of the Saxon genitive, etc.; see Section 3.1).

Furthermore, the on-line results of the L2ers somewhat displayed similarity to those of the low-span English native speaker group, but again, the slight local attachment advantage observed for L2ers is not statistically significant. For this reason, a study with a larger number of subjects is clearly needed to further investigate the effect of working memory capacity on L2ers’ RC attachment preferences. Also, as pointed out in Frenck-Mestre (2005), it may be critical to use an on-line measure such as eye-tracking that is more revealing of initial stages of parsing.
6.3. Concluding remarks

The research reported in the present thesis investigated the relation between working memory resources and RC attachment preferences in English native speakers and advanced Japanese L2ers of English. Generally, it was found that working memory capacity interacted to some extent with aspects of L1 English off-line processing, but not with L1 English on-line processing, L1 Japanese off-line processing, L2 English off-line processing, or L2 English on-line processing. However, further studies with (a) a larger number of subjects, (b) refined techniques with more precise measures of initial and later stages of processing, and (c) an independent test of the syntax and pragmatics of the Saxon genitive form and of its relation to working memory capacity, are necessary to investigate more thoroughly the suggestions made in this research.
APPENDIX A

SENTENCES USED IN THE JAPANESE READING SPAN TEST

The sentences are presented in the order in which they appeared in the reading span test. The underlined words are the target words for each sentence. The slashes indicate the segmentation pattern.

2-sentence condition

工房では/職人が/シャツを/手織りしている。
結婚記念日に/腕時計を/お母さんは/お父さんが/買ってくれたと/驚いていた。
撮影現場で/カメラマンは/フィルムが/アシスタントを/忘れたと/しっかりと。
夏休み前に/課題図書を/先生が/紹介した。
知事選挙にて/マスコミが/世論を/飲んだ。
悲しいことに/現地人は/登山客が/きれいな山を/汚してしまうと/嘆いている。
会議で/奨学金を/学長は/大学院生らが/獲得したと/発表した。
昔のフランスでは/芸術家は/学校教育を/決して信用する。
最近/手紙を/若い人は/書かなくなった。
厨房では/シェフは/ウェイターが/料理を/氷いだと/言っていた。

3-sentence condition

繁華街で/警察官を/指名手配中の男が/殴った。
夏休みあけに/読書感想文が/子供たちを/提出した。
懐かしきに/祖父は/祖母が/愛犬を/可愛がったことを/思い出していた。
控え室で/店長は/アルバイトが/ちょっとも/仕事をすると/嘆いている。
毎年毎年/テレビ局は/視聴率競争を/繰り広げている。
朝食には/目玉焼きを/父は/必ず食べる。
飼育場で/ライオンが/飼育係を/読んだ。
受付で/風邪を/看護婦は/薬剤師が/用意していると/言っていた。
教室では/マンガ本を/男の子たちが/飛んでいた。
中間決算によると/売上高が/過去最高を/記録した。
試乗会で/新車を/営業マンが/売りさばいていた。
職員会議で/評価を/先生方は/子供たちが/めったに引かないと/心配していた。
発表会で/ピアノを/女の子が/読んだ。
ありがとうことに/財布を/警察は/親切な人が/既に届けてくれたと/伝えてくれた。
受験直前に/家庭教師は/宿題が/教え子を/忘れると/怒った。
4-sentence condition

昨日/祖母が/大根の種を/蒔いていた。
舞台稽古で/舞台監督は/脚本家が/台本を/飲んだと/伝えた。
突然/授業を/大学院生は/先生が/休講にしたと/伝えてきた。
大慌てで/大金を/専務が/泳いだ。
大寒波で/ブドウ畑が/大きな被害を/受けた。
去年の誕生日に/パーティーを/友達が/極秘に計画してくれた。
味付けに/料理長が/料理酒を/泣いた。
驚くことに/栄養士を/トマトは/たくさん含んでいる。
展示会で/デザイナーが/奇麗なニットを/紹介した。
院中では/キリンを/エサが/食べていた。
突然/バスガイドは/運転手が/道を/間違えたと/説明した。
法廷で/弁護士を/裁判長は/証拠が/出したと/述べた。
かわいそうになることに/友達が/恋人が/めったに約束を/守ると/悲しんでいた。
最近では/クッキーを/製菓会社が/競って発売している。
オーケーションで/絵画が/実業家を/買い落とした。
雪国では/雪だるまを/保母さんは/子ども達が/喜んで作ると/おしゃべった。
昨夜/台風が/関東地方を/襲った。
最近では/企画書を/社長が/決して受け入れる。
ニュースで/レポーターは/野球選手が/ファンを/大事にしていると/言っていた。
テレビで/パリを/芸能記者は/有名女優が/こっそり塗っていたと/しゃべっていた。

5-sentence condition

驚いたことに/新刊本を/店長は/お客が/粗末に扱っているのを/気にも留めなかった。
試合中に/ファンが/相手チームを/眠った。
年末に/年賀状を/父は/必死に書いている。
県議会で/市民団体は/知事が/少年法改正案を/採択することを/要求した。
チケット売り場で/販売員が/新作のチケットを/売り始めた。
教会で/神父様は/みんなが/平和を/飲んでいると/お話した。
幸せそうに/紙飛行機を/おじいちゃんは/子供たちが/飛ばすのを/じっと見ていた。
試写会で/舞台挨拶を/映画スターが/着た。
新聞調査によると/サラリーマンは/余暇を/十分に持てないている。
ここ数年/交通規則を/警察は/若いドライバーが/食べると/発表した。
卒業式で/感謝状を/私の姉が/読み上げた。
アンコールを求めて/観客が/拍手を/決して止めた。
国会では/予算案を/議員たちが/話し合っていた。
真夏になると/カフェが/バリの風景を/明るくする。
テーブル席で/ビールを/お客は/ウェインが/もっとも持ってくると/わめっていた。
予定日を過ぎても/原稿を/作家が/もっとも渡す。
報告会議で/秘書は/社員たちが/給料値上げを/要求していると/報告した。
学会で/小判を/隊長は/調査隊が/泳いだと/発表した。
道端で/占い師が/手相を/占っている。
展示会で/デザイナーは/モデルが/服を/飛んだと/言っていた。
散歩の途中で/犬を/穴が/ほりだした。
昨日の新聞で/新聞記者は/首相が/フランスを/訪問するだろうと/記していたようだ。
対談番組で/ゲストを/司会者は/お土産が/持ってきたと/はしゃいだ。
診察後に/お医者さんは/息子が/風邪を/こじらせてしまったと/おっしゃっていた。
残念なことに/アルバイトを/先生方は/近頃の学生が/歩いていたと/嘆いていた。
APPENDIX B

SENTENCES USED IN THE ENGLISH READING SPAN TEST

The sentences are presented in the order in which they appeared in the reading span test. The underlined words are the target words for each sentence. The slashes indicate the segmentation pattern.

2-sentence condition

It was / the snow / that excited the skiers.
The award / pleased / the actor / that the review upset.
The guitarist / bought / the CD / that dropped the boy.
It was / the ball / that the boy threw at the window.
It was / the passenger / that delighted the music.
The painter / praised / the architect / that designed the museum.
The mayor / supported / the candidate / that the issue worried.
It was / the customer / that pleased the price.
It was / the composer / that the opera amused.
The rain / ended / the game / that played the children.

3-sentence condition

It was / the flower / that the girl cherished.
It was / the war / that protested against the leader.
The dinner / disgusted / the manager / that owned the restaurant.
The burglar / stole / the diamond / that watched the guard.
It was / the bible / that the priest dropped on the floor.
It was / the banker / that deposited the cash.
It was / the researcher / that interested the lecture.
The map / guided / the explorers / that the storm frightened.
It was / the princess / that the apple ate.
It was / the prisoner / that escaped from the jail.
It was / the audience / that the comedy entertained.
The discussion / followed / the lectures / that the participants bored.
It was / the carpenter / that the house built.
The coach / trained / the athlete / that the letter surprised.
The fire / burnt / the magazine / that read the family.
4-sentence condition

It was / the movie / that impressed the lady.
The secretary / sent / the money / that requested the author.
The land / excited / the sailors / that the journey exhausted.
It was / the wave / that the surfers frightened.
It was / the politician / that enacted the law.
It was / the scientist / that the experiment excited.
It was / the computer / that fixed the student.
It was / the beach / that the tourists visited.
It was / the student / that fixed the computer.
It was / the report / that the policeman astonished.
The poem / amused / the musician / that wrote the song.
The violinist / composed / the melody / that the dancers excited.
The earthquake / destroyed / the restaurant / that owned the manager.
It was / the surfer / that the wave frightened.
It was / the cash / that deposited the banker.
The nurse / greeted / the patient / that the medicine relaxed.
It was / the lecture / that interested the researcher.
It was / the girl / that the flowers cherished.
The pianist / scolded / the boy / that dropped the CD.
The assistant / brought / the medicine / that the patient relaxed.

5-sentence condition

The fan / praised / the dancer / that the melody excited.
It was / the jail / that escaped from the prisoner.
It was / the apple / that the princess ate.
The terrorism / shocked / the family / that read the magazine.
It was / the bomb / that killed the spy.
The professor / praised / the museum / that designed the architect.
The activity / entertained / the participants / that the lectures bored.
It was / the opera / that the composer excited.
It was / the leader / that protested against the war.
The lightning / preceded / the storm / that the explorers frightened.
It was / the house / that the carpenter built.
It was / the skier / that excited the snow.
It was / the policeman / that the report astonished.
It was / the price / that pleased the customer.
The captain / wrote / the letter / that the athlete surprised.
It was / the priest / that the bible dropped on the floor.
The victory / delighted / the children / that played the game.
The minister / supported / the issue / that the candidate worried.
It was / the music / that delighted the passenger.
The noise / ruined / the song / that wrote the musician.
It was / the comedy / that the audience entertained.
The thief / poisoned / the guard / that watched the diamond.
The weather / interrupted / the journey / that the sailors exhausted.
The editor / hit / the author / that requested the money.
The magazine / featured / the review / that the actor upset.
APPENDIX C

TEST SENTENCES USED IN EXPERIMENTS 1 AND 3

1a The doctor that the uncle of the bishop who injured himself last summer trusted was concerned about the infection.
1b The doctor said that the uncle of the bishop who injured himself last summer was concerned about the infection.

2a The babysitter that the sister of the schoolgirl who burned herself the other day adored was very nice.
2b The babysitter said that the sister of the schoolgirl who burned herself the other day was very nice.

3a The lawyer that the uncle of the councilman who treated himself after the accident consulted had watched the news.
3b The lawyer thought that the uncle of the councilman who treated himself after the accident had watched the news.

4a The journalist that the nephew of the congressman who calmed himself after the tragedy contacted was waiting for the doctor.
4b The journalist reported that the nephew of the congressman who calmed himself after the tragedy was waiting for the doctor.

5a The neighbor that the mother of the ballerina who found herself in a lot of trouble consulted phoned the police.
5b The neighbor said that the mother of the ballerina who found herself in a lot of trouble phoned the police.

6a The gardener that the grandmother of the seamstress who treated herself to an ice-cream cone disliked was sitting on the front porch.
6b The gardener said that the grandmother of the seamstress who treated herself to an ice-cream cone was sitting on the front porch.

7a The chef that the daughter of the waitress who enjoyed herself a lot admired was out on the balcony.
7b The chef thought that the daughter of the waitress who enjoyed herself a lot was out on the balcony.

8a The detective that the brother of the doorman who cut himself on the broken glass contacted was shocked by the accident.
8b The detective believed that the brother of the doorman who cut himself on the broken glass was shocked by the accident.
9a The director that the sister of the actress who introduced herself at the party admired was very intelligent.
9b The director noticed that the sister of the actress who introduced herself at the party was very intelligent.

10a The judge that the niece of the widow who defended herself from false accusations respected lived in a little village.
10b The judge knew that the niece of the widow who defended herself from false accusations lived in a little village.

11a The merchant that the nephew of the fisherman who taught himself how to surf disliked grew up on a small island.
11b The merchant knew that the nephew of the fisherman who taught himself how to surf grew up on a small island.

12a The tenant that the niece of the landlady who distinguished herself in the community despised went to a prestigious university.
12b The tenant heard that the niece of the landlady who distinguished herself in the community went to a prestigious university.

13a The knight that the aunt of the princess who amused herself at the party praised was terribly talkative.
13b The knight noticed that the aunt of the princess who amused herself at the party was terribly talkative.

14a The designer that the mother of the bride who embarrassed herself at the bridal show trusted was highly respected in the community.
14b The designer believed that the mother of the bride who embarrassed herself at the bridal show was highly respected in the community.

15a The agent that the father of the actor who hated himself for lying paid was known for shady deals.
15b The agent heard that the father of the actor who hated himself for lying was known for shady deals.

16a The writer that the son of the repairman who educated himself at night loved always worked late.
16b The writer noticed that the son of the repairman who educated himself at night always worked late.

17a The teacher that the brother of the schoolboy who prepared himself for the speech admired graduated from college with honors.
17b The teacher knew that the brother of the schoolboy who prepared himself for the speech graduated from college with honors.
18a The attorney that the aunt of the landlady who bankrupted herself for the second time respected was searching for a solution.
18b The attorney heard that the aunt of the landlady who bankrupted herself for the second time was searching for a solution.

19a The housekeeper that the nephew of the businessman who reclined himself on the couch praised was brought up in England.
19b The housekeeper thought that the nephew of the businessman who reclined himself on the couch was brought up in England.

20a The policeman that the sister of the lady who protected herself from an intruder approached stayed at the crime scene for a while.
20b The policeman noticed that the sister of the lady who protected herself from an intruder stayed at the crime scene for a while.

21a The servant that the grandfather of the emperor who hid himself from the public disliked was very self-involved and uncaring.
21b The servant thought that the grandfather of the emperor who hid himself from the public was very self-involved and uncaring.

22a The manager that the uncle of the salesman who invited himself to the party greeted was drinking lots of champagne.
22b The manager noticed that the uncle of the salesman who invited himself to the party was drinking lots of champagne.

23a The pilot that the mother of the stewardess who entertained herself most evenings met was a complete gentleman.
23b The pilot heard that the mother of the stewardess who entertained herself most evenings was a complete gentleman.

24a The physician that the son of the waiter who hurt himself with the knife called had gone out of business.
24b The physician heard that the son of the waiter who hurt himself with the knife had gone out of business.

25a The caterer that the aunt of the hostess who complimented herself too many times hired looked unhappy at the function.
25b The caterer noticed that the aunt of the hostess who complimented herself too many times looked unhappy at the function.

26a The reporter that the grandfather of the fireman who blamed himself for the accident despised did not tell the truth.
26b The reporter said that the grandfather of the fireman who blamed himself for the accident did not tell the truth.
27a The governor that the daughter of the nanny who served herself at the banquet flattered was very pleased with the food.
27b The governor noticed that the daughter of the nanny who served herself at the banquet was very pleased with the food.

28a The mayor that the father of the mailman who devoted himself to political campaigns supported was very popular among the local residents.
28b The mayor heard that the father of the mailman who devoted himself to political campaigns was very popular among the local residents.

29a The nurse that the niece of the maid who rescued herself from the burning house appreciated was very calm and patient.
29b The nurse thought that the niece of the maid who rescued herself from the burning house was very calm and patient.

30a The janitor that the grandfather of the foreman who relaxed himself by reading a book complimented was very picky about hygiene.
30b The janitor heard that the grandfather of the foreman who relaxed himself by reading a book was very picky about hygiene.

31a The professor that the daughter of the midwife who worried herself about tuition worshiped was trying to find financial support.
31b The professor believed that the daughter of the midwife who worried herself about tuition was trying to find financial support.

32a The accountant that the father of the craftsman who saved himself from financial crises consulted was extremely supportive.
32b The accountant said that the father of the craftsman who saved himself from financial crises was extremely supportive.
APPENDIX D

THE CLOZE TEST USED IN THE PRESENT STUDY

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

NOTE: Spelling will not count against you as long as the scorer can read the word.
EXAMPLE: The boy walked up the street. He stepped on a piece of ice.
He fell (1) down but he didn’t hurt himself.

MAN AND HIS PROGRESS

Man is the only living creature that can make and use tools. He is the most teachable of living beings, earning the name of Homo sapiens. (1) ever restless brain has used the and the wisdom of his ancestors (3) improve his way of life. Since (4) is able to walk and run (5) his feet, his hands have always (6) free to carry and to use (7) . Man’s hands have served him well (8) his life on earth. His development, (9) can be divided into three major (10) , is marked by several different ways (11) life.

Up to 10,000 years ago, (12) human beings lived by hunting and (13) . They also picked berries and fruits, (14) dug for various edible roots. Most (15) , the men were the hunters, and (16) women acted as food gatherers. Since (17) women were busy with the children, (18) men handled the tools. In a (19) hand, a dead branch became a (20) to knock down fruit or to (21) for tasty roots. Sometimes, an animal (22) served as a club, and a (23) piece of stone, fitting comfortably into (24) hand, could be used to break (25) or to throw at an animal. (26) stone was chipped against another until (27) had a sharp edge. The primitive (28) who first thought of putting a (29) stone at the end of a (30) made a brilliant discovery: he (31) joined two things to make a (32) useful tool, the spear. Flint, found (33)
many rocks, became a common cutting in the Paleolithic period of man's. Since no wood or bone tools survived, we know of this man his stone implements, with which he kill animals, cut up the meat, scrape the skins, as well as pictures on the walls of the where he lived during the winter. the warmer seasons, man wandered on steppes of Europe without a fixed, always foraging for food. Perhaps the carried nuts and berries in shells skins or even in light, woven. Wherever they camped, the primitive people fires by striking flint for sparks using dried seeds, moss, and rotten for tinder. With fires that he kindled himself, man could keep wild animals away and could cook those that he killed, as well as provide warmth and light for himself.

Answer keys

<table>
<thead>
<tr>
<th>Blank</th>
<th>Exact</th>
<th>Acceptable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>his</td>
<td>man's, our, the</td>
</tr>
<tr>
<td>2</td>
<td>knowledge</td>
<td>accomplishments, culture, cunning, examples, experience(s), hands, ideas, information, ingenuity, instinct, intelligence, mistakes, nature, power, skill(s), talent, teaching, technique, thought, will, wit, words, work</td>
</tr>
<tr>
<td>3</td>
<td>to</td>
<td>upon, using, with</td>
</tr>
<tr>
<td>4</td>
<td>man</td>
<td>he</td>
</tr>
<tr>
<td>5</td>
<td>on</td>
<td>felt, hung, remained</td>
</tr>
<tr>
<td>6</td>
<td>been</td>
<td>adequately, carefully, conventionally, creatively, diligently, efficiently, freely, implements, objects, productively, readily, them, things, weapons</td>
</tr>
<tr>
<td>7</td>
<td>tools</td>
<td>all, for, improving, in, through, throughout, with</td>
</tr>
<tr>
<td>8</td>
<td>during</td>
<td>also, basically, conveniently, easily, historically, however, often, since, that, thus</td>
</tr>
<tr>
<td>9</td>
<td>which</td>
<td>areas, categories, divisions, eras, facets, groups, parts, phases, sections, stages, steps, topics, trends</td>
</tr>
<tr>
<td>10</td>
<td>periods</td>
<td>for, in, through, towards</td>
</tr>
<tr>
<td>11</td>
<td>of</td>
<td>early, hungry, many, most, only, primitive, the, these</td>
</tr>
<tr>
<td>12</td>
<td>all</td>
<td>fishing</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>farming, foraging, gathering, killing, scavenging, scrounging, sleeping, trapping</td>
</tr>
<tr>
<td>14</td>
<td>and</td>
<td>often, ravenously, some, the</td>
</tr>
<tr>
<td>15</td>
<td>often</td>
<td>always, emphatically, important, nights, normally, of, times, trips</td>
</tr>
<tr>
<td>16</td>
<td>the</td>
<td>all, house, many, most, older, their, younger</td>
</tr>
<tr>
<td>17</td>
<td>the</td>
<td>all, many, married, most, often, older, primate, these</td>
</tr>
<tr>
<td>18</td>
<td>the</td>
<td>all, constructive, many, most, older, primate, tough, younger</td>
</tr>
</tbody>
</table>
### APPENDIX E

READING SPAN SCORES FROM ALL PARTICIPANTS

<table>
<thead>
<tr>
<th>Subject</th>
<th>Experiment 1 L1 English reading span (Max = 70)</th>
<th>Experiment 2 L1 English reading span (Max = 70)</th>
<th>Experiments 3 and 4 Subject L2 English reading span (Max = 70)</th>
<th>L1 Japanese reading span (Max = 70)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>28</td>
<td>E29</td>
<td>J1</td>
<td>49</td>
</tr>
<tr>
<td>E2</td>
<td>31</td>
<td>E30</td>
<td>J2</td>
<td>43</td>
</tr>
<tr>
<td>E3</td>
<td>36</td>
<td>E31</td>
<td>J3</td>
<td>26</td>
</tr>
<tr>
<td>E4</td>
<td>37</td>
<td>E32</td>
<td>J4</td>
<td>38</td>
</tr>
<tr>
<td>E5</td>
<td>38</td>
<td>E33</td>
<td>J5</td>
<td>35</td>
</tr>
<tr>
<td>E6</td>
<td>38</td>
<td>E34</td>
<td>J6</td>
<td>37</td>
</tr>
<tr>
<td>E7</td>
<td>42</td>
<td>E35</td>
<td>J7</td>
<td>41</td>
</tr>
<tr>
<td>E8</td>
<td>43</td>
<td>E36</td>
<td>J8</td>
<td>34</td>
</tr>
<tr>
<td>E9</td>
<td>46</td>
<td>E37</td>
<td>J9</td>
<td>44</td>
</tr>
<tr>
<td>E10</td>
<td>47</td>
<td>E38</td>
<td>J10</td>
<td>54</td>
</tr>
<tr>
<td>E11</td>
<td>47</td>
<td>E39</td>
<td>J11</td>
<td>66</td>
</tr>
<tr>
<td>E12</td>
<td>47</td>
<td>E40</td>
<td>J12</td>
<td>43</td>
</tr>
<tr>
<td>E13</td>
<td>49</td>
<td>E41</td>
<td>J13</td>
<td>56</td>
</tr>
<tr>
<td>E14</td>
<td>50</td>
<td>E42</td>
<td>J14</td>
<td>32</td>
</tr>
<tr>
<td>E15</td>
<td>51</td>
<td>E43</td>
<td>J15</td>
<td>61</td>
</tr>
<tr>
<td>E16</td>
<td>51</td>
<td>E44</td>
<td>J16</td>
<td>41</td>
</tr>
<tr>
<td>E17</td>
<td>53</td>
<td>E45</td>
<td>J17</td>
<td>35</td>
</tr>
<tr>
<td>E18</td>
<td>54</td>
<td>E46</td>
<td>J18</td>
<td>45</td>
</tr>
<tr>
<td>E19</td>
<td>55</td>
<td>E47</td>
<td>J19</td>
<td>41</td>
</tr>
<tr>
<td>E20</td>
<td>56</td>
<td>E48</td>
<td>J20</td>
<td>41</td>
</tr>
<tr>
<td>E21</td>
<td>56</td>
<td>E49</td>
<td>J21</td>
<td>26</td>
</tr>
<tr>
<td>E22</td>
<td>56</td>
<td>E50</td>
<td>J22</td>
<td>36</td>
</tr>
<tr>
<td>E23</td>
<td>57</td>
<td>E51</td>
<td>J23</td>
<td>36</td>
</tr>
<tr>
<td>E24</td>
<td>58</td>
<td>E52</td>
<td>J24</td>
<td>23</td>
</tr>
<tr>
<td>E25</td>
<td>58</td>
<td>E53</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>E26</td>
<td>58</td>
<td>E54</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>E27</td>
<td>64</td>
<td>E55</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>E28</td>
<td>66</td>
<td>E56</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E57</td>
<td></td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E58</td>
<td></td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E59</td>
<td></td>
<td>65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E60</td>
<td></td>
<td>65</td>
</tr>
</tbody>
</table>

| Mean    | 49.00                                         | 46.38                                         | 40.96                                          | 47.04                              |
| SD      | 9.58                                          | 12.16                                         | 10.56                                          | 8.00                               |

158
APPENDIX F

ANOVA TABLE FROM EXPERIMENT 1

Tests of Within-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>complexity</td>
<td>Sphericity Assumed</td>
<td>1.00</td>
<td>185.95</td>
<td>2.382</td>
<td>0.135</td>
<td>0.084</td>
<td>0.318</td>
</tr>
<tr>
<td>complexity * span level</td>
<td>Sphericity Assumed</td>
<td>1.00</td>
<td>18.544</td>
<td>0.238</td>
<td>0.630</td>
<td>0.009</td>
<td>0.076</td>
</tr>
<tr>
<td>Error (complexity)</td>
<td>Assumed</td>
<td>26.00</td>
<td>78.056</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tests of Between-Subjects Effects

Transformed Variable: Average

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>87578.354</td>
<td>1.00</td>
<td>87578.35</td>
<td>79.433</td>
<td>0.000</td>
<td>0.753</td>
<td>1.000</td>
</tr>
<tr>
<td>SPANLEV</td>
<td>7131.925</td>
<td>1.00</td>
<td>7131.92</td>
<td>6.469</td>
<td>0.017</td>
<td>0.199</td>
<td>0.687</td>
</tr>
<tr>
<td>Error</td>
<td>28666.066</td>
<td>26.00</td>
<td>1102.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX G

TEST SENTENCES USED IN EXPERIMENTS 2 AND 5

1a The doctor that the sister of the bishop who injured himself last summer trusted was concerned about the infection
1b The doctor that the sister of the bishop who injured herself last summer trusted was concerned about the infection
1c The doctor said that the sister of the bishop who injured himself last summer was concerned about the infection
1d The doctor said that the sister of the bishop who injured herself last summer was concerned about the infection

2a The babysitter that the brother of the schoolgirl who burned herself the other day adored was very nice.
2b The babysitter that the brother of the schoolgirl who burned himself the other day adored was very nice.
2c The babysitter said that the brother of the schoolgirl who burned herself the other day was very nice
2d The babysitter said that the brother of the schoolgirl who burned himself the other day was very nice

3a The lawyer that the aunt of the councilman who treated himself after the accident consulted was watching the news.
3b The lawyer that the aunt of the councilman who treated herself after the accident consulted was watching the news.
3c The lawyer thought that the aunt of the councilman who treated himself after the accident was watching the news
3d The lawyer thought that the aunt of the councilman who treated herself after the accident was watching the news

4a The journalist that the niece of the congressman who calmed himself after the tragedy contacted was waiting for the doctor.
4b The journalist that the niece of the congressman who calmed herself after the tragedy contacted was waiting for the doctor.
4c The journalist reported that the niece of the congressman who calmed himself after the tragedy was waiting for the doctor
4d The journalist reported that the niece of the congressman who calmed herself after the tragedy was waiting for the doctor

5a The neighbor that the uncle of the ballerina who found herself in a lot of trouble consulted was a big gossiper.
5b The neighbor that the uncle of the ballerina who found himself in a lot of trouble consulted was a big gossiper.
5c The neighbor said that the uncle of the ballerina who found herself in a lot of trouble was a big gossiper.
5d The neighbor said that the uncle of the ballerina who found himself in a lot of trouble was a big gossiper.

6a The gardener that the brother of the seamstress who treated herself to an ice-cream cone disliked was waiting on the porch.
6b The gardener that the brother of the seamstress who treated himself to an ice-cream cone disliked was waiting on the porch.
6c The gardener thought that the brother of the seamstress who treated herself to an ice-cream cone was waiting on the porch.
6d The gardener thought that the brother of the seamstress who treated himself to an ice-cream cone was waiting on the porch.

7a The chef that the husband of the waitress who enjoyed herself a lot admired was out on the balcony.
7b The chef that the husband of the waitress who enjoyed himself a lot admired was out on the balcony.
7c The chef thought that the husband of the waitress who enjoyed herself a lot was out on the balcony.
7d The chef thought that the husband of the waitress who enjoyed himself a lot was out on the balcony.

8a The detective that the mother of the doorman who cut himself on the broken glass contacted was shocked by the accident.
8b The detective that the mother of the doorman who cut herself on the broken glass contacted was shocked by the accident.
8c The detective believed that the mother of the doorman who cut himself on the broken glass was shocked by the accident.
8d The detective believed that the mother of the doorman who cut herself on the broken glass was shocked by the accident.

9a The director that the husband of the actress who introduced herself at the party admired was very intelligent.
9b The director that the husband of the actress who introduced himself at the party admired was very intelligent.
9c The director noticed that the husband of the actress who introduced herself at the party was very intelligent.
9d The director noticed that the husband of the actress who introduced himself at the party was very intelligent.

10a The judge that the brother of the widow who defended herself from false accusations respected was from a little village.
10b The judge that the brother of the widow who defended himself from false accusations respected was from a little village.
10c The judge knew that the brother of the widow who defended herself from false accusations was from a little village.
10d The judge knew that the brother of the widow who defended himself from false accusations was from a little village.
11a The merchant that the daughter of the fisherman who taught himself how to surf disliked was brought up on an island.
11b The merchant that the daughter of the fisherman who taught herself how to surf disliked was brought up on an island.
11c The merchant knew that the daughter of the fisherman who taught himself how to surf was brought up on an island.
11d The merchant knew that the daughter of the fisherman who taught herself how to surf was brought up on an island.

12a The tenant that the nephew of the landlady who distinguished herself in the community despised was a graduate of MIT.
12b The tenant that the nephew of the landlady who distinguished himself in the community despised was a graduate of MIT.
12c The tenant heard that the nephew of the landlady who distinguished herself in the community was a graduate of MIT.
12d The tenant heard that the nephew of the landlady who distinguished himself in the community was a graduate of MIT.

13a The knight that the uncle of the princess who amused herself at the party praised was terribly talkative.
13b The knight that the uncle of the princess who amused himself at the party praised was terribly talkative.
13c The knight noticed that the uncle of the princess who amused herself at the party was terribly talkative.
13d The knight noticed that the uncle of the princess who amused himself at the party was terribly talkative.

14a The designer that the father of the bride who embarrassed herself at the bridal show trusted was respected in the community.
14b The designer that the father of the bride who embarrassed himself at the bridal show trusted was respected in the community.
14c The designer believed that the father of the bride who embarrassed herself at the bridal show was respected in the community.
14d The designer believed that the father of the bride who embarrassed himself at the bridal show was respected in the community.

15a The agent that the aunt of the actor who hated himself for lying paid was known for shady deals.
15b The agent that the aunt of the actor who hated herself for lying paid was known for shady deals.
15c The agent heard that the aunt of the actor who hated himself for lying was known for shady deals.
15d The agent heard that the aunt of the actor who hated herself for lying was known for shady deals.
16a The writer that the daughter of the repairman who educated himself at night loved was always working till late.
16b The writer that the daughter of the repairman who educated herself at night loved was always working till late.
16c The writer heard that the daughter of the repairman who educated himself at night was always working till late.
16d The writer heard that the daughter of the repairman who educated herself at night was always working till late.

17a The teacher that the sister of the schoolboy who prepared himself for the speech admired was a graduate of Harvard.
17b The teacher that the sister of the schoolboy who prepared herself for the speech admired was a graduate of Harvard.
17c The teacher knew that the sister of the schoolboy who prepared himself for the speech was a graduate of Harvard.
17d The teacher knew that the sister of the schoolboy who prepared herself for the speech was a graduate of Harvard.

18a The attorney that the son of the landlady who bankrupted herself for the second time respected was searching for a solution.
18b The attorney that the son of the landlady who bankrupted himself for the second time respected was searching for a solution.
18c The attorney heard that the son of the landlady who bankrupted herself for the second time was searching for a solution.
18d The attorney heard that the son of the landlady who bankrupted himself for the second time was searching for a solution.

19a The housekeeper that the niece of the businessman who reclined himself on the couch praised was brought up in England.
19b The housekeeper that the niece of the businessman who reclined herself on the couch praised was brought up in England.
19c The housekeeper thought that the niece of the businessman who reclined himself on the couch was brought up in England.
19d The housekeeper thought that the niece of the businessman who reclined herself on the couch was brought up in England.

20a The policeman that the nephew of the lady who protected herself from an intruder approached was observing the crime scene.
20b The policeman that the nephew of the lady who protected himself from an intruder approached was observing the crime scene.
20c The policeman noticed that the nephew of the lady who protected herself from an intruder was observing the crime scene.
20d The policeman noticed that the nephew of the lady who protected himself from an intruder was observing the crime scene.
21a The servant that the grandmother of the emperor who hid himself from the public disliked was very self-involved and uncaring.
21b The servant that the grandmother of the emperor who hid herself from the public disliked was very self-involved and uncaring.
21c The servant thought that the grandmother of the emperor who hid himself from the public was very self-involved and uncaring.
21d The servant thought that the grandmother of the emperor who hid herself from the public was very self-involved and uncaring.

22a The manager that the wife of the salesman who invited himself to the party greeted was drinking lots of champagne.
22b The manager that the wife of the salesman who invited herself to the party greeted was drinking lots of champagne.
22c The manager noticed that the wife of the salesman who invited himself to the party was drinking lots of champagne.
22d The manager noticed that the wife of the salesman who invited herself to the party was drinking lots of champagne.

23a The pilot that the grandfather of the stewardess who entertained herself most evenings met was a complete gentleman.
23b The pilot that the grandfather of the stewardess who entertained himself most evenings met was a complete gentleman.
23c The pilot heard that the grandfather of the stewardess who entertained herself most evenings was a complete gentleman.
23d The pilot heard that the grandfather of the stewardess who entertained himself most evenings was a complete gentleman.

24a The physician that the mother of the waiter who hurt himself with the knife called was almost ready for retirement.
24b The physician that the mother of the waiter who hurt herself with the knife called was almost ready for retirement.
24c The physician heard that the mother of the waiter who hurt himself with the knife was almost ready for retirement.
24d The physician heard that the mother of the waiter who hurt herself with the knife was almost ready for retirement.

25a The caterer that the husband of the hostess who complimented herself too many times hired was feeling unhappy at the function.
25b The caterer that the husband of the hostess who complimented himself too many times hired was feeling unhappy at the function.
25c The caterer noticed that the husband of the hostess who complimented herself too many times was feeling unhappy at the function.
25d The caterer noticed that the husband of the hostess who complimented himself too many times was feeling unhappy at the function.

26a The reporter that the wife of the fireman who blamed himself for the accident despised was not telling the truth.
26b The reporter that the wife of the fireman who blamed herself for the accident despised was not telling the truth.

26c The reporter said that the wife of the fireman who blamed himself for the accident was not telling the truth.

26d The reporter said that the wife of the fireman who blamed herself for the accident was not telling the truth.

27a The governor that the uncle of the nanny who served herself at the banquet flattered was very pleased with the food.

27b The governor that the uncle of the nanny who served himself at the banquet flattered was very pleased with the food.

27c The governor noticed that the uncle of the nanny who served herself at the banquet was very pleased with the food.

27d The governor noticed that the uncle of the nanny who served himself at the banquet was very pleased with the food.

28a The mayor that the sister of the mailman who devoted himself to political campaigns supported was very popular among local residents.

28b The mayor that the sister of the mailman who devoted herself to political campaigns supported was very popular among local residents.

28c The mayor heard that the sister of the mailman who devoted himself to political campaigns was very popular among local residents.

28d The mayor heard that the sister of the mailman who devoted herself to political campaigns was very popular among local residents.

29a The nurse that the son of the maid who rescued herself from the burning house appreciated was very calm and patient.

29b The nurse that the son of the maid who rescued himself from the burning house appreciated was very calm and patient.

29c The nurse thought that the son of the maid who rescued herself from the burning house was very calm and patient.

29d The nurse thought that the son of the maid who rescued himself from the burning house was very calm and patient.

30a The janitor that the grandmother of the foreman who relaxed himself by reading a book complimented was very picky about hygiene.

30b The janitor that the grandmother of the foreman who relaxed herself by reading a book complimented was very picky about hygiene.

30c The janitor heard that the grandmother of the foreman who relaxed himself by reading a book was very picky about hygiene.

30d The janitor heard that the grandmother of the foreman who relaxed herself by reading a book was very picky about hygiene.

31a The professor that the son of the midwife who worried herself about tuition worshiped was trying to find financial support.

31b The professor that the son of the midwife who worried himself about tuition worshiped was trying to find financial support.
31c  The professor believed that the son of the midwife who worried herself about tuition was trying to find financial support.
31d  The professor believed that the son of the midwife who worried himself about tuition was trying to find financial support.

32a  The accountant that the niece of the craftsman who saved himself from financial crises consulted was extremely supportive.
32b  The accountant that the niece of the craftsman who saved herself from financial crises consulted was extremely supportive.
32c  The accountant said that the niece of the craftsman who saved himself from financial crises was extremely supportive.
32d  The accountant said that the niece of the craftsman who saved herself from financial crises was extremely supportive.
APPENDIX H
ANOVA TABLES FROM EXPERIMENT 2

**Comprehension accuracy data**

Tests of Within-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Mean Eta Squared</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity</td>
<td>Assumed Sphericity</td>
<td>0.088</td>
<td>1.0</td>
<td>0.088</td>
<td>2.823</td>
<td>0.103</td>
<td>0.086</td>
</tr>
<tr>
<td>Complexity * reading span Error (Complexity)</td>
<td>Assumed Sphericity</td>
<td>0.115</td>
<td>1.0</td>
<td>0.115</td>
<td>3.705</td>
<td>0.064</td>
<td>0.110</td>
</tr>
<tr>
<td>Attachment</td>
<td>Assumed Sphericity</td>
<td>0.934</td>
<td>30.0</td>
<td>0.031</td>
<td>0.182</td>
<td>0.673</td>
<td>0.006</td>
</tr>
<tr>
<td>Attachment * reading span Error (attachment)</td>
<td>Assumed Sphericity</td>
<td>0.001</td>
<td>1.0</td>
<td>0.001</td>
<td>0.052</td>
<td>0.822</td>
<td>0.002</td>
</tr>
<tr>
<td>Complexity * Attachment</td>
<td>Assumed Sphericity</td>
<td>0.278</td>
<td>1.0</td>
<td>0.278</td>
<td>13.569</td>
<td>0.001</td>
<td>0.311</td>
</tr>
<tr>
<td>Complexity * Attachment * Reading span Error (Complexity * Attachment)</td>
<td>Assumed Sphericity</td>
<td>0.197</td>
<td>1.0</td>
<td>0.197</td>
<td>9.631</td>
<td>0.004</td>
<td>0.243</td>
</tr>
</tbody>
</table>

Tests of Between-Subjects Effects

Transformed Variable: Average

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>61.221</td>
<td>1.0</td>
<td>61.221</td>
<td>8</td>
<td>0.000</td>
<td>0.969</td>
<td>1.000</td>
</tr>
<tr>
<td>RST</td>
<td>0.505</td>
<td>1.0</td>
<td>0.505</td>
<td>7.811</td>
<td>0.009</td>
<td>0.207</td>
<td>0.772</td>
</tr>
<tr>
<td>Error</td>
<td>1.940</td>
<td>30.0</td>
<td>0.065</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Reading time in Region 9

Tests of Within-Subjects Effects (Region 9)

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPLX</td>
<td>Sphericity Assumed</td>
<td>9698.06</td>
<td>1.0</td>
<td>9698.06</td>
<td>0.104</td>
<td>0.749</td>
<td>0.003</td>
<td>0.061</td>
</tr>
<tr>
<td>Error(COMP LX)</td>
<td>Sphericity Assumed</td>
<td>2883536.76</td>
<td>31.0</td>
<td>93017.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATTACH</td>
<td>Sphericity Assumed</td>
<td>179793.50</td>
<td>1.0</td>
<td>179793.50</td>
<td>4.403</td>
<td>0.044</td>
<td>0.124</td>
<td>0.529</td>
</tr>
<tr>
<td>Error(ATTACH H)</td>
<td>Sphericity Assumed</td>
<td>1265736.53</td>
<td>31.0</td>
<td>40830.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMPLX* ATTACH</td>
<td>Sphericity Assumed</td>
<td>41424.75</td>
<td>1.0</td>
<td>41424.75</td>
<td>0.415</td>
<td>0.524</td>
<td>0.013</td>
<td>0.096</td>
</tr>
<tr>
<td>Error(COMP LX*ATTACH H)</td>
<td>Sphericity Assumed</td>
<td>3094121.50</td>
<td>31.0</td>
<td>99810.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tests of Within-Subjects Effects (EC condition only)

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATTACH</td>
<td>Sphericity Assumed</td>
<td>24307.90</td>
<td>1.0</td>
<td>24307.90</td>
<td>0.411</td>
<td>0.526</td>
<td>0.013</td>
<td>0.095</td>
</tr>
<tr>
<td>Error(ATTACH H)</td>
<td>Sphericity Assumed</td>
<td>1834304.16</td>
<td>31.0</td>
<td>59171.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tests of Within-Subjects Effects (SC condition only)

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATTACH</td>
<td>Sphericity Assumed</td>
<td>196910.35</td>
<td>1.0</td>
<td>196910.35</td>
<td>2.417</td>
<td>0.130</td>
<td>0.072</td>
<td>0.325</td>
</tr>
<tr>
<td>Error(ATTACH H)</td>
<td>Sphericity Assumed</td>
<td>2525553.87</td>
<td>7</td>
<td>81469.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Reading time in Region 12

Tests of Within-Subjects Effects (Region 12)

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity Sphericity Assumed</td>
<td>1206197.70</td>
<td>1.0</td>
<td>1206197.70</td>
<td>18.40</td>
<td>0.00</td>
<td>0.37</td>
<td>0.99</td>
</tr>
<tr>
<td>Error (Complexity) Sphericity Assumed</td>
<td>2032022.11</td>
<td>31.0</td>
<td>65549.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attachment Sphericity Assumed</td>
<td>3377.29</td>
<td>1.0</td>
<td>3377.29</td>
<td>0.18</td>
<td>0.68</td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>Error (Attachment) Sphericity Assumed</td>
<td>592064.19</td>
<td>31.0</td>
<td>19098.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complexity * Sphericity Assumed</td>
<td>50151.21</td>
<td>1.0</td>
<td>50151.21</td>
<td>2.08</td>
<td>0.16</td>
<td>0.06</td>
<td>0.29</td>
</tr>
<tr>
<td>Error (Complexity * Attachment) Sphericity Assumed</td>
<td>745687.92</td>
<td>31.0</td>
<td>24054.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX I

BACKGROUND QUESTIONNAIRE USED FOR THE JAPANESE SUBJECTS

This survey is used to collect information about your language learning background. The information you provide here will only be used for research purposes and will not be disclosed to anybody but the researcher himself. Please fill out your answers in the given space, or circle the appropriate answer if there are options to choose from.

1. Your home city, state/country

2. List all of your second languages

3. Length of exposure to English Age of first exposure [ ] Length

4. If you have lived and/or studied in countries where English is spoken as a primary language, please specify how long you were there and supply the total length of residence in English-speaking countries.

   How long (in total) you stayed there [ ]

5. How do you self-rate overall proficiency of your English on a 5-point scale below?

   1 Beginning 2 low intermediate 3 intermediate 4 advanced 5 near-native


   I speak my first language [ ] % of the day.
   I speak my second language [ ] % of the day

7. If you have taken a TOEFL test, please specify your most recent TOEFL score as well as when you took one.

   When ( ) Score ( ) (CBT/Paper both OK)

Thank you very much for your cooperation!
## APPENDIX J

### RELEVANT BACKGROUND INFORMATION OF THE JAPANESE SUBJECTS

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age at the time of the experiment</th>
<th>Age of first exposure to English</th>
<th>Length of residence in English-speaking environments (in years)</th>
<th>Cloze test (Max = 50)</th>
<th>English reading span (Max = 70)</th>
<th>Japanese reading span (Max = 70)</th>
</tr>
</thead>
<tbody>
<tr>
<td>J1</td>
<td>34</td>
<td>8</td>
<td>16</td>
<td>47</td>
<td>49</td>
<td>52</td>
</tr>
<tr>
<td>J2</td>
<td>24</td>
<td>6</td>
<td>9.5</td>
<td>46</td>
<td>43</td>
<td>54</td>
</tr>
<tr>
<td>J3</td>
<td>25</td>
<td>10</td>
<td>2</td>
<td>45</td>
<td>35</td>
<td>50</td>
</tr>
<tr>
<td>J4</td>
<td>26</td>
<td>10</td>
<td>8</td>
<td>45</td>
<td>26</td>
<td>34</td>
</tr>
<tr>
<td>J5</td>
<td>31</td>
<td>12</td>
<td>6.5</td>
<td>45</td>
<td>38</td>
<td>51</td>
</tr>
<tr>
<td>J6</td>
<td>32</td>
<td>12</td>
<td>2.5</td>
<td>44</td>
<td>37</td>
<td>38</td>
</tr>
<tr>
<td>J7</td>
<td>37</td>
<td>12</td>
<td>2.5</td>
<td>44</td>
<td>41</td>
<td>62</td>
</tr>
<tr>
<td>J8</td>
<td>20</td>
<td>12</td>
<td>8</td>
<td>43</td>
<td>34</td>
<td>46</td>
</tr>
<tr>
<td>J9</td>
<td>28</td>
<td>12</td>
<td>3</td>
<td>42</td>
<td>66</td>
<td>61</td>
</tr>
<tr>
<td>J10</td>
<td>31</td>
<td>10</td>
<td>2.5</td>
<td>42</td>
<td>44</td>
<td>53</td>
</tr>
<tr>
<td>J11</td>
<td>27</td>
<td>12</td>
<td>5</td>
<td>42</td>
<td>54</td>
<td>44</td>
</tr>
<tr>
<td>J12</td>
<td>35</td>
<td>8</td>
<td>3</td>
<td>40</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>J13</td>
<td>25</td>
<td>12</td>
<td>2.5</td>
<td>40</td>
<td>56</td>
<td>52</td>
</tr>
<tr>
<td>J14</td>
<td>42</td>
<td>10</td>
<td>15</td>
<td>39</td>
<td>32</td>
<td>50</td>
</tr>
<tr>
<td>J15</td>
<td>26</td>
<td>13</td>
<td>3.5</td>
<td>39</td>
<td>61</td>
<td>57</td>
</tr>
<tr>
<td>J16</td>
<td>28</td>
<td>12</td>
<td>6</td>
<td>37</td>
<td>41</td>
<td>44</td>
</tr>
<tr>
<td>J17</td>
<td>30</td>
<td>10</td>
<td>2.5</td>
<td>37</td>
<td>45</td>
<td>46</td>
</tr>
<tr>
<td>J18</td>
<td>36</td>
<td>13</td>
<td>7.5</td>
<td>37</td>
<td>35</td>
<td>43</td>
</tr>
<tr>
<td>J19</td>
<td>44</td>
<td>12</td>
<td>14</td>
<td>36</td>
<td>41</td>
<td>40</td>
</tr>
<tr>
<td>J20</td>
<td>23</td>
<td>13</td>
<td>4.5</td>
<td>35</td>
<td>41</td>
<td>51</td>
</tr>
<tr>
<td>J21</td>
<td>30</td>
<td>13</td>
<td>1</td>
<td>35</td>
<td>26</td>
<td>35</td>
</tr>
<tr>
<td>J22</td>
<td>33</td>
<td>12</td>
<td>1.5</td>
<td>34</td>
<td>36</td>
<td>41</td>
</tr>
<tr>
<td>J23</td>
<td>31</td>
<td>12</td>
<td>3</td>
<td>34</td>
<td>23</td>
<td>32</td>
</tr>
<tr>
<td>J24</td>
<td>22</td>
<td>12</td>
<td>4</td>
<td>34</td>
<td>36</td>
<td>50</td>
</tr>
</tbody>
</table>

| Mean    | 30.00                           | 11.17                           | 5.56                                                          | 40.08                 | 40.96                           | 47.04                           |
| SD      | 6.01                            | 1.81                            | 4.30                                                          | 4.25                  | 10.56                           | 8.00                            |
## APPENDIX K

### ANOVA TABLES FROM EXPERIMENT 3

#### English reading span and complexity as factors

**Tests of Within-Subjects Effects**

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPLX</td>
<td>Sphericity Assumed</td>
<td>0.016</td>
<td>1</td>
<td>0.016</td>
<td>3.525</td>
<td>0.074</td>
</tr>
<tr>
<td>COMPLX * SPAN</td>
<td>Sphericity Assumed</td>
<td>0.016</td>
<td>1</td>
<td>0.016</td>
<td>3.525</td>
<td>0.074</td>
</tr>
<tr>
<td>Error (COMP LX)</td>
<td>Sphericity Assumed</td>
<td>0.097</td>
<td>22</td>
<td>0.004</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Tests of Between-Subjects Effects**

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>11.709</td>
<td>1</td>
<td>11.709</td>
<td>62.49</td>
<td>0.000</td>
<td>0.740</td>
</tr>
<tr>
<td>SPAN</td>
<td>0.501</td>
<td>1</td>
<td>0.501</td>
<td>2.675</td>
<td>0.116</td>
<td>0.108</td>
</tr>
<tr>
<td>Error</td>
<td>4.122</td>
<td>22</td>
<td>0.187</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Japanese reading span and complexity as factors

**Tests of Within-Subjects Effects**

Measure: MEASURE_1

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPLX</td>
<td>Sphericity Assumed</td>
<td>0.018</td>
<td>1</td>
<td>0.018</td>
<td>3.669</td>
<td>0.069</td>
</tr>
<tr>
<td>COMPLX * JSPAN</td>
<td>Sphericity Assumed</td>
<td>0.002</td>
<td>1</td>
<td>0.002</td>
<td>0.480</td>
<td>0.496</td>
</tr>
<tr>
<td>Error (COMP LX)</td>
<td>Sphericity Assumed</td>
<td>0.110</td>
<td>22</td>
<td>0.005</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Tests of Between-Subjects Effects**

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>11.38801</td>
<td>1</td>
<td>11.388</td>
<td>54.25</td>
<td>0.000</td>
<td>0.712</td>
</tr>
<tr>
<td>JSPAN</td>
<td>0.006075</td>
<td>1</td>
<td>0.006</td>
<td>0.029</td>
<td>0.866</td>
<td>0.001</td>
</tr>
<tr>
<td>Error</td>
<td>4.617517</td>
<td>22</td>
<td>0.210</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX L

TEST SENTENCES USED IN EXPERIMENT 4

1a 看護婦が親族が医者が昨夏けがをした牧師の叔父に信用されていたと言ったと主張した。
1b 医者が昨夏けがをした牧師の叔父に信用されていたと親族が言ったと看護婦が主張した。

2a 母親が息子がベビーシッターが先日やけどをした少女の妹に好かれていたと気づいたと立っていた。
2b ベビーシッターが先日やけどをした少女の妹に好かれていたと息子が気づくと母親が立っていた。

3a 記者が秘書が弁護士が事故後回復に努めた議員の叔父に相談をされたと答えたと報道した。
3b 弁護士が事故後回復に努めた議員の叔父に相談をされたと秘書が答えたと記者が報道した。

4a 弁護士が手伝いさんがジャーナリストが悲劇の後だいぶ落ち着いた国会議員の甥から電話を受けたと話ししていたと聞き分けた。
4b ジャーナリストが悲劇の後だいぶ落ち着いた国会議員の甥から電話を受けたと手伝いさんが話していたと弁護士が聞き分けた。

5a 探偵が親戚がお隣さんが災難に見舞われたバレリーナの母親に相談をされたと話していたと突き止めた。
5b お隣さんが災難に見舞われたバレリーナの母親に相談をされたと親戚が話していたと探偵が突き止めた。

6a お隣さんが弟が庭師がアイスクリームを好んで食べた裁縫職人の祖母に嫌われていたと思っていたと言って。
6b 庭師がアイスクリームを好んで食べた裁縫職人の祖母に嫌われていたと弟子が言っていたとお隣さんが言った。

7a オーナーが客さんがシェフが楽しそうにしていたウェイトレスの娘さんに尊敬されていたと噂していたと聞いた。
7b シェフが楽しそうにしていたウェイトレスの娘さんに尊敬されていたとお客さんが尊敬していたとオーナーが聞いた。

8a 警官が強盗がガラスで怪我をした従業員の兄から連絡を受けたと気づいたと報告していた。
8b 刑事がガラスで怪我をした従業員の兄から連絡を受けたと強盗が気づいたと警官が報告していた。
9a 社長がマネージャーが映画監督がパーティーで自己紹介していた女優の姉に賞賛されていたと勘違いしたと思っていた。
9b 映画監督がパーティーで自己紹介していた女優の姉に賞賛されていたとマネージャーが勘違いしたと社長が思っていた。
10a 検事が探偵が裁判官が根拠のない非難を浴びた未亡人の経っ子に尊敬されていたと聞きつけたと発言した。
10b 裁判官が根拠のない非難を浴びた未亡人の経っ子に尊敬されていたと探偵が聞きつけたと検事が発言した。
11a 村役人が村長が商人がサーフィンを独学で学んだ漁師の甥っ子に嫌われていたと知っていたと思っていた。
11b 商人がサーフィンを独学で学んだ漁師の甥っ子に嫌われていたと村長が知っていたと村役人が思っていた。
12a 管理人が船員が住人が近所で有名人だった大家さんの娘っ子に見下されていたと噂していたと証言した。
12b 住人が近所で有名人だった大家さんの娘っ子に見下されていたと船員が噂していたと管理人が証言した。
13a 王様が大臣が騎士が晩餐会で楽しそうにしていてお姫様の叔母に称賛されたと気づいていたと思っていた。
13b 騎士が晩餐会で楽しそうにしていてお姫様の叔母に称賛されたと大臣が気づいていたと王様が思っていた。
14a 父親が写真家がデザイナーがプライダルフェアで恥をかいた花嫁の母親に信頼されていたといったことがあったと言われていたと父親が言っていた。
14b デザイナーがプライダルフェアで恥をかいた花嫁の母親に信頼されていたと写真家が聞いたと父親が言っていた。
15a レポーターが検事が代理人が嘘を偽った俳優の父親に見取されていたと主張したと報道した。
15b 代理人が嘘を偽った俳優の父親に見取されていたと検事が主張したとレポーターが報道した。
16a 隣人は管理人がその作家が夜な夜な独学に励んでいた修理工の息子に敬愛されていたと話ししていたと思い出した。
16b その作家が夜な夜な独学に励んでいた修理工の息子に敬愛されていたと管理人が話ししていたと隣人が思い出した。
17a 父兄が事務員がその教師が作文発表の準備をしていた少年のお兄ちんに尊敬されていったと噂していたと聞きつけた。
17b その教師が作文発表の準備をしていた少年のお兄ちんに尊敬されていったと事務員が噂していたと父兄が聞きつけた。
18a タイピストが裁判官がその弁護士が再び破産してしまった家主さんの叔母に敬われていたと知っていったと書いた。
18b その弁護士が再び破産してしまった家主さんの叔母に敬われていたと裁判官が知って
いたとタイピストが書いた。

19a 掃除婦が社長がお手伝いさんがソファで伸びをしていたビジネスマンの甥っ子にほめ
られたと思い込んでいたと話していた。
19b お手伝いさんがソファで伸びをしていたビジネスマンの甥っ子にほめられたと社長が
思い込んでいたと掃除婦が話していた。

20a 親族が新聞記者が警官が侵入者を自力で倒した後女の姉に助けを求められたと報道し
たと言っていった。
20b 警官が侵入者を自力で倒した後女の姉に助けを求められたと新聞記者が報道したと親
族が言っていた。

21a 将軍が近辺が召使が公の場に姿を現さなかった皇帝の祖父に懐われていたと悩んでい
たと伝え聞いた。
21b 召使が公の場に姿を現さなかった皇帝の祖父に懐われていたと側近が悩んでいたと將
軍が伝え聞いた。

22a 専務が秘書が支配人が勝手にパーティーに来たセールスマンの叔父に挨拶されたと思
い込んでいたと言った。
22b 支配人が勝手にパーティーに来たセールスマンの叔父に挨拶されたと秘書が思い込
んでいたと専務が言った。

23a 親戚のお隣さんがパイロットが毎晩喜びに楽しんでいたスチュワーデスの母親に紹介
されたと噂していたと聞いた。
23b パイロットが毎晩喜びに楽しんでいたスチュワーデスの母親に紹介されたとお隣さん
が噂していたと親戚が聞いた。

24a 受付係が看護婦が外科医がナイフで怪我をしたウェイターの息子から電話を受けたと
伝え聞かしていたと言っていた。
24b 外科医がナイフで怪我をしたウェイターの息子から電話を受けたと看護婦が伝え聞か
えたと受付係が言っていた。

25a 召使がゲストが仕出屋が自画自賛ばかりしていた主催者の叔母に雇われていたと気づ
いたと思った。
25b 仕出屋が自画自賛ばかりしていた主催者の叔母に雇われていたとゲストが気づいたと
召使が思った。

26a 消防局長がカメラマンがレポーターが事故に責任を感じていた消防士の祖父に軽蔑さ
れていたと気づいたと言った。
26b レポーターが事故に責任を感じていた消防士の祖父に軽蔑されていたとカメラマンが
気づいたと消防局長が言った。

27a 側近がマスコミが知事が晩餐会で自分で酒をついていた乳母の娘におだてられたとは
やし立てたと言っていた。
27b 知事が晩餐会で自分で酒をついていた乳母の娘におだてられたとマスコミがはやし立て
たと側近が言っていた。

175
28a 記者が市議会議員が市長が政治運動に熱心だった郵便配達人の父親に支援されていたと勘違いしていたと報道した。
28b 市長が政治運動に熱心だった郵便配達人の父親に支援されていたと市議会議員が勘違いしていたと記者が報道した。
29a 婦長が消防士が看護婦が火事から逃げ出したメイドの姪っ子に感謝されたと言っていたと聞いた。
29b 看護婦が火事から逃げ出したメイドの姪っ子に感謝されたと消防士が言っていたと婦長が聞いた。
30a お上さんが弟子が管理人が本を読んでリラックスしていた親方の祖父にほめられたと勘違いしたと言っていった。
30b 管理人が本を読んでリラックスしていた親方の祖父にほめられたと弟子が勘違いしたとお上さんが言っていった。
31a 大学院生が助手が教授が学費のことを心配していた助産婦の娘に尊敬されていたといい込んでいたと噂していた。
31b 教授が学費のことを心配していた助産婦の娘に尊敬されていたと助手が思い込んでいたと大学院生が噂していた。
32a 息子が弟子が会計士が財政危機から脱出した職人の父親から相談を受けたと発見したと聞いた。
32b 会計士が財政危機から脱出した職人の父親から相談を受けたと弟子が発見したと息子が聞いた。
## APPENDIX M

### ANOVA TABLE FROM EXPERIMENT 4

#### Tests of Within-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPLX Assumed Sphericity</td>
<td>0.051</td>
<td>1</td>
<td>0.051</td>
<td>5.309</td>
<td>0.031</td>
<td>0.194</td>
</tr>
<tr>
<td>COMPLX Sphericity</td>
<td>0.002</td>
<td>1</td>
<td>0.002</td>
<td>0.212</td>
<td>0.649</td>
<td>0.010</td>
</tr>
<tr>
<td>* SPAN Assumed Sphericity</td>
<td>0.211</td>
<td>22</td>
<td>0.010</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Tests of Between-Subjects Effects

**Transformed Variable: Average**

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>7.471</td>
<td>1</td>
<td>7.471</td>
<td>3</td>
<td>0.000</td>
<td>0.784</td>
</tr>
<tr>
<td>SPAN</td>
<td>0.165</td>
<td>1</td>
<td>0.165</td>
<td>1.762</td>
<td>0.198</td>
<td>0.074</td>
</tr>
<tr>
<td>Error</td>
<td>2.057</td>
<td>22</td>
<td>0.094</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## APPENDIX N

### ANOVA TABLES FROM EXPERIMENT 5

#### Region 9

Tests of Within-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPLX</td>
<td>Assumed Sphericity</td>
<td>1</td>
<td>13212.669</td>
<td>669</td>
<td>0.610</td>
<td>0.026</td>
</tr>
<tr>
<td>Error(COMPLX)</td>
<td>Assumed Sphericity</td>
<td>23</td>
<td>498267.099</td>
<td>787</td>
<td>0.610</td>
<td></td>
</tr>
<tr>
<td>ATTACH</td>
<td>Assumed Sphericity</td>
<td>1</td>
<td>11861.073</td>
<td>073</td>
<td>0.243</td>
<td>0.010</td>
</tr>
<tr>
<td>Error(ATTACH)</td>
<td>Assumed Sphericity</td>
<td>23</td>
<td>1124679.243</td>
<td>098</td>
<td>0.243</td>
<td></td>
</tr>
<tr>
<td>COMPLX*</td>
<td>Assumed Sphericity</td>
<td>1</td>
<td>27540.718</td>
<td>718</td>
<td>1.299</td>
<td>0.053</td>
</tr>
<tr>
<td>ATTACH</td>
<td>Assumed Sphericity</td>
<td>23</td>
<td>487451.456</td>
<td>542</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Region 12

Tests of Within-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMPLX</td>
<td>Assumed Sphericity</td>
<td>1</td>
<td>187065.8473</td>
<td>8.473</td>
<td>0.000</td>
<td>0.498</td>
</tr>
<tr>
<td>Error(COMPLX)</td>
<td>Assumed Sphericity</td>
<td>23</td>
<td>1886859.790</td>
<td>382</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATTACH</td>
<td>Assumed Sphericity</td>
<td>1</td>
<td>38845.920</td>
<td>920</td>
<td>1.840</td>
<td>0.074</td>
</tr>
<tr>
<td>Error(ATTACH)</td>
<td>Assumed Sphericity</td>
<td>23</td>
<td>485649.629</td>
<td>201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMPLX*</td>
<td>Assumed Sphericity</td>
<td>1</td>
<td>17179.473</td>
<td>473</td>
<td>0.809</td>
<td>0.034</td>
</tr>
<tr>
<td>ATTACH</td>
<td>Assumed Sphericity</td>
<td>23</td>
<td>488211.905</td>
<td>605</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


