UNVEILING LINGUISTIC COMPETENCE BY FACILITATING

PERFORMANCE

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

DOCTOR OF PHILOSOPHY

IN

SECOND LANGUAGE STUDIES

MAY 2014

By
Kitaek Kim

Dissertation Committee:
Bonnie D. Schwartz, Chairperson
William O’Grady, Chairperson
Kamil Ud Deen
Theres Grüter
Ho-min Sohn
ACKNOWLEDGMENTS

I am very glad to have an opportunity to express my deepest gratitude to all those people who have contributed to this dissertation. Without their help, I would never have completed it.

First of all, I would like to give my deepest thanks to Bonnie D. Schwartz and William O’Grady, who are my academic parents and role models. I have been constantly impressed (sometimes, overwhelmed) by their superb intellect, keen insight, marvelous preciseness, and vast knowledge. Bonnie trained me to think critically and logically, always pushing me to do my best and making me fully use my genetic endowment—i.e., my brain. William taught me how to develop rough ideas by feeding in appropriate input and feedback regularly and systematically.

I also thank my committee members Kamil Ud Deen, Theres Grüter, and Ho-min Sohn for their helpful comments on the dissertation. It was Kamil who encouraged me to develop the Korean OSV study that was initially a final project for his class. Working with Theres has been great; from her I learned a great deal about psycholinguistic methods and statistical analyses. Ho-min’s wonderful insight about the Korean language always provides me ground for new research questions.

I owe much to other faculty members at UH. I thank Amy J. Schafer for giving me the valuable chance to learn about psycholinguistics by working with her. I am also indebted to Georganne Nordstrom. It was great working with her in the Writing Center. I wish as well to thank Yuko Otsuka, who introduced me to Austronesian syntax in her class and extended my knowledge in syntax. I wish to thank J. D. Brown and Thom Hudson for teaching me about issues in L2 pedagogy in their classes. In addition, I wish to thank Seongah Im for introducing me to the principles of statistics and Catherine Sophian for teaching me various aspects of cognitive development.

My special thanks go to my teachers at Seoul National University: Professor Hyun-Kwon Yang taught me syntax, second language acquisition, and how to live a meaningful life. I also thank Professor Kil-Joong Kim, Professor Oryang Kwon, Professor Moonsu Shin, Professor Jin-Wan Kim, Professor Hyunkee Ahn,
Professor Byungmin Lee, Professor Sun-Young Oh, and Professor Heejeong Ko for their warm hearts and helpful encouragement.

I also want to thank many people I met in Hawai‘i and in Seoul for their support: Hyunah Ahn, Marvin Chandra, Yunchuan Chen, Il Kyu Cho, Min Young Cho, Na Young Choi, Daniela Culinovic, Jonathan Guo, Keiko Hata, Seung Jin Hong, Hyun-Jin Hwangbo, Junko Imai, Seong-ryong Ji, Hanbyul Jung, Chae-Eun Kim, Da Young Kim, Dongmin Kim, Hyunwoo Kim, Ji hyun Kim, Jonny Kim, Jung Hee Kim, Sangki Kim, Sunju Kim, Elaine Lau, Josephine Lee, On-Soon Lee, Sumire Matsuyama, Mari Miyao, Kichun Nam, Jeongyeon Park, Kyae-Sung Park, Mapo Seagulls, Yukyoung Shin, Aya Takeda, Nozomi Tanaka, Zhijun Wen, Gordon West, and Soo Jung Youn. In addition, I sincerely appreciate Professor Sook Whan Cho, Professor Hyun-Sook Chung, Professor Hye-ryeong Hahn, Professor Sang-Keun Shin, and Dr. Jai-Ho Yoo for their warm encouragement and helpful advice to me. I also thank Professor Hyun Sook Ko and Professor Jee Hyun Ma for their kind guidance as I settled down in Hawai‘i. In addition, I appreciate Laurie Durand for carefully reading my dissertation.

My wholehearted thanks and love go to my wife Kum-Jeong for her constant support and love. My biggest asset in life is her, my best friend, colleague, and wife. My deepest gratitude also goes to my parents and parents-in-laws for their support in every way.
ABSTRACT

This dissertation looks at the question of whether the source of language learners’ poor performance is a deficit in knowledge or a performance problem. Two constructions are examined, both being cases in which learners have been reported to show deficits in linguistic competence: (1) Korean OSV (e.g., Elmo-lul Big Bird-ka anayo. ‘Elmo, Big Bird is hugging.’), where knowledge of both scrambling and case markers is necessary for target performance; (2) the English tough construction (English TC; e.g., The doll is easy to see e), where knowledge of tough movement is necessary for target performance (i.e., the object interpretation only). If poor performance (operationalized as chance performance or below) is due to a failure to apply knowledge rather than the absence of knowledge, good performance (i.e., significantly-above-chance performance) should emerge with the effective manipulation of non-grammatical factors to facilitate learners’ application of knowledge. I experimentally tested this hypothesis in two studies, employing both (a) sentence-internal manipulations and (b) sentence-external manipulations. The Korean OSV study collected data from Korean Heritage children (KHC) \( n = 31 \) as well as two groups of native Korean-speaking children, a younger group \( n = 21 \) and an older group \( n = 23 \). The English TC study collected data from native Korean-speaking adult L2ers of English \( n = 49 \) and native Korean-speaking child L2ers of English \( n = 30 \), along with adult native speaker controls \( n = 10 \).

Study 1: Knowledge of scrambling and case in Korean was assessed, first, via picture-selection comprehension tasks that (a) manipulated the prosodic salience of case markers and (b) manipulated context to make the direct object a natural scrambled topic, and, second, via a production task eliciting case markers. The results suggested three possible sources of poor performance. Many participants showed good performance only in the manipulation conditions (vs. the baseline condition), suggesting that their poor performance in the baseline condition could be due to a performance problem (e.g., perception failure or a heavy processing load). Some participants showed absence or errors of case in the production task, which suggests that their poor performance could come from a deficit in knowledge of case or from a mapping problem between a Case feature and case morphology.
Study 2: Knowledge of the English TC was assessed via Truth-Value Judgment Tasks (TVJT) that (a) manipulated verb transitivity to make the infinitival object gap more salient or less salient and (b) manipulated context to avoid or strengthen bias toward the (erroneous) subject interpretation. A Korean TVJT was also conducted to find the interpretation tendency for the Korean translation equivalents of the English TC. Comparisons between the Korean and the English TVJT results revealed a strong subject-interpretation bias in both languages for lower proficiency adult and child L2ers but only in Korean for higher proficiency adult L2ers, which suggests that L2ers do transfer their L1 subject interpretation but can subsequently unlearn it—despite an (ostensible) absence of negative evidence. However, (most of) the high proficiency adult L2ers still showed (below) chance performance in the error-inducing conditions, which suggests that non-grammatical factors can conceal knowledge of the English TC that some of them actually have.

The two studies show that poor performance on the part of language learners—here, Heritage learners and L2 learners—is not necessarily due to a deficit in knowledge, which in turn suggests that linguistic competence can be obscured due to performance reasons.
TABLE OF CONTENTS

Acknowledgments
Abstract
List of Tables .................................................................v
List of Figures ................................................................. vi
List of Abbreviations ......................................................... ix

CHAPTER I: Introduction .........................................................1

CHAPTER II: Linguistic Competence and Performance.........................4
  2.1 Linguistic Competence and Psycholinguistic Performance .................4
  2.2 Sources of Poor Performance in Comprehension .................................9
  2.3 Poor Performance and L2 Acquisition Theories ......................................11
  2.4 Assessing Competence from Performance Data .......................................12
    2.4.1 Deciding the Cut-off Point for Comprehension Data ..........................15
    2.4.2 Deciding the Cut-off Point for Production Data .................................17
  2.5 Facilitating Performance and Testing the Presence of Knowledge .............21
  2.6 Summary of Chapter 2 ......................................................................21

CHAPTER III: Comparing Different Learner Populations .......................23
  3.1 Defining Learner Populations .........................................................23
  3.2 Grouping Learners by Proficiency ....................................................25
  3.3 Summary of Chapter 3 ..................................................................27

CHAPTER IV: Korean Scrambled OSV ...........................................28
  4.1 Competence and Performance: The Korean OSV Pattern .......................28
    4.1.1 Grammar and Processing: Scrambling ..........................................29
    4.1.2 Grammar and Processing: Case ..................................................34
  4.2 Previous Acquisition Studies ..........................................................41
  4.3 Methodological Innovations ............................................................44
<table>
<thead>
<tr>
<th>Section Number</th>
<th>Section Title</th>
<th>Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.1</td>
<td>Picture-selection Comprehension Tasks</td>
<td>44</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Case-marker Elicitation Production Task</td>
<td>46</td>
</tr>
<tr>
<td>4.4</td>
<td>Research Questions and Hypotheses</td>
<td>48</td>
</tr>
<tr>
<td>4.5</td>
<td>The Current Study</td>
<td>48</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Picture-selection Comprehension Tasks</td>
<td>49</td>
</tr>
<tr>
<td>4.5.1.1</td>
<td>Participants</td>
<td>49</td>
</tr>
<tr>
<td>4.5.1.2</td>
<td>Materials</td>
<td>50</td>
</tr>
<tr>
<td>4.5.1.3</td>
<td>Results</td>
<td>53</td>
</tr>
<tr>
<td>4.5.1.3.1</td>
<td>Comprehension Data</td>
<td>53</td>
</tr>
<tr>
<td>4.5.1.3.2</td>
<td>Assessing Competence from Comprehension Data</td>
<td>60</td>
</tr>
<tr>
<td>4.5.2</td>
<td>Case-marker-elicitation Production Task</td>
<td>66</td>
</tr>
<tr>
<td>4.5.2.1</td>
<td>Participants</td>
<td>66</td>
</tr>
<tr>
<td>4.5.2.2</td>
<td>Materials</td>
<td>66</td>
</tr>
<tr>
<td>4.5.2.3</td>
<td>Results</td>
<td>66</td>
</tr>
<tr>
<td>4.5.2.3.1</td>
<td>Production Data</td>
<td>66</td>
</tr>
<tr>
<td>4.5.2.3.2</td>
<td>Assessing Competence from Production Data</td>
<td>74</td>
</tr>
<tr>
<td>4.5.3</td>
<td>Source of Poor Performance of Each Individual in the Baseline</td>
<td>79</td>
</tr>
<tr>
<td>4.5.4</td>
<td>Discussion</td>
<td>83</td>
</tr>
</tbody>
</table>

**CHAPTER V: The English Tough Construction**

<table>
<thead>
<tr>
<th>Section Number</th>
<th>Section Title</th>
<th>Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Competence and Performance: The English TC</td>
<td>85</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Grammar and Processing: Tough Movement</td>
<td>85</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Tough predicates in Korean</td>
<td>89</td>
</tr>
<tr>
<td>5.1.3</td>
<td>Acquisition of the English TC by Korean-speaking Learners</td>
<td>91</td>
</tr>
<tr>
<td>5.2</td>
<td>Previous Acquisition Studies</td>
<td>93</td>
</tr>
<tr>
<td>5.2.1</td>
<td>L1 Acquisition</td>
<td>93</td>
</tr>
<tr>
<td>5.2.2</td>
<td>L2 Acquisition</td>
<td>96</td>
</tr>
<tr>
<td>5.3</td>
<td>Methodological Innovations</td>
<td>98</td>
</tr>
<tr>
<td>5.3.1</td>
<td>TVJT</td>
<td>99</td>
</tr>
<tr>
<td>5.4</td>
<td>Research Questions and Hypotheses</td>
<td>101</td>
</tr>
</tbody>
</table>
CHAPTER VI: General Discussion and Conclusion

6.1 Summary of Main Findings

6.2 What Obscures Linguistic Knowledge?

6.3 Theoretical Implications

6.4 Learnability Problems

6.5 Concluding Remarks

Appendix A. Norming Studies for Prosody Manipulation

Appendix B. Proficiency Data for the Korean OSV Study

Appendix C. Items Used for the Korean OSV study

Appendix D. OSV Accuracy on the Two Manipulation Tasks

Appendix E. OSV Accuracy vs. SOV Inaccuracy

Appendix F. Stimuli for the Case-marker-elicitation Task

Appendix G. Analysis of the Closest Korean Equivalent to the English TC
LIST OF TABLES

4.1 Participant grouping by proficiency .................................................................50
4.2 Testing the presence of TL knowledge (with the comparison-against-chance
diagnostic) ...........................................................................................................60
4.3 Testing the presence of TL knowledge (with the comparison-against-baseline
diagnostic) ..........................................................................................................62
4.4 Subject-markings in Sessions 1 and 2, case-marker elicitation task: KHC........67
4.5 Object-markings in Sessions 1 and 3, case-marker elicitation task: KHC ........67
4.6 Subject-markings in Sessions 1 and 2, case-marker elicitation task: cL1ers ......68
4.7 Object-markings in Sessions 1 and 3, case-marker elicitation task: cL1ers.......68
4.8 Testing the presence of TL knowledge: Subject-marking...............................75
4.9 Testing the presence of TL knowledge: Object-marking...............................75
4.10 Individual evaluation: KHC..............................................................................81
4.11 Individual evaluation: cL1ers .........................................................................82
5.1 Comparing the English TC and Korean surface equivalents.........................91
5.2 Participant grouping by proficiency .................................................................104
5.3 Testing the presence of TL knowledge (with the comparison-against-chance
diagnostic) ..........................................................................................................131
5.4 Testing the presence of TL knowledge (with the comparison-against-baseline
diagnostic) ..........................................................................................................133
5.5 Testing the presence of TL knowledge (with the comparison-against-baseline
diagnostic) ..........................................................................................................134
LIST OF FIGURES

2.1 Sentence-internal and sentence-external processing in sentence comprehension ..6
2.2 Good performance: Significantly-above-chance performance .......................14
2.3 Scope of performance to test the presence of knowledge: Comprehension data .17
2.4 Scope of performance to test the presence of knowledge: Production data ......20
3.1 Different language learner populations ..........................................................25
4.1 An example context-condition item (Song et al., 1997, p. 114, (6)) .............43
4.2 An example context-condition item in the current study ................................45
4.3 Examples of the case-marker-elicitation task .................................................47
4.4 An example of the picture-selection task ......................................................52
4.5 Accuracy rates on scrambled OSV sentences by the 3 KHC proficiency groups ...................................................................................................................54
4.6 Accuracy rates on scrambled OSV sentences by the 3 young cL1ers proficiency groups ...................................................................................................................56
4.7 Prosody–baseline difference in the KHC ......................................................58
4.8 Context–baseline difference in the KHC ......................................................58
4.9 Prosody–baseline difference in the cL1ers .....................................................59
4.10 Context–baseline difference in the cL1ers ...................................................59
4.11 Performance of individual KHC for the OSV items from the manipulations tasks ......................................................................................................................64
4.12 Performance of individual cL1ers for the OSV items from the manipulations tasks ......................................................................................................................64
4.13 Categorizing the participants based on the number of correct responses ....65
4.14 Subject-markings by individual KHC in Session 1: Non-focused subject ....70
4.15 Subject-markings by individual KHC in Session 2: Focused subject ..........70
4.16 Object-markings by individual KHC in Session 1: Non-focused object .......71
4.17 Object markings by individual KHC in Session 3: Focused object ...............71
4.18 Subject-markings by individual cL1ers in Session 1: Non-focused subject ....72
4.19 Subject-markings by individual cL1ers in Session 2: Focused subject .........72
4.20 Object-markings by individual cL1ers in Session 1: Non-focused object ....73
4.21 Object-markings by individual cL1ers in Session 3: Focused object ...................73
4.22 Number of correct subject-markings: KHC ..........................................................76
4.23 Number of correct object-markings: KHC ...........................................................77
4.24 Number of correct subject-markings: cL1ers .......................................................77
4.25 Number of correct object-markings: cL1ers .........................................................77
4.26 Categorizing the participants: Knowledge of subject-marking ............................78
4.27 Categorizing the participants: Knowledge of object-marking ..............................79
5.1 An example of a neutral context TVJT story ......................................................100
5.2 An example of an agent-biased context TVJT story ..........................................101
5.3 Acceptability ratings for verbs in transitive vs. intransitive sentences: 
   English native speakers and Korean native speakers .........................................106
5.4 Acceptability ratings for verbs in transitive vs. intransitive sentences: 
   aL2ers-Low .........................................................................................................109
5.5 Acceptability ratings for verbs in transitive vs. intransitive sentences: 
   aL2ers-Mid .........................................................................................................111
5.6 Acceptability ratings for verbs in transitive vs. intransitive sentences: 
   aL2ers-High ........................................................................................................112
5.7 Acceptability ratings for verbs in transitive vs. intransitive sentences: 
   cL2ers-Low .........................................................................................................114
5.8 Acceptability ratings for verbs in transitive vs. intransitive sentences: 
   cL2ers-Mid .........................................................................................................115
5.9 Acceptability ratings for verbs in transitive vs. intransitive sentences: 
   cL2ers-High ........................................................................................................117
5.10 Object interpretation (%) in the English/Korean TVJT: Control groups ............119
5.11 Object interpretation (%) in the English/Korean TVJT: aL2ers-Low .................122
5.12 Object interpretation (%) in the English/Korean TVJT: aL2ers-Mid .................124
5.13 Object interpretation (%) in the English/Korean TVJT: aL2ers-High ...............125
5.14 Object interpretation (%) in the English/Korean TVJT: cL2ers-Low .................126
5.15 Object interpretation (%) in the English/Korean TVJT: cL2ers-Mid .................127
5.16 Object interpretation (%) in the English/Korean TVJT: cL2ers-High ...............129
5.17 Object interpretation (%) in the English TVJT ..................................................130
5.18 Number of correct object interpretations in the English TVJT by aL2ers ........135
5.19 Number of correct object interpretations in the English TVJT by cL2ers ........136
5.20 Number of correct object interpretations in the English TVJT by aL2ers:
  Grouping by context ...........................................................................................137
5.21 Number of correct object interpretations in the English TVJT by aL2ers:
  Grouping by verb transitivity .............................................................................137
5.22 Number of correct object interpretations in the English TVJT by cL2ers:
  Grouping by context ...........................................................................................137
5.23 Number of correct object interpretations in the English TVJT by cL2ers:
  Grouping by verb transitivity .............................................................................138
5.24 Categorizing the participants: Knowledge of the English TC .....................139
6.1 Optimal mapping and non-optimal mapping ......................................................152
LIST OF ABBREVIATIONS

The following abbreviations are used to label the linguistic terms employed in this dissertation.

ACC  Accusative case particle
CL   Classifier
DAT  Dative case particle
DECL Declarative sentence-type suffix
HON  Honorific suffix
LOC  Locative case particle
NML  Nominalizer suffix
NOM  Nominative case particle
PASS Passive suffix
PAST Past tense suffix
PL   Plural suffix or particle
Q    Question marker, i.e., interrogative sentence-type suffix
SES  Sentence ender suffix
TOP  Topic particle
CHAPTER I
INTRODUCTION

This dissertation examines two constructions in which learners have been reported to show deficits in linguistic competence: (a) the scrambled Korean Object-Subject-Verb pattern (Korean OSV; e.g., Elmo-lul Big Bird-ka anayo. ‘Elmo, Big Bird hugs.’) where knowledge of both scrambling and the distinguishing function of case markers1 is necessary for target performance; (b) the English tough construction (English TC; e.g., The doll is easy to see e) where knowledge of tough movement regulates target performance, i.e., the “object interpretation” only. If poor performance (operationalized as chance performance or below) is due to failure in the application of knowledge, not absence of knowledge, good performance (i.e., significantly-above-chance performance) should emerge with the effective manipulation of non-grammatical factors to facilitate learners’ application of knowledge. I experimentally test this, in each study, using both (a) sentence-internal manipulations and (b) sentence-external manipulations.

To interpret the scrambled Korean OSV pattern, knowledge of scrambling and overt morphosyntactic elements, namely, case markers, is necessary for target-like performance. Song, O’Grady, Cho, and Lee (1997) reported that Korean Heritage children2 (KHC) in the United States (ages 3–8) showed difficulty in understanding scrambled Korean OSV sentences. In a picture-selection comprehension test, the KHC largely miscomprehended OSV as canonical SOV instead, even when a context sentence was used to make the scrambled sentence sound natural. This suggests that the KHC did not always make use of knowledge of scrambling and case to establish the syntactic relations of the arguments. Rather, they interpreted the sentences linearly, with the first NP as the subject and the

---

1 De Hoop and Malchuvok (2008) explain that case serve two overlapping functions in language. On the one hand, case can serve a “distinguishing function” in that it distinguishes between a verb’s two core arguments (its subject and its direct object). On the other hand, in some languages, it can have an identifying function in that it helps identify the thematic role of the NP on which it appears. In Manipuri, for example, ergative case appears only on agents.

2 Heritage speakers are “child and adult members of a linguistic minority who grew up exposed to their home language and the majority language” (Montrul, 2010, p. 4). It is common for their command of the heritage language to fall short of the native-speaker level of their parents (Montrul, 2006).
second NP as the direct object. Song et al. (1997) attributed this poor performance to “deficits in children’s linguistic knowledge” (p. 119).

For the English TC, knowledge of tough movement is responsible for target-like performance. Native English-speaking children, even at age 8, have been claimed to show syntactic deficits in the English TC (C. Chomsky, 1969). Consider the following two sentences (N. Chomsky, 1964, pp. 61-65).

(1) a. The boy is eager to please.
   b. The boy is easy to please.

The two sentences look identical except for the adjectives in them, eager and easy. However, they are structurally very different: The boy is understood in (1a) as a “pleaser” but in (1b) as a “pleasee.” Sentences like that in (1b) have been called “tough constructions.” Native English-speaking children erroneously interpreted sentences of the type in (1b) as the type in (1a), as late as age 8 (C. Chomsky, 1969). To understand the English TC, it is necessary to know tough movement, in which the matrix subject is linked to the object gap in the subordinate clause.

There are two conflicting accounts for why learners make such errors: The first is that the learners have a deficit in the linguistic knowledge underlying the construction, that is, knowledge of scrambling and case in Korean and of tough movement in English. The second account is that the learners do have adult-like knowledge, but for extraneous reasons, they are not able to use the knowledge effectively in the course of comprehension or production. The current study addresses this issue.

The structure of this dissertation is as follows. Chapter 2 gives the background for why this dissertation deals with this competence vs. performance issue. It first presents two opposing theoretical approaches to the issue, the Representational Deficit Approach and the Computational Difficulty Approach. It then discusses how to assess knowledge from performance data. Chapter 3 gives the background for comparing different learner population, that is, native children vs. Heritage children and child L2ers vs. adult L2ers. Chapter 4 provides the details and the results of the study that was conducted to test whether observed poor performance on the scrambled Korean OSV by KHC and native
Korean children is due to a knowledge deficit or performance-related phenomena. Chapter 5 provides the details and the results of the study that was conducted to test whether native Korean child L2ers and adult L2ers are able to acquire the English TC and assesses whether their poor performance comes from a knowledge deficit or from performance problems. Chapter 6 concludes the dissertation. It discusses the research findings and considers the following two questions: (a) Why is the learner’s linguistic knowledge sometimes obscured? (b) What are the theoretical implications of the findings from the two studies?
CHAPTER II
LINGUISTIC COMPETENCE AND PERFORMANCE

This chapter discusses the relation between linguistic competence and performance. I will first define the relevant linguistic concepts: competence, performance, grammar, and the parser. I will then discuss the two possible sources of poor performance—a deficit in competence and a problem in performance. Next, two opposing theoretical approaches on how to view poor performance will be introduced: One approach attributes poor performance to a deficit in competence, while the other attributes it to a performance failure. I will then introduce diagnostics for assessing linguistic competence through performance data, which are based on rejecting the null hypothesis. Finally, I will propose a method to test the source of poor performance, that is, whether the source is a deficit in knowledge or a performance failure.

2.1 Linguistic Competence and Psycholinguistic Performance

Linguistic competence is “the speaker/hearer’s knowledge of his language” (N. Chomsky, 1965, p. 4) or “tacit knowledge of the generative grammatical system that underlies language” (Lust, 2006, p. 124). This knowledge corresponds to the grammatical information used in parsing and producing a language, or any representation of linguistic information (Carroll, 2001, p. 24). This grammatical information, within, e.g., the Principles and Parameters framework, refers to universal principles and parametric variation, where scholars have distinguished it from descriptive grammar (Schwartz, 1986; VanPatten, 2011). Linguistic competence can be inferred from language performance, defined as “the actual use of language in concrete situations” (N. Chomsky, 1965, p. 4) or “behavior with language, e.g., speaking, comprehending or making judgments about language in real time” (Lust, 2006, p. 124).

Performance, however, can fail to reveal linguistic competence. In some cases, competence can be obscured due to the intervention of grammatically irrelevant conditions such as “memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic)” in the application of the knowledge (N. Chomsky, 1965, p. 3). This suggests that the mechanisms specialized for language performance (i.e., the
performance systems) are distinct from the underlying repository of linguistic knowledge (i.e., the competence system).

The standard view of the competence system vs. the performance systems has been posited as *the grammar vs. the parser* (or *the producer* for production) (for discussion, see e.g., Crocker, Pickering, & Clifton, 2000; Phillips, 2004). The grammar is assumed to constitute a *lexicon* (i.e., a mental dictionary), *syntax* (i.e., computational system forming a syntactic structure), an *LF component* (i.e., responsible for mapping the syntactic structure into a corresponding Logical Form representation), and a *PF component* (i.e., responsible for mapping the syntactic structure into a Phonetic Form representation) (N. Chomsky, 1995), which together are the “knowledge which the parser puts to use” (Pritchett, 1992, p. 2). The parser is a mechanism “designed to encode the signal in various representational formats,” which is “tuned to the frequency of particular structures in the language the hearer hears” (Carroll, 2001, p. 10).¹ I adopt this standard view which looks at the grammar and the parser as two separate systems.²

Since a primary goal of language acquisition research has been “to assess the Chomskyan notion of grammatical competence” (Demuth, 1996), assessing knowledge of syntax has received particular attention among acquisition researchers. However, knowledge of syntax is an abstract notion, which cannot be physically revealed; it is only inferred from language performance, by testing whether a learner correctly parses or produces a target language (TL) structure. Thus, to assess whether a language learner has certain syntactic knowledge, researchers examine the learner’s language behavior, such as comprehension and production. Correct performance suggests the presence of the

---

¹ This characteristic of the parser suggests that processing *routines* can be generated when particular structures are processed frequently. O’Grady (2005, p. 10) emphasizes that computational routines should be distinguished from grammatical rules. He defines routines as “real-time processes,” which say “how the structure is built,” and rules as “patterns of elements,” which say “what the structure is.” He further notes that “the emergence and strengthening of routines” facilitates processing.

² There are two other views for the competence system vs. the performance systems: First, according to Phillips (2004), grammatical knowledge is a real-time system for constructing sentences. That is, there is not a distinction between the grammar and the parser (and the producer). In the standard view, the grammar is time-independent whereas the parser and the producer are time-dependent systems. Second, according to O’Grady (2005, 2013), there is no grammar but only the parser. The parser maps a form onto a semantic representation without building syntactic representations.
knowledge, assuming that the knowledge is used when the parser or the producer processes the TL structure.

However, it is worth noting that comprehension and production are psycholinguistic processes in which processing syntax is merely one of the multiple processes that a language user employs to comprehend or produce a sentence. For example, comprehension involves (at least) the following processes: (a) sound processing (e.g., perception and segmentation of speech signals), (b) syntax/semantics processing, and (c) sentence-external information processing. Figure 2.1 illustrates the sources of information and types of processing involved in sentence comprehension that I assume.³

![Figure 2.1](image-url)

*Figure 2.1. Sentence-internal and sentence-external processing in sentence comprehension (adapted from Truscott & Sharwood-Smith, 2004, p. 2, Figure 1)*

In sentence-internal processing, the first stage is to process the sound of language (i.e., the sound of morphemes and words). The acoustic properties of the speech signals must be perceived and subsequently segmented into phonemes, syllables, morphemes, and words. Only when the speech signals have been coded in the auditory processor and

---

³ Figure 2.1 exhibits sentence-internal processing and sentence-external processing in the course of comprehension. According to O’Grady (to appear), these are termed “sentence-level (or structural) processing” and “pragmatic processing,” respectively. Sentence-level processing is concerned with “the form and composition of the utterance, including the lexical properties of its component words, their linear relationship to each other, and their morphological form.” Pragmatic processing is concerned with “linguistic and situational context” information.
mentally represented in the system can the syntax/semantics processor extract additional information from them (Carroll, 2001, p. 10). In fact, it is common for acquisition studies assessing knowledge of syntax to presuppose that the speech signals are correctly turned into mental representations, and that these mental representations then arrive in the correct form in the syntax/semantics processor.

The second stage of sentence-internal processing is to process the syntax/semantics of language. In this stage, two types of syntactic knowledge are used from the parser: lexicon and syntax. In recent theories of syntax (e.g., the Minimalist Program; N. Chomsky, 1995), a syntactic derivation requires (a) subcategorization information for lexical items and (b) combinatorial operations (i.e., Merge and Move). Two lexical items are merged in such a way as to satisfy, e.g., the subcategorization requirements of a lexical item. More technically speaking, when, e.g., a constituent merges with another constituent in its complement position of an X’-schema, the uninterpretable syntactic features of the constituent must be checked off. Only interpretable features can be spelled out. For example, the transitive verb *tickle* has an uninterpretable feature [uN], so that only an NP can be merged with the verb as its complement because the NP can check off the [uN] feature in *tickle*. This uninterpretable feature comes from the subcategorization information of the verb *tickle*. In this case, the knowledge of syntax to assess in language learners is whether they have come to know this syntactic requirement of this lexical item and are able to build the relevant X’-structure.

Notably, syntactic operations such as Case assignment/checking and movement are subject to the computational system for which subcategorization information is not mainly responsible. For Case assignment, when a functional projection such as TP or vP merges with a lexical projection, syntactic Case features must be checked off. For example, T bears an uninterpretable Case feature [uCase:NOM]. An NP checks off the uninterpretable Case feature, determining the Case value [uCase:NOM]. Movement phenomena also occur via checking off uninterpretable features in functional projections.

4 Interpretable features refer to the features that have an effect on semantic interpretation, while uninterpretable features do not contribute to semantic interpretation. Only when every uninterpretable feature is checked off are the remaining interpretable features spelled out and mapped to relevant morphemes or words.
For example, it has been proposed that I to C movement in English (e.g., *Is John a student?*) occurs in order to check off the strong \([+Q]\) feature under C.\(^5\)

In sentence-external processing, both linguistic and non-linguistic information can be involved in message analysis, while the parser deals with the internal structure of a sentence. Relevant sentence-external sources of information can include (a) “sense-semantic information,” (b) “discourse-context information,” and (c) “information about discourse focus” (Crocker et al., 2000, pp. 7–9).\(^6\) Sense-semantic information is related to *plausibility*. For example, for the sentence *When the ball rolled the man jumped*, the intransitive reading of *rolled* is plausible, while the transitive reading of *rolled* is implausible. Such plausibility information helps hearers to process the intransitive reading. Discourse-context information can contribute to disambiguation. For example, for the sentence *I saw the man with the binoculars*, if a context is presented where only one man has been mentioned, the possessor of the *binoculars* will more likely be *I*; however, if a context describes more than one man, the possessor of the *binoculars* will more likely be *the man*. Information about discourse focus can facilitate processing of a target sentence by organizing given (topic) and new (focus) information within Information Structure.\(^7\) Listeners are actively searching for a focus when processing speech, and the identification of focus facilitates the rapid comprehension of the message (Cutler & Fodor, 1979; Cutler & Foss, 1977). Information about discourse focus distinguishes the topic and the focus of the target sentence, and compatible topic-focus ordering of the arguments in the target sentence reduces the processing burden (e.g., Hwang, 2008). By organizing Information Structure, information about discourse focus can also promote the perception of sound forms, as has been attested by experiments on the processing of focal accent and Information Structure; the facilitating effect appears

---

\(^5\) The strength of such features varies parametrically. For example, \([+Q]\) in English is strong and thus it motivates the movement of a *wh*-phrase; in contrast, \([+Q]\) in Korean is weak and thus it does not motivate the movement of a *wh*-phrase.

\(^6\) Crocker et al. (2000) list relevant sources of information for language processing as follows: (a) “syntactic category information” (e.g., verb, intransitive, subject), (b) “grammatical features” (e.g., case, number, gender), (c) “grammatical rules,” (d) “sense-semantic information,” (e) “discourse-context information,” (f) “prosody and punctuation,” and (g) “information about discourse focus.”

\(^7\) Given information has often been identified with the topic of a sentence and new information with the focus of a sentence, because what is under discussion (i.e., given) is typically the topic of the current utterance (for discussion, see Birner & Ward, 2009).
when the focal accent falls on the words corresponding to new information, but
 disappears when the accent falls on the words corresponding to old information (Bock & Mazella, 1983; Terken & Nooteboom, 1987). In this regard, Carroll (2006, p. 20) asserts that “we are not noticing sound forms because they are objectively salient[,] they are perceptually salient because we expect them to be filling certain informational functions.”

2.2 Sources of Poor Performance in Comprehension

The fact that sentence comprehension comprises multiple processes suggests that when a comprehension failure occurs, it is not easy to identify where the failure comes from. In fact, if a hearer fails to understand a sentence although the hearer knows (the meaning of) each word in the sentence, there are at least three possible reasons for the breakdown in comprehension. First, the hearer might be unsuccessful in processing the speech signal. Second, the hearer might not know the syntax of the sentence and thus be unable to process it correctly. Third, non-linguistic information might have interfered with the processing of the sentence. Let us address this issue by considering the processing of the sentence The dog is bitten by a man.

The first stage of processing The dog is bitten by a man is to perceive the acoustic properties of the speech signal in the sentence. The auditory processor of the hearer must process a variety of acoustic information, including degree of stridency, formants, fundamental frequency, voicing, and so on, from the sound of the utterance. This processing is not simple, because for the same sentence, the acoustic information differs between different speakers (e.g., different fundamental frequencies). Even when one speaker repeats a sentence s/he already uttered, the two utterances’ acoustic properties are highly unlikely to be exactly the same (the so-called problem of perceptual constancy). Next, the phonological processor segments the sentence into seven phonological words ([də], [dɔɡ], [ɪz], [bɪtən], [baj], [ə], [mæn]), in which each phonetic representation is composed of one or two syllables, following the hearer’s phonotactic rules. The speech signal itself does not exhibit such discrete strings, but the hearer’s phonological, morphological, and lexical knowledge segments the speech signals in such a way. If a hearer failed to perceive a certain ‘morpheme’, the information of the morpheme would
not be delivered to the syntax/semantics processor, which in turn could result in a failure in comprehension of the morpheme’s properties.\(^8\)

The syntax/semantics processor receives the information from the phonological processor. The syntax/semantics processor extracts the syntactic information in the sentence, using knowledge of the lexicon and syntax. Here, the grammatical relation between the arguments is determined: *A dog* serves as the subject, while *a man* serves as the complement of the preposition *by*. In addition, the agreement relation between the subject and the auxiliary as well as the definite/indefinite information is processed. If there is a deficit in knowledge, the syntax/semantics processor will bypass the relevant information, because the processor does not have sources in the grammar that would allow it to draw such information from the sentence.

The parsed linguistic information is integrated with sentence-external information for comprehension. For example, world knowledge would suggest the ‘biter’ being *the dog* and the ‘bitee’ being *a man*. The hearer might *anticipate* such a semantic relation even before completing the sentence’s parsing. Such anticipation can result in a failure in comprehension if the parser does not fully process the syntactic information in the sentence, i.e., when *good enough* parsing (Ferreira & Patson, 2007) happens.\(^9\) However, if a context were to permit the hearer to anticipate *a man* as the ‘biter’, the hearer might find processing the sentence easier. For example, a speaker’s introduction of a humorous situation, such as “I saw a funny comedy show,” might be helpful for the hearer to correctly process the target sentence *The dog is bitten by a man*. Furthermore, a felicitous context to make the first NP topical would reduce the processing burden for the target sentence. For example, a speaker’s mentioning of *a dog* in a preceding sentence (e.g., “There is a cute dog here”) might facilitate the processing of the target sentence.

---

\(^8\) However, according to Bonnie D. Schwartz, it is not always bottom-up processing. For instance, when a native hearer hears all of the words except [ɔ], the hearer would ‘stick in’ the [ɔ] or a [ðɔ].

\(^9\) In *good enough* parsing, “the language comprehension system creates syntactic and semantic representations that are merely ‘good enough’… given the task that the comprehender needs to perform” (Ferreira & Patson, 2007, p. 71).
2.3 Poor Performance and L2 Acquisition Theories

To assess whether a language learner has certain syntactic knowledge, researchers examine the learner’s language behavior, such as comprehension and production. Correct performance suggests the presence of the knowledge. Poor performance, however, has been viewed in two ways: One view attributes poor performance to an absence of the knowledge. The other view assumes intact linguistic knowledge, but attributes the poor performance to a failure (due to various extraneous reasons) to apply that knowledge. For convenience, I follow Hopp (2007), labeling the two views as the Representational Deficit Approach (RDA) and the Computational Difficulty Approach (CDA).10

Proposals claiming a deficit in knowledge for adult L2 learners have been formulated based on L2 learners’ performance errors (e.g., Bley-Vroman, 1990; Clahsen & Muysken, 1986; Eubank, 1993/1994, 1994; Hawkins & Chan, 1997; Hawkins & Hattori, 2006; Hawkins & Liszka, 2003; Tsimpli, 2003; Tsimpli & Dimitrakopoulou, 2007). This approach often links competence deficits with critical period effects reflecting learners’ maturational changes, as, for instance, in a claim made by Beck (1999, p. 316): “The morphosyntactic features that require or prohibit thematic verb raising become impaired during the course of maturation.” One piece of evidence in favor of this approach comes from adult L2ers’ morphological difficulties (e.g., Beck, 1997). This approach claims that such difficulties are a sign of incomplete acquisition of morphosyntactic knowledge. Another piece of evidence stems from adult L2ers’ low accuracy and inconsistency in grammaticality judgment tasks (e.g., Johnson & Newport, 1989, 1991).

The CDA attributes (at least some) poor performance to a performance failure. For example, L2 studies under the umbrella of the Missing Surface Inflection Hypothesis (Chondrogianni & Marinis, 2012; Haznedar & Schwartz, 1997; Prévost, 2001; Prévost & White, 2000; Tran, 2005) separate morphology from syntax, arguing that production errors in morphology do not necessarily suggest impaired syntax. Under this view, the limited processing resources of learners as well as L1 influence have received particular attention as sources of failure to use knowledge (Grüter & Crago, 2011; Hopp, 2006).

---

This approach considers (adult) L2ers’ morphological errors to result from difficulties in accessing, retrieving, or controlling morphological knowledge, assuming that L2ers know more than they actually demonstrate knowing. More recently, Schwartz (2010) proposes that the Missing Surface Inflection Hypothesis can also explain errors in comprehension, pointing to a performance failure in comprehension.

Proficiency matters for the CDA, while it does not matter for the RDA. Proponents of the RDA note that L2ers’ poor performance could be due to a performance failure, and proponents of the CDA acknowledge that L2ers in especially early developmental stages could have a deficit in TL knowledge. A crucial difference between the two approaches is placed on L2ers with advanced TL proficiency. The RDA contends that L2ers with even very advanced proficiency cannot have TL knowledge if that knowledge is not present in their L1. However, the CDA predicts that lower proficiency L2ers may not have TL knowledge, while higher proficiency L2ers may well be able to come to have it.

To test the two approaches through experimental research, an important question to address is this: Does the analysis set a reasonable cut-off point for the accuracy rate that indicates the presence of TL knowledge? Jiang (2004) claimed that whatever accuracy rate in the production of morphemes is achieved in an experiment, if the accuracy is not at ceiling, the experimental results cannot contribute to resolving the debate between the RDA and the CDA. For example, an accuracy rate of 80% can be viewed as a sign of a deficit in TL knowledge, supporting the RDA, or as a sign of the presence of TL knowledge, supporting the CDA. Thus, to test whether TL knowledge is present in a learner’s grammar from experimental research, setting a reasonable cut-off point is crucial. The following section addresses this issue.

2.4 Assessing Competence from Performance Data

In a language task assessing competence, there are cognitive processes that necessarily occur in addition to the linguistic processes. For example, a comprehension task such as a Truth-Value Judgment task (TVJT) requires the following cognitive and linguistic processes: (a) participants must understand the story, such as “the characters, their motives, the order of events, the paths of causation”; (b) participants must retain the narrative in memory; and (c) participants must parse the target sentence and arrive at an
interpretation (de Villers & Roeper, 1996, p. 172). In a production task, too, participants must employ complex processes, such as planning, searching for and choosing appropriate items in the lexicon, encoding syntactic and phonological rules, and using the muscles in the articulators (e.g., jaw, tongue, lungs, etc.) (Levelt, 1989). The complexity of cognitive and linguistic processes in comprehension and production tasks suggests that when non-target performance occurs, it is hard to identify where it originates.

The necessary use of cognitive processes in a language task suggests that an assessment of competence does not require correct performance 100% of the time. It is possible that participants give a wrong response not because of a deficit in linguistic knowledge but due to a problem in a cognitive process. For example, participants may give a wrong response in a TVJT because they fail to retain the narrative in memory. As such, it might be acceptable to state that good performance suggests the existence of competence. Here, however, a non-alignment problem arises. Competence is a categorical concept, viewed in terms of whether or not rather than to what extent. Meanwhile, the notion of good performance is not a categorical concept, but a gradient concept in that it deals with to what extent the performance is good. In a language acquisition study dealing with performance data, it is quite a complicated problem to determine just how good the performance on a task should be in order to be taken as indicating the associated competence.

To solve the non-alignment problem, researchers have attempted to find motivated and reasonable criteria to define good performance. Although what criterion is most appropriate depends on the particular study and aspect of language under investigation (Stromswold, 1996), it is worthwhile to seek principled ways of deciding the criterion, across data types (i.e., group data, individual data).

A principled way of determining good performance is to decide on a performance rate that rejects a null hypothesis. There are two types of hypothesis, i.e., the experimental hypothesis (H₁) and the null hypothesis (H₀). The null hypothesis stating that an effect is absent is the opposite of the experimental hypothesis. In statistics, the experimental hypothesis cannot be proven, but the null hypothesis can be rejected. Only when data give us confidence to reject the null hypothesis can our experimental hypothesis be supported (Field, 2009, p. 27). In a task testing for presence of target
knowledge, the experimental hypothesis is the presence of knowledge, and the null hypothesis is the lack of knowledge. To reject the null hypothesis (i.e., to support the experimental hypothesis), we should predict what would happen when participants do not know the target grammar. If participants’ linguistic behavior is significantly different from the prediction, we may reject the null hypothesis.

If participants do not know the target grammar, at least two scenarios can be considered: First, the participants give completely random responses to the test items. In this scenario, chance performance should appear. For example, Gordon (1996, p. 228) explained that in a TVJT, statistical analyses should be carried out to “determine whether the responses were significantly different from chance.” If chance is considered to occur at the point of response, the expected chance value would be 50% since there are only two possible responses (i.e., True and False), just like tossing a coin to see whether it comes up heads or tails.\(^\text{11}\) If the accurate response rate is significantly higher than chance, the null hypothesis—participants do not know the target grammar—can be rejected. In other words, significantly-above-chance performance is categorized as good performance, which suggests associated competence. Figure 2.2 demonstrates this logic (when there are two possible responses).

---

**Figure 2.2.** Good performance: Significantly-above-chance performance.

---

\(^{11}\) Gordon notes that there is another type of chance: chance at the point of choice. Gordon explains that if chance is at the point of choice, the expected chance value might not be 50% if there are more than two choices. For example, for the sentence *The dog said that the monkey is washing himself, the monkey is the antecedent of himself.* If there are four animals (e.g., *a dog, a monkey, a cat, and a squirrel*) in the context, the chance of responding correctly is 25%.
In the second scenario, there would be no difference between the performance in a condition under investigation and the performance in a baseline condition. This scenario is inspired by Grimshaw and Rosen’s (1990) idea about assessing children’s knowledge of Binding Principle B. See the following excerpt.

If a subject does not know Principle B, then in terms of grammaticality, the local and the nonlocal antecedent are equivalent. The crucial assumption is that if there is no knowledge of Principle B, the BT[KK: Binding Theory]-ungrammatical condition will be treated just like the BT-grammatical condition—without Principle B, both are equally grammatical. If subjects treat the two conditions differently, they are detecting a difference in the acceptability of the sentences in the two conditions, and we can reason from this that they must have knowledge of the binding theory. Thus, this crucial comparison does not require a high rate of performance on any given condition. (p. 191)

That is, if the accuracy rate in a condition under investigation is significantly higher than the accuracy rate in a baseline condition, the null hypothesis—participants do not know the target grammar—can be rejected.

2.4.1 Deciding the cut-off point for comprehension data.

As discussed above, when analyzing numerical performance data, there are potentially two diagnostics for calculating the cut-off point that suggests participants possess the target knowledge. The first diagnostic considers the cut-off point to be the point above which the accuracy rate is significantly higher than chance performance (henceforth, the “comparison-against-chance diagnostic”). The second diagnostic assesses whether participants’ behavior in a treatment condition is significantly different from their behavior in a baseline condition (henceforth, the “comparison-against-baseline diagnostic”).

To determine the presence of knowledge, I use the less controversial diagnostic, which is the comparison-against-chance diagnostic. Identifying the chance value can be obvious (e.g., 50% when there are two possible responses). However, identifying a
baseline is not. It requires a reasonable explanation of why it is a baseline (which can be controversial). In addition, for the data that will be reported in this dissertation, the comparison-against-chance diagnostic is a more conservative way of looking at the cut-off for presence of knowledge.\textsuperscript{12} The use of the more conservative diagnostic is advantageous in that it minimizes controversy with respect to the cut-off point.

However, a question arises: If groups/participants do not pass the comparison-against-chance criterion, can we simply conclude that they do not have the target knowledge, given that competence is a categorical concept, viewed in terms of presence or absence (i.e., “+” or “−”)? As discussed above, we submitted that there are at least two ways (when testing a null hypothesis) to test for the presence of competence (i.e., comparison-against-chance and comparison-against-baseline). What if the performance of groups/participants falls below the more conservative rejection criterion (here, the comparison-against-chance) but above the less conservative rejection criterion (here, the comparison-against-baseline)? In this situation, we may not be able to decide either presence or absence of knowledge. Therefore, to cover this situation, we would need an additional category, possibility of knowledge, which can be marked with a question mark (“?”). In this dissertation, I also present results in terms of the comparison-against-baseline diagnostic, when I discuss the idea that we cannot simply conclude an absence of knowledge if participants do not pass the comparison-against-chance diagnostic.

I will consider there to be a deficit in target knowledge when the participants pass neither the comparison-against-chance nor the comparison-against-baseline diagnostic. That is, I use three categories for the results of my tests of the presence of TL knowledge: “+” (presence of knowledge), “?” (possibility of knowledge), and “−” (absence of knowledge). Figure 2.3 represents this logic graphically.

\textsuperscript{12} In a task in which two responses are possible, if the accuracy rate in a baseline condition is lower than 50%, the more conservative diagnostic is the comparison-against-chance diagnostic. However, if the accuracy rate in a baseline condition is higher than 50%, the more conservative diagnostic is the comparison-against-baseline diagnostic.
In a study such as this, it is important to examine individual data, in addition to group data. Averaged results, which reflect the good and poor performances of all participants, potentially mask underlying contrasts. I claim the presence of knowledge in an individual when his/her performance passes the comparison-against-chance diagnostic, which is the less controversial (and more conservative) diagnostic for the data in this dissertation. To assess whether the participants’ response patterns are significantly different from chance performance or from baseline performance at the $p < .05$ level, I calculate the $p$-value based on a binomial distribution (Orfitelli & Hyams, 2012). This is because the number of tokens obtained from an individual participant is not large. I consider an absence of knowledge in an individual to be demonstrated when his/her performance passes neither the comparison-against-chance diagnostic nor the comparison-against-baseline diagnostic. When a participant passes the comparison-against-baseline criterion but does not pass the comparison-against-chance criterion, I put a question mark (?), claiming evidence for neither presence nor absence of knowledge.

**2.4.2 Deciding the cut-off point for production data.**

Comprehension has been regarded as generally preceding production in language acquisition (Clark, 1993, p. 246), and therefore comprehension tasks (e.g., TVJT, picture-selection tasks, act-out tasks, etc.) are often used to test syntactic knowledge. In this vein,
Demuth (1996) suggests that grammatical competence can be underestimated when examining production data.

If comprehension precedes production, we might expect better performance in comprehension tasks than in production tasks. However, this is not always the case. Unsworth (2007) reported that child L2ers of Dutch showed more target-like production than comprehension in the domain of direct object scrambling, a property of Dutch that requires syntactic, semantic, and discourse knowledge. To explain this discrepancy, Unsworth took up Miller and Schmitt’s (2004) suggestion that limited discourse integration is responsible for children’s comprehension failures.

If it is not conclusive that comprehension tells us more about competence than production does, it may be useful to conduct both comprehension and production tasks. In particular, if it is necessary for participants to integrate various pieces of morphosyntactic knowledge in a comprehension task, it may be worthwhile to test each piece of knowledge in a separate task, because a problem might then be located in the integration of the various sources of knowledge rather than being considered the lack of either type of knowledge. The separate task, no matter whether it is a comprehension task or a production task, should (ideally) be cognitively less demanding than the comprehension task requiring the integration of the various pieces of morphosyntactic knowledge.

In naturalistic production data, the response is typically not a choice from a limited number of options, making it harder to determine the cut-off point that suggests the presence of knowledge. Researchers seem to have been unable to agree on a cut-off point for production, and a variety of proposals have been put forth. For example, Bickerton (1981) and Snyder and Stromswold (1997) consider that the initial (non-rote) emergence of a linguistic phenomenon indicates acquisition. Snyder and Stromswold stress that age of first productive use is highly correlated with other measures of acquisition, based on Stromswold’s (1989) study. However, it might be hard to determine whether that first use is based on knowledge or imitation or routine. In the early morpheme order studies with L1 learners (Brown, 1973), 90% accurate use of target morphology in obligatory contexts (in 3 consecutive sessions) was considered to demonstrate the acquisition of the target item. However, the particular percentage required is arbitrary. Stromswold (1996, p. 44)
points out that “there is nothing magical about 90% correct use in obligatory context.” In a study by Vainikka and Young-Scholten (1994), the criterion for acquisition in production is an accuracy level of 60%; that is, if lexical items/inflectional morphology associated with the grammar occur in 60% or more of obligatory contexts, acquisition was assumed to have occurred. However, as White (2003, p. 77) points out, this 60% criterion is, again, “arbitrary and apparently unmotivated.”

In fact, naturalistic speech data have challenges in terms of the absence of TL phenomena. First, it is hard to collect sufficient data, as large amounts of data may not include the target construction or morphemes. Second, it is hard to determine whether the absence of, e.g., a particular grammatical construction is “due to lack of linguistic ability, lack of exposure to the construction, or lack of appropriate discourse contexts in the sample” (Demuth, 1996, p. 19).

The problem of absence can be reduced if an elicited-production task is employed. In Berko’s (1958) now famous research, children between the ages of 4 and 7 were shown pictures of an object, such as a cartoon bird, and were then informed that the object was a “wug.” Next, a second picture was presented in which there were two of the objects. Children were asked to finish the experimenter’s sentence: “There are two ___.” The results indicated that most children produced “wugs,” suggesting that they knew that the plural suffix -s attaches to nouns.

The use of elicited-production data has another advantage: By avoiding the absence problem, it allows us to consider a more reasonable way of deciding the cut-off point for assessing knowledge. The way used to decide the cut-off point with production data, however, should be different from the way used with comprehension data. For comprehension data, I use the notions of chance and baseline to assess presence and absence of knowledge; however, it is impossible to define chance and baseline performance for production data. Therefore, for assessing knowledge with production data, I rely on a different diagnostic, which compares the rate of correct responses with the rate of incorrect responses, based on the following assumptions:13 (a) If groups/participants have TL knowledge, they will produce more correct responses than incorrect responses. (b) If groups/participants do not have TL knowledge, they will

13 This idea came from Robert Bley-Vroman and Kamil Ud Deen.
produce more incorrect responses than correct responses. Let us apply this idea for assessing knowledge to the group production data and the individual production data.

For the group production data, I define presence of knowledge as when the group’s accuracy rate is significantly above the group’s inaccuracy rate; absence of knowledge as when the group’s inaccuracy rate is significantly above their accuracy rate (in other words, the accuracy rate is significantly below the inaccuracy rate); and possibility of knowledge as when the group’s performance fits the definition of neither presence of knowledge nor absence of knowledge. Figure 2.4 represents this logic graphically.

![Figure 2.4. Scope of performance to test the presence of knowledge: Production data.](image)

Now let us apply this logic to individual production data. Given that the number of tokens obtained from an individual participant is not large, it seems impossible to find a statistical difference between the accuracy rate and the inaccuracy rate. Therefore, I set cut-off points for individual production based on intuition. In this dissertation, I present production data in which a participant was given 6 tokens for subject-case-marking and 6 tokens for object-case-marking. To test knowledge of subject-case-marking and object-case-marking, for each context, presence of knowledge is operationalized as 5 or 6 correct (out of 6; 83.3%–100%); absence of knowledge is operationalized as 0 or 1 correct (out of 6; 0%–16.7%); and possibility of knowledge is operationalized as between 2 and 4 correct (out of 6; 33.3%–66.7%). Intuitively, out of 6 tokens, 5 or 6 correct is well above 1 or 0 incorrect (i.e., the number of accurate responses is well above the number of inaccurate responses) and 0 or 1 correct is well below 5 or 6 incorrect (i.e., the number of accurate
responses is well below the number of inaccurate responses). However, 2–4 correct is not very different from the number of incorrect responses (i.e., 4–2 incorrect).

2.5 Facilitating Performance and Testing the Presence of Knowledge

As discussed above, poor performance can be viewed in two ways: Poor performance may come from a deficit in knowledge, or it may come from a failure in the application of knowledge rather than an absence of knowledge. Two different predictions can be made according to the two approaches. If the source of poor performance is not a deficit of knowledge, significantly-above-chance performance should be obtained through the effective manipulation of non-grammatical factors to facilitate the learners’ application of knowledge. In contrast, if the source of poor performance is a deficit in knowledge, manipulation of non-grammatical factors would not affect obtaining significantly-above-chance performance.

One way of manipulating non-grammatical factors is to administer a cognitively undemanding task, which requires fewer processing resources. All tasks designed to assess knowledge require participants to engage in both linguistic and cognitive processes. If a task consumes too many resources with cognitive processes, the remaining resources may be insufficient for linguistic processes. Another way is to manipulate either sentence-internal factors or sentence-external factors to promote learners’ use of their knowledge, considering that comprehension (and production) involves both sentence-internal processing and sentence-external processing. These manipulations are expected to encourage the application of knowledge. By employing them, the current study tests whether or not the knowledge under investigation is present.

2.6 Summary of Chapter 2

This chapter introduced the theoretical debates between the RDA and the CDA on the competence vs. performance issue: In essence, the RDA attributes poor performance to a deficit in competence, while the CDA attributes it to a performance failure.

This chapter also evaluated two diagnostics for assessing linguistic competence through comprehension data: the comparison-against-chance diagnostic and the

---

14 This applies to comprehension tasks only.
comparison-against-baseline diagnostic. To claim presence of knowledge (+), I rely on the comparison-against-chance diagnostic. The comparison-against-baseline diagnostic distinguishes possibility of knowledge (?) from absence of knowledge (−). To assess linguistic competence through production data, however, the notions of chance and baseline are impossible to define. Therefore, I rely on the diagnostic which compares the rate of correct responses with the rate of incorrect responses for production data.

This chapter proposed a method to test the source of poor performance (i.e., a deficit in knowledge vs. a performance failure). It hypothesized that if poor performance is due to failure in the application of knowledge, not absence of knowledge, good performance should result with the effective manipulation of non-grammatical factors to facilitate learners’ application of knowledge. I test this, using experimental (a) sentence-internal manipulations and (b) sentence-external manipulations.

We now turn to Chapter 3, which defines different learner populations—L1 children, Heritage language children, L2 children, and L2 adults—and discusses how to compare the learner groups.
CHAPTER III
COMPARING DIFFERENT LEARNER POPULATIONS

This chapter addresses the issues involved in making comparisons between different populations, namely L1 children (child L1ers), Heritage language children (HLC), L2 children (child L2ers), and L2 adults (adult L2ers). I will first define each learner group, each of which is under investigation in this dissertation. I will then discuss issues on measuring TL proficiency, which is important when comparing different learner populations.

3.1 Defining Learner Populations

Child L1ers, child L2ers, and adult L2ers are defined based on their first exposure to a target language (TL): Child L1ers are exposed to their L1 from birth. For child L2ers, early studies have defined them as those whose first non-native TL exposure occurs between 0 and 7 (Johnson & Newport, 1989) or between 0 and 9 (Penfield & Roberts, 1959). However, later studies have distinguished child L2ers from simultaneous bilingual speakers (2L1ers), where child L2ers are those whose first non-native TL exposure occurs between the ages 4 and 7 (Schwartz, 2004). This range was chosen because children at the age of 4 know the bulk of their first language grammar (Guasti, 2002), and children who begin L2 acquisition before age 8 have shown native-like performance on tasks testing various morphosyntactic phenomena (Johnson & Newport, 1989, 1991). Schwartz (2004) claims that the native-like performance of these child L2ers suggests that they are “utilizing the same acquisition processes as children use in L1 acquisition” (p. 48).

The boundary between child and adult L2 acquisition is not clear-cut. A conservative strategy is to exclude participants whose first non-native TL exposure is between the age of 7 and 13, because “this period may be considered to mark the fuzzy boundary between children and adults” (Unsworth & Blom, 2010, p. 202). For Haznedar (2013), the boundary for adult L2ers is considered to be 7–10 years old. This is based on the debate with respect to the Critical Period Hypothesis (Lenneberg, 1967), which proposes that certain physiological changes in the human brain around the time of puberty preclude the
possibility of target-like language attainment (for an empirical study testing this hypothesis, among others, see Abrahamsson & Hyltenstam, 2009). That is, for Haznedar, adult L2ers are defined as those who start to learn the TL after the alleged critical period for language acquisition ends.

On the other hand, HLC are defined based on the place where they grow up. They are “child and adult members of a linguistic minority who grew up exposed to their home language and the majority language” (Montrul, 2010, p. 4). Under this definition, both 2L1ers and child L2ers can be called HLC if they are raised as a linguistic minority. For example, Korean immigrants living in the U.S. as children, no matter whether they were born in Korea or in the U.S., can be called HLC with respect to Korean. However, 2L1ers and child L2ers cannot be considered HLC if their home language is a majority language. For instance, if an L1-Korean child learns L2 English from an English-speaking kindergarten in Korea, that child is called a child L2er but not an HLC. However, to avoid an overlap between child L2ers and HLC, I only test HLC whose first exposure to the majority language occurs between 0 and 4.

See Figure 3.1 for a summary of these criteria: For the definition of child L2ers, I follow Schwartz (2004); for the definition of adult L2ers, I follow Haznedar (2013). However, to make the gap between child L2ers and adult L2ers clear in the array of participants, I exclude learners whose first TL exposure is between the age of 7 and 8. As for HLC, as discussed above, I consider them to be children whose first exposure to the majority language occurs between 0 and 4.
3.2 Grouping Learners by Proficiency

To compare different learner populations, it is necessary to group the learners by TL proficiency and compare the different learner populations within each proficiency group, to see whether the response patterns of one learner population are relatively parallel to those of another learner population. That is, measuring TL proficiency is crucial for comparing different learner populations.

I adopt an in-house assessment method, using a picture-narration task for proficiency measurement (Song & Schwartz, 2009; Unsworth, 2005; Whong-Barr & Schwartz, 2002), instead of, say, a traditional (written) cloze test often used in adult L2 studies, because it is inappropriate for children (Whong-Barr & Schwartz, 2002). In the picture-narration task, an L2 proficiency score is calculated by measuring both language complexity (mean length of T-unit) and grammatical accuracy (proportion of error-free T-units). Although complexity may be a satisfactory index for L1 acquirers, a combinatory measure of both complexity and accuracy is a better means of measuring L2 proficiency because the accuracy measure prevents learners’ proficiency from being overestimated (Larsen-Freeman, 1983, p. 295). In addition, I avoid using a proficiency test that is “explicitly designed to measure the abilities under investigation” (Thomas, 1994, p. 322). For example, knowledge of case markers and scrambling is necessary for understanding
the Korean OSV pattern; therefore, errors with case markers and scrambling should not be regarded as errors in an accuracy measure for a study on Korean OSV.

A challenge for an in-house assessment is deciding how to divide the participants into different proficiency groups. One way is to divide participants at a clear gap in the array of scores. For example, in Kobayashi and Rinnert’s (1992) study, the cut-off point for partitioning participants into high and low groups was “the only clear gap” (p. 206) in the distribution of scores. However, this method often fails to control group size; for example, Kobayashi and Rinnert had a low-proficiency group half the size of their high-proficiency group. The other way is to divide participants into high, mid, and low groups, “getting the cut-off points at whatever yield groups of approximately equal size” (Thomas, 1994, p. 323), which avoids the problem of asymmetrical sizes of participant groups. Thomas notes that while this way has been the most conventional way of grouping participants, the cut-off points are arbitrary. Thomas suggests that a less arbitrary means of grouping participants is on the basis of their score distributions, for example, as with Sasaki’s (1990, p. 345) method of creating groups of four proficiency levels: lowest (test scores more than one standard deviation [1SD] below the mean of the total pool); mid-lower (scores between the mean and 1SD below it); mid-higher (scores between the mean and 1SD above); and highest (test scores more than 1SD above the mean). However, Sasaki’s method is also disadvantageous in that the numbers of participants in the two middle groups are likely to be about twice as large as those in the high and low groups, because in a normal distribution the area between mean -1SD and mean +1SD constitutes 68.3% of the cases.15

In this dissertation, I follow the method of participant grouping used in Unsworth (2005). Unsworth follows the conventional method—dividing participants into low, medium, and high groups—but deciding the cut-off points in a way that is not arbitrary. She transformed the raw proficiency scores into z-scores and then partitioned her participants into low (z-score below -0.5), mid (z-score between -0.5 and 0.5), and high groups (z-score more than 0.5). Thus, each group includes a similar number of

15 In fact, the problem of defining proficiency groups is entirely avoided if raw proficiency scores are incorporated as a variable to predict the experimental performance data. However, this practice is often less informative in an examination of developmental patterns—how learners’ L2 performance changes from one stages to the next. Grouping the participants can be more informative when examining the differences at each developmental stage.
participants (e.g., z-scores between -0.5 and 0.5 constitute about 38% of the whole data set), and at the same time, participants are grouped rationally.

For the method of calculating proficiency scores from picture-narration production data, I follow Whong-Barr and Schwartz (2002) and Song and Schwartz (2009): The raw proficiency scores come from the participants’ scores on both a complexity measure and an accuracy measure. The complexity measure is the mean word length of T-unit. The accuracy measure is a percentage of error-free T-units.

3.3 Summary of Chapter 3

This chapter defined four language learner groups—child L1ers, HLC, child L2ers, and adult L2ers. It explained that child L1ers, child L2ers, and adult L2ers are defined based on their first exposure to a target language while HLC are defined based on the place where they grow up. It then discussed how to compare different learner groups, focusing on measuring learners’ proficiency.

We now turn to Chapter 4, which presents the details and results of the first study in which the two target learner groups are HLC and child L1ers of Korean.
This chapter is concerned with the acquisition and processing of the Korean OSV pattern. I will first describe the linguistic competence needed to process the Korean OSV pattern, which relies on knowledge of scrambling, abstract Case, and morphological case. I will then discuss performance-related sources of poor performance—perception failure, heavy processing load, and mapping problems. Subsequently, I will introduce the novel methods this study used to test knowledge of the Korean OSV pattern via comprehension and production tasks. Finally, I will report the results of experiments using these tasks with young native Korean-speaking children (henceforth, young cL1ers), older native Korean-speaking children (henceforth, old cL1ers), and Korean heritage-speaker children (henceforth, KHC).

4.1 Competence and Performance: The Korean OSV Pattern

Korean, a canonically SOV language, uses case\(^1\) to mark grammatical relations. However, on occasion, Korean speakers may omit case markers, relying on word order to signal syntactic relations (the first NP is the subject; the second is the direct object). Thus the case-marked SOV sentence in (1a) has the same meaning as the caseless sentence in (1b).\(^2\)

---

\(^1\) In this dissertation, “case” generally means morphological case whereas “Case” refers to syntactic Case.

\(^2\) However, in a transitive sentence, omitting both the nominative and accusative markers (as in (1b)) sounds somewhat unnatural.
(1) a. Wensungi-ka kay-lul cha.
   monkey-NOM dog-ACC kick
   ‘A monkey is kicking a dog.’

b. Wensungi kay cha.
   monkey dog kick
   ‘A monkey is kicking a dog.’

c. Wensungi-lul kay-ka cha.
   monkey-ACC dog-NOM kick
   ‘A monkey, a dog is kicking.’

Korean also allows scrambling, “the phenomenon of variable word order within a clause” (Corver & van Riemsdijk, 1994, p. 1), which makes case-marker information crucial for interpretation. Thus, thanks to case, the OSV sentence in (1c) has a different meaning from the sentence in (1b), despite the similarity in word order. Without the case marking, (1c) would be interpreted as ‘a monkey is kicking a dog.’

It has been claimed that scrambling and case marking can be separated in derivation (e.g., distributed morphology; Halle & Marantz, 1993). Scrambling is constrained in narrow syntax (Ko, 2007; Neeleman & Weerman, 2009; Saito, 1985); morphological case markers, triggered by syntactic Case features (in a transitive sentence, +T for the subject and +v for the object), are inserted at or after Spell-Out (N. Chomsky, 2001). Section 4.1.1 deals with grammar and processing in scrambling; Section 4.1.2 discusses knowledge of abstract Case and morphological case and their processing.

4.1.1 Grammar and processing: Scrambling.

The syntax of scrambling has been viewed in two ways: a “base generation approach” and a “movement approach” (Corver & van Riemsdijk, 1994, p. 1). The former approach claims that the variable word order comes from the free generation of constituents, assuming that there is no single basic order of constituents. The latter argues that the variety of alternate word order arrangements in a clause is the result of movement, derived from one underlying word order.
Neeleman and Weerman (2009) argued for the base generation approach, focusing on the role of morphological case in grammar. They compared three types of case systems—the “highly deflected” case system (e.g., Dutch; there is no dative-accusative distinction) (p. 282), the less deflected case system (e.g., German; dative-accusative distinction), and the rich morphological case system (e.g., Japanese, Korean)—and claimed that “scrambling across arguments is allowed as long as the argument appearing unexpectedly low in the structure carries a formal marker that identifies it” (p. 282).

However, it seems that the claim of the base generation approach does not fully explain a constraint in scrambling in Korean and Japanese, languages with rich morphological case system. The examples in (2) and (3) are Saito’s (1985), translated from Japanese into Korean by Ko (2007).

   John-NOM beer-ACC 3-CL bottle drink-PAST-DECL
   ‘John drank three bottles of beer.’

      beer-ACC John-NOM 3-CL bottle drink-PAST-DECL
      ‘John drank three bottles of beer.’
      (Ko, 2007, p. 50, (1a-b))

   student-PL-NOM 3-CL person beer-ACC drink-PAST-DECL
   ‘Three students drank beer.’

      student-PL-NOM beer-ACC 3-CL person drink-PAST-DECL
      ‘Three students drank beer.’
      (Ko, 2007, p. 51, (2a-b))

---

3 Neeleman and Weerman do not provide a definition for this term; however, the following excerpt indicates their idea on the differences between the three case systems: “Since Japanese has a formal nominative marker, scrambling across the subject is possible. The German case system does not allow this, but the presence of dative in German allows scrambling across the indirect object. In a highly deflected language like Dutch, no scrambling across arguments is possible as no identifying morphological markers exist” (p. 282).
A numeral quantifier (NQ) followed by a classifier expresses quantity. An NQ can be separated from its host NP, as in (2a-b); in (2b), the subject may intervene between the object and the object-oriented NQ (NQ_{obj}). In contrast, as in (3a-b), the object cannot intervene between the subject and the subject-oriented NQ (NQ_{subj}). Saito argues that the ungrammaticality of the Japanese analogue of (3b) is because the subject cannot move to the left of the moved object, as illustrated in (4).

\[
\begin{array}{c}
\text{[S Haksayng-tul-i]}_2 [s \text{ maykewu-lul} t_1 [s t_2 \text{ sey-myeng} [v \text{ masi-ess-ta.}]]]) \\
\text{student-PL-NOM beer-ACC 3-CLperson drink-PAST-DECL} \\
\text{‘Three students drank beer.’} \\
\end{array}
\]

(Ko, 2007, p. 51, (3))

Therefore, I consider scrambling in Japanese and Korean to be a movement operation in syntax, following Saito (1985) and Ko (2007). Scrambling is one way for flexible word-order languages to encode Information Structure. In Korean and Japanese, scrambled OSV sentences sound somewhat unnatural out of context, but are natural when the context makes the direct object topical. In other words, the scrambled (object) NP is defocused, and the subject carries focus, which corresponds to Information Structure that provides given information (the scrambled NP) first and then new information (the subject). This can be shown in pairs of a \textit{wh}-question and a subsequent response. Consider (5) and (6).

\footnotesize{4} Ko (2007) presents an alternative account for the ungrammaticality of (3b) based on the theory of Cyclic Linearization of syntactic structure at the phonology-syntax interface.

\footnotesize{5} It is also interesting to note that Korean OSV is not frequently produced by native speakers: Korean OSV sentences constitute only 1.2\% of the sentences in the Seyjong corpus (Kwon, Polinsky, & Kluender, 2006).
(5) a. Nwu-ka Elmo-lul ttayly-ess-ni?
who-NOM Elmo-ACC beat-PAST-Q
‘Who beat Elmo?’
b. Elmo-lul Bert-ka ttayly-ess-e.
Elmo-ACC Bert-NOM beat-PAST-SES
‘Elmo, Bert beat.’
Bert-NOM Elmo-ACC beat-PAST-SES
‘Bert beat Elmo.’

(6) a. Bert-ka nwukwu-lul ttayly-ess-ni?
Bert-NOM who-ACC beat-PAST-Q
‘Who did Bert beat?’
b. # Elmo-lul Bert-ka ttayly-ess-e.
Elmo-ACC Bert-NOM beat-PAST-SES
‘Elmo, Bert beat.’

In both (5b) and (6b), the subject is new and focused, while the object is old and non-focused (or topical). The scrambled sentence in (5b) is felicitous because (5a) asks about the subject, and thus the subject should be focused. However, the scrambled sentence in (6b) is infelicitous because (6a) asks about the object, and thus the object should be focused. It is worthwhile to note that between scrambled and non-scrambled sentences, this discourse effect only holds for scrambled sentences. Consider (5c), which is a non-scrambled sentence. (5a) asks about the subject, and thus the subject should be focused; however, in (5c), the subject is in the non-focused position. Despite this, (5c) is a natural way of responding to the question (5a).

The scrambled order is harder to process than the non-scrambled order, according to the prediction of the Minimal Chain Principle (De Vincenzi, 1991, p. 13): “Avoid postulating unnecessary chain member at S-structure, but do not delay required chain members.” This principle dictates that the parser constructs a structure as parsimoniously as possible, but in the scrambled OS order, the object NP has moved to the left of the
subject, which is structurally less parsimonious than the canonical SO order. That is, the scrambled order requires additional processing effort compared to the canonical order. The greater difficulty of processing the scrambled order than the canonical order has been reported in studies on a variety of languages (Bader & Bayer, 2006; Hopp, 2009; Hwang, 2008; Sekerina, 2003).

The processing of Korean OSV sentences with no context can be described as follows: Following the Minimal Chain Principle, the parser attempts to build an SOV structure because it contains minimal phrase structure. When the parser encounters the accusative-marked first NP, the parser would not immediately reanalyze the SO order to the OS order because (since Korean is a null-argument language) the parser can build a representation with a canonical SO order in which a null subject occurs in initial position. Only when the parser encounters the nominative-marked NP would the parser confidently reanalyze the SO order to the OS order. The processing cost of this reanalysis has been demonstrated by research showing a garden-path effect (i.e., reading slowdown) on the nominative-marked second NP in Korean OSV sentences (but not on the accusative-marked first NP) (Hwang, 2008).6

The processing cost of Korean OSV sentences can be alleviated with the use of felicitous contexts. When a context makes the first NP of an OSV sentence topical, an SO order sentence with a null subject is unnatural. That is, the parser would not build a representation with a canonical SO order in which a null subject is placed in the subject position. This suggests that when a context makes the first NP of the scrambled sentence topical, the parser would reanalyze the SO order to the OS order upon encountering the accusative-marked first NP. Hwang’s (2008) online self-paced reading experiment with native Korean-speaking adults showed that whereas participants processed OSV sentences more slowly than canonical SOV sentences, their reading of OSV was facilitated by a context that made a sentence-initial direct object more topical. This context effect appeared on both the first NP and the second NP of scrambled OSV (as well as SOV) sentences, with faster reading times on the two NPs when the context made the first NP serve as the topic (rather than the focus).

---

6 Different results, however, have been found for the processing of scrambled OS order in German; German is not a null subject language, and the garden path effect appears on the first, accusative-marked NP (Bayer & Bader, 2006).
4.1.2 Grammar and processing: Case.

Case marks a structural position in a tree. It is distinct from theta roles, which denote thematic relations. Theta roles are assigned to participants in an event. For example, consider (7a-e).

(7) a. I pushed him.
   b. He was pushed.
   c. The door opened.
   d. I opened the door.
   e. The gift arrived.

The masculine third person singular is the theme in (7a-b). However, it is in the object position in (7a) but in the subject position in (7b). This is also seen in causative/inchoative alternations, as in (7c-d). In (7c-d), the *door* is the theme, but in (7c) it is in the subject position, whereas in (7d) it is in the object position. In an unaccusative construction, as in (7e), the theme (i.e., the *gift*) is placed in the subject position. In short, the function of Case is purely syntactic and thus uninterpretable. Abstract Case is assigned to a structural position without there necessarily being any theta relation between the assigner and the assignee; hence, abstract Case is independent from the choice of a particular predicate or other lexical head.

Morphological case has been viewed as a spell-out version of abstract Case. That is, morphological case is how abstract Case is morphologically realized or inserted at or after Spell-Out. The Case features that regulate the syntactic distribution of NPs are the same features that are overtly realized as morphological case. See (8a-b) for example.

(8) a. The butler attacked the robber.
   b. *He* attacked *him*.

   (Haegeman, 1994, p. 155, (1))

In (8a), the subject is *the butler* and the direct object is *the robber*. There is no morphological marking to signal Case. In (8b), the subject is the pronoun *he* where
nominative Case is marked; the object is the accusative marked *him*. The marking of nominative and accusative in (8b) suggests that abstract Cases are assigned to the subject and object in (8a), although there are no morphological markings in the sentence. In fact, the Case Filter proposes that all overt NPs must be assigned abstract Case.

In a Korean active transitive sentence, the nominative marker -i/-ka occurs on the subject while the accusative marker -ul/-lul occurs on the direct object. These morphological case markers are instances of morphological case. They are believed to be the morphological realization of abstract Case assigned by Case assigners, as evident in passive sentences. As one of the properties of passive constructions is that the verb loses its ability to assign accusative Case to its argument, the accusative marker -ul/-lul cannot appear. See (9a-c) for examples.

    John-NOM Mary-ACC hug-PAST-DECL
    ‘John hugged Mary.’

    Mary-NOM John-ACC hug-PASS-PAST-DECL
    ‘Mary was hugged by John.’

    Mary-NOM John-DAT hug-PASS-PAST-DECL
    ‘Mary was hugged by John.’

(9a) is a transitive active sentence. The transitive verb an- ‘hug’ takes a subject and a direct object. The nominative marker -i occurs with the subject John while the accusative marker -lul occurs with the object Mary. (9b) is a passive sentence in which the thematic relations are the same as (9a). The nominative marker -ka occurs on the subject Mary. The accusative marker -lul cannot be used with John because the passive verb an-ki ‘be hugged’ lost its ability to assign the accusative Case. The dative marker -hanthey can occur with John, as in (9c).

In recent theories of syntax, the connection between abstract Case and morphological case has become more tenuous following the claim that syntax and morphology are
separated in derivation (e.g., distributed morphology; Halle & Marantz, 1993). It has
been argued that morphological case is triggered by syntactic Case features (in a
transitive sentence, +T for the subject and +v for the object) in narrow syntax, and
inserted at or after Spell-Out (N. Chomsky, 2001). In this view, morphological case is
related to abstract Case, but abstract Case is not the sole factor for determining
morphological case. Look at instances of so-called quirky case in Korean. For example,
consider (10a-b).

    John-DAT Mary-NOM good-DECL
    ‘John likes Mary.’

    John-DAT Mary-NOM feared-DECL
    ‘John is scared of Mary.’

In (10a), the predicate coh- ‘good’ takes two thematic roles, experiencer and theme. In
(10b) the predicate mwusep- ‘feared’ takes two thematic roles, experiencer and stimulus.
Here, the marker -ka co-occurs with the non-subjects—namely, Mary. If -ka is
exclusively the morphological realization of nominative Case, Mary should be assigned
nominative Case from +T and thus appear in the subject position. However, the subject of
the sentences is the dative-marked NP (i.e., John).7 This can be tested on the basis of
honorification agreement, which is a reliable subjecthood test in Korean (Hong, 1994).
Consider (11a-b) and (12a-b).

7 In fact, the nominative -ka/-i can also appear on John, the subject. However, when this happens,
the marker -ka/-i occurs twice in a single sentence, as in (ia-b).

    John-NOM Mary-NOM like-DECL
    ‘John likes Mary.’

    John-NOM Mary-NOM feared-DECL
    ‘John is scared of Mary.’
    professor-DAT Mary-NOM like-HON-DECL
    ‘The professor likes Mary.’
    John-DAT professor-NOM like-HON-DECL
    ‘John likes the professor.’

    professor-DAT Mary-NOM feared-HON-DECL
    ‘The professor is scared of Mary.’
    John-DAT professor-NOM feared-HON-DECL
    ‘John is scared of the professor.’

Both in (11a-b) and (12a-b), the honorific marker -si can agree only with kyoswunim ‘professor’. This confirms that kyoswunim ‘professor’ is the subject even though it is marked with the dative -hanthey, and that Mary is not the subject although it is marked with -ka.

It is noteworthy that the predicate coh- ‘good’ or mwusep- ‘feared’ in Korean is an intransitive verb (or adjective). As such, it does not allow accusative Case. The quirky case disappears if the predicate becomes a transitive verb. See (13a-c).

    John-NOM Mary-ACC good-do-DECL
    ‘John likes Mary.’
    John-NOM Mary-NOM good-do-DECL
    John-DAT Mary-NOM good-do-DECL
By adding the light verb *ha*- ‘do’, the intransitive predicate *coh* ‘good’ becomes the transitive verb *cohaha*- ‘like’. Thus, the complex verb assigns accusative Case, and the accusative -*lul* occurs with the direct object, as seen in (13a); the direct object, *Mary*, cannot take the marker -*ka*, as in (13b). In addition, the dative -*hanthey* is no longer allowed to occur on the subject, *John*, as in (13c).

The accusative marker -*ul/-lul* also does not always mark only the direct object, as can be seen from the use of the marker -*ul/-lul* with a non-argument. See (14a-b) for an example.

\[
\text{(14) Chelswu-ka kwail-ul sakwa-lul mek-ess-ta.}
\]

Chelswu-NOM fruit-ACC apple-ACC eat-PAST-DECL

‘Chelswu ate an apple (as for fruit).’

In (14a), the transitive verb *mek*- ‘eat’ is a two-place predicate. The external argument is *Chelswu* and the internal argument is *sakwa* ‘apple’; the nominative -*ka* and the accusative -*ul/-lul* occur on *Chelswu* and *sakwa* ‘apple’, respectively. However, note that *kwail* ‘fruit’ also occurs with -*ul*. If -*ul* is exclusively a realization of accusative Case, where is it assigned from? The transitive verb *mek*- ‘eat’ has already assigned accusative Case to *sakwa* ‘apple’. If Korean case is always a spell-out version of abstract Case, the example in (14) is unexplainable.

Sim (2005) claims that case morphology is basically a realization of abstract Case but can be used as a morphological resource to mark a topic within VP (see Schütze, 2001, for a relevant account). According to Sim, when -*ul/-lul* appears with a non-argument as in (14), it is not case morphology but merely a use of the morphological resource -*ul/-lul* to link a topic that occurs within VP to a syntactic constituent that corresponds to the comment part of the topic. That is, -*ul/-lul* is used to mark the topical property of *kwail* ‘fruit’ within VP. Sim claims that topics in Korean can occur in two different positions: in the domain of CP (topics with the topic marker -(n)un in the canonical topic position) and in the domain of VP (topics with the marker -*ul/-lul*), noting that Frey (2004) proposes that topics in German can occur in two different positions: in the domain of CP and in the domain of TP. See (15) and (16a-b).
(15) *Mwues-un Chelswu-ka sakwa-lul mek-ess-ni?
        what-TOP Chelswu-NOM apple-ACC eat-PAST-Q

‘As for what, Chelswu ate apples?’

(Sim, 2005, p. 107, (13))

(16) a. Chelswu-ka kwail-ul mwues-ul mek-ess-ni?
        Chelswu-NOM fruit-ACC what-ACC eat-PAST-Q

‘As for fruits, what did Chelswu eat?’

b. *Chelswu-ka mwues-ul sakwa-lul mek-ess-ni?
        Chelswu-NOM what-ACC apple-ACC eat-PAST-Q

‘As for what, Chelswu ate apples?’

(Sim, 2005, p. 107, (14))

Sim argues that the ungrammaticality of (15) and (16b) is due to the fact that mwues
‘what’, which cannot be a topic, is placed in the topic positions in the CP domain and VP
domain, respectively. In contrast, in (16a), mwues ‘what’ is in situ, not in the topic
position of the VP domain, which makes the sentence grammatical.

The use of case markers is also relevant to focus. Alternative semantics (Rooth, 1992)
defines focus as follows: “Focus indicates the presence of alternatives that are relevant
for the interpretation of linguistic expression (Féry & Krifka, 2008, p. 4).” Lee (2007,
2011) reported that case markers in Korean are rarely dropped when the argument they
mark is contrastively focused. For example, Lee (2011) found that native Korean adults
gave low acceptability ratings to sentences in which the object is contrastively focused
and case-elided. This has also been proposed for Japanese. Masunaga (1988, p. 147)
observed that whenever the pertinent NP is “deemphasized” or “defocused,” the case
marker can be deleted. In addition, the nominative case marker -ga in Japanese cannot be

---

8 According to Féry and Krifka (2008, p. 1), “the meaning of a question identifies a set of
alternative propositions, the answer picks out one of these, and the focus within the answer
signals the alternative propositions inherent in the question.” They give the following example of
answers to the question “What did you see on the road?”

(ii) a. We saw a TIGER on the road.
    b. #We saw a tiger on the ROAD.
dropped when the referent it marks is contrasted with other objects of the same type (Yatabe, 1999).

According to McFadden (2004, p. 9), there are three possible explanations for the extent to which “the DP-licensing condition known as syntactic Case is related to morphological case-marking”:

(a) “Morphological case is the direct spell-out of syntactic Case features.”
(b) “Morphological case is related to syntactic Case features but can also be affected by other factors.”
(c) “Morphological case is determined without reference to syntactic Case.”

In this dissertation, I follow the second explanation, which is compatible with N. Chomsky’s (2001) view. Morphological case is related to abstract Case, as seen, e.g., from the unavailability of the accusative -ul/-lul in a passive construction, as in (9b); but, other factors (e.g., intransitive verbs assigning two thematic roles; focus) affect the mapping of abstract Case onto case morphology.

Processing case is not simple. Syntax and morphology (and often, discourse) interact with each other in such processing. The parser extracts the syntactic information (e.g., subject, object) that the case markers deliver. However, the parser does not extract the accusative information when -ul/-lul occurs with non-arguments because it is merely a use of case morphology to deliver the topical property of a non-argument, as Sim (2005) points out. In addition, discourse properties (e.g., contrastive focus) are relevant to case morphology, as discussed by Lee (2007, 2011). In producing case markers, the producer must choose correct case markers for syntax-morphology mapping, out of tens of postpositional particles (e.g., -to ‘also’, -man ‘only’). In addition, the producer does not drop the case marker when the pertinent NP is contrastively focused.

---

9 Bonnie D. Schwartz raises an interesting question as to how the parser knows this before it gets to the verb in final position.
10 Sohn (1999, pp. 212-215) lists 46 case particles, 19 delimiters, and 1 conjunctive particle. This number excludes phonological variants of a particle such as when the nominative marker -ka is realized as -i following a noun ending with a consonant.
4.2 Previous Acquisition Studies

Children acquiring their native language appear to master appropriate use of case markers and word order together by around age 3;6 (S. Cho et al., 2002; S. Kim, O'Grady, & Cho, 1995; Y.-j. Kim, 1997; Slobin & Bever, 1982). Native Korean-speaking children start producing Korean case markers at a quite early age. Based on a study of 5 monolingual children, Y.-j. Kim (1997) reported that the first case marker to emerge in production is the nominative marker -ka (between 1;8 and 2;0). The accusative marker -lul emerges slightly later; 3 children in her study began to produce it between 1;11 and 2;3, and the remaining 2 children did so between 2;6 and 2;8. Chung (1994) suggested a developmental sequence for the nominative and accusative case markers in Korean: (a) nominative marker only (Stage 1); (b) overextension of nominative marker to non-subject NPs (Stage 2); (c) nominative marker to the first NP and accusative marker to the second NP (Stage 3); and (d) adult-like use with occasional errors (Stage 4). Here, it is worthwhile to note that children overgeneralize the use of the nominative marker, using it for the accusative. This is different from the developmental sequence in the acquisition of English where children often use accusative pronoun subjects until age 2;0 (Schütze, 1997). Schütze (1997) argued that such overgeneralization occurs because the default case in English is accusative.

Korean case markers and argument NPs may be dropped relatively freely when the information is recoverable from contexts. Arguments may be realized in three different forms: (a) case-marked overt NPs, (b) zero-marked overt NPs, and (c) phonetically null NPs (Clancy, 1995). In the speech of the children in Y.-j. Kim’s (1997) study, the rate of accusative case drop was much higher than that of nominative case drop. Cho (1981) reported that adults are more likely to drop accusative markers than nominative markers, based on longitudinal data from 3 mother-child dyads: The mothers used nominative case on subjects a little more than half the time (56.8%); the accusative was used for overt direct objects even less frequently—about 10% of the time (pp. 45–46).

The work by Cho (1981) and by Chung (1994) suggests that monolingual Korean children are able to interpret scrambled OSV sentences by approximately 4 years of age. Prior to that time, there is a tendency to interpret the first NP as subject and the second NP as object (Cho, 1981). Cho reported that the mothers dropped nominative case for the other 43.2% of the time.
NP as direct object. For example, in Chung's (1994) act-out comprehension task, 24 native Korean-speaking children were divided into four groups by age: Group 1 (2;8–3;0), Group 2 (3;3–3;5), Group 3 (3;7–3;10) and Group 4 (4;1–4;6). The rates of accurate responses for canonical SOV and scrambled OSV sentences (e.g., *hama-lul yemso-ka chaeseyo* ‘The hippo, the goat kicked’, p. 188, (B6)) are as follows: SOV 67% vs. OSV 33% in Group 1; SOV 83% vs. OSV 50% in Group 2; SOV 100% vs. OSV 72% in Group 3; and SOV 100% vs. OSV 78% in Group 4.

In Korean and Japanese, scrambled OSV sentences sound somewhat unnatural out of context, but are natural when the context makes the direct object topical. Otsu (1994) was the first to test whether context helps children understand scrambled sentences. He reported that in an act-out task, a context sentence helped monolingual Japanese children (ages 3–4) understand OSV sentences. This suggests that children may actually have knowledge of scrambling and case but are poor at using that knowledge in the absence of an appropriate context. S. Kim, O’Grady and Cho (1995) identified a similar effect in Korean-monolingual children (*n* = 68, ages 2–8). Using a picture-selection task, they found that the presence of context led to a noticeable improvement in the children’s comprehension of OSV sentences.

Song et al. (1997) asked whether KHC in the United States (*n* = 28, ages 3–8) understand scrambled Korean OSV sentences and whether a context sentence would bring about improvement in comprehension. In their picture-selection test, however, the KHC showed poor comprehension of OSV sentences, regardless of the presence or absence of a context sentence. In the no-context condition, the accuracy rates were 34.8% for OSV and 75.9% for SOV; in the context condition, the accuracy rates were 29.5% for OSV and 81.3% for SOV. In addition, Song et al. reported data from 13 older KHC (ages 7–12) who were given a picture-selection task with no context. Their accuracy rates were 25% for OSV and 94% for SOV. However, the older participants did not complete a picture-selection task with context. This is because the task with these children was intended to test an instructional effect, not a context effect.

---

12 In addition, Song et al. reported data from 13 older KHC (ages 7–12) who were given a picture-selection task with no context. Their accuracy rates were 25% for OSV and 94% for SOV. However, the older participants did not complete a picture-selection task with context. This is because the task with these children was intended to test an instructional effect, not a context effect.
<table>
<thead>
<tr>
<th>Spoken stimuli</th>
<th>Visual stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Context</strong></td>
<td><img src="image1.png" alt="Image" /></td>
</tr>
<tr>
<td>Yeki wensungi isse. Kwiyepci? here monkey exist be cute</td>
<td>‘Here is a monkey. He is cute, isn’t he?’</td>
</tr>
<tr>
<td><strong>Test sentence</strong></td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td>I wensungi-lul kay-ka cha. This monkey-ACC dog-NOM kick</td>
<td>‘This monkey, a dog is kicking.’</td>
</tr>
</tbody>
</table>

*Figure 4.1. An example context-condition item (Song et al., 1997, p. 114, (6)).*

Song et al.’s (1997) results can be understood in either of two ways: (a) the KHC may have deficits in relevant knowledge of either scrambling or the distinguishing function of case, or (b) the KHC may have such knowledge, but were unable to use it effectively in online comprehension. For the latter account, three factors might have contributed to the KHC’s failure to use knowledge that they had. The first is that the children might have failed to perceive case markers because the markers were relatively weak acoustically (O’Grady, Kwak, Lee, & Lee, 2011). If the case markers in an OSV sentence such as (17a) were not perceived, the children would “hear” the utterance as if it were the canonical SOV pattern in (17b). This in turn would create the impression that they did not have adult-like knowledge of (the distinguishing function of) case when in fact the issue is one of perception, not knowledge.

(17) a. Elmo-lul Big Bird-ka ana-yo.
   Elmo-ACC Big Bird-NOM hug-SES
   ‘Elmo, Big Bird is hugging.’

b. Elmo Big Bird ana-yo.
   Elmo Big Bird hug-SES
   ‘Elmo is hugging Big Bird.’
Second, the children’s failure to apply their knowledge might be due to the context used in Song et al. (1997), which may not have been as felicitous as the authors anticipated. For example, let us revisit Figure 4.1: “Here is a monkey. He is cute, isn’t he? This monkey, a dog is kicking.” Pointing out that a monkey is cute does not relate to the consequence that a dog is kicking the monkey. That is, this context may not be sufficient to meet sense-plausibility. It might be possible that the children spent their processing resources to figure out why the one animal suddenly kicked the other animal. Third, the KHC’s poor performance may have been caused by a problem of mapping the case markers onto a Case feature. This mapping may not be simple, given that syntactic Case features are not the sole factor determining morphological case.

4.3 Methodological Innovations

This section introduces the methodological innovations that were used to adapt Song et al.’s (1997) picture-selection comprehension task for the experiments reported in this dissertation. The purpose of the innovations was to facilitate perception of case markers and to create felicitous contexts. In addition, a case-marker-elicitation task, which was designed to test knowledge of case and examine the mapping of a Case feature to a particular case marker, is described.

4.3.1 Picture-selection comprehension tasks.

For this dissertation research, I made two methodological innovations to adapt the picture-selection task employed by Song et al. (1997): One is an attempt at sentence-internal facilitation and the other is an attempt at sentence-external facilitation. First, I manipulated the prosodic salience of the case markers in the test sentences (for both OSV and SOV sentences). Because prominence is signaled by a number of acoustic features, including a rise in fundamental frequency (F0), an increase in intensity, and an increase in duration (Ladd, 1996), I manipulated the pitch, intensity, and duration of the case markers using Praat (Boersma & Weenink, 2008) and Goldwave. In order to ensure that this manipulation did not result in unnatural sentences, I conducted an analysis of the speech production of 6 native Korean-speaking adults, and I elicited acceptability judgments from 94 native Korean adults. Based on these two norming procedures,
I manipulated the case markers by increasing the pitch by a factor of 1.1, the intensity by a factor of 2, and the duration by a factor of 2 (see Appendix A for the details).

For the sentence-external manipulation, I provided a felicitous context to make scrambled OSV sentences sound more natural, adapting the contexts in Hwang’s (2008) study, but also using slides including color and animation. This context was also used for SOV sentences. See Figure 4.2 for a sample context and test sentence.

<table>
<thead>
<tr>
<th>Spoken stimuli</th>
<th>Visual stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Context</strong></td>
<td></td>
</tr>
<tr>
<td>Na-nuncikum konghang-ey isse-yo.</td>
<td>![Context Image]</td>
</tr>
<tr>
<td>I-TOP now airport-at stay-SES</td>
<td></td>
</tr>
<tr>
<td>‘I am in an airport now.’</td>
<td></td>
</tr>
<tr>
<td>Ceki Big Bird-nun pihayngki-eyse nayly-esse-yo.</td>
<td>![Context Image]</td>
</tr>
<tr>
<td>there Big Bird-TOP airplane-from land-PST-SES</td>
<td></td>
</tr>
<tr>
<td>‘There, Big Bird landed in his plane.’</td>
<td></td>
</tr>
<tr>
<td><strong>Test sentence</strong></td>
<td></td>
</tr>
<tr>
<td>(OSV item)</td>
<td>![Test Image A]</td>
</tr>
<tr>
<td>I Big Bird-lul Elmo-ka ana-yo.</td>
<td>![Test Image B]</td>
</tr>
<tr>
<td>this Big Bird-ACC Elmo-NOM hug-SES</td>
<td></td>
</tr>
<tr>
<td>‘This Big Bird, Elmo is hugging.’</td>
<td></td>
</tr>
<tr>
<td>(SOV item)</td>
<td></td>
</tr>
<tr>
<td>I Big Bird-ka Elmo-lul ana-yo.</td>
<td></td>
</tr>
<tr>
<td>this Big Bird-NOM Elmo-ACC hug-SES</td>
<td></td>
</tr>
<tr>
<td>‘This Big Bird is hugging Elmo.’</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 4.2. An example context-condition item in the current study.*

In the first context sentence, the speaker introduces the setting: an airport. Note that this is relevant to the upcoming hugging event between *Big Bird* and *Elmo* in the test sentence. Hugging is often used in greeting or parting, and so an airport is a place where it is frequently observed. The second context sentence is to make the first NP in the scrambled test sentence topical. Because the first NP in the subsequent scrambled test sentence has
already been mentioned in the preceding sentence, the scrambled sentence sounds more natural; the object NP is a discourse topic. In addition, in order to make Big Bird still more prominent, I depicted him in color while Elmo is de-emphasized in black and white. Furthermore, in the animation of this item, Big Bird moves his head from side to side, while Elmo is stationary. These manipulations were expected to lead the participants to concentrate more on Big Bird than on Elmo as they listened to the context sentences, making Big Bird the more natural candidate for the referent of the first NP in the scrambled sentence.

The use of a felicitous context reduces the processing burden. As reported in Hwang’s (2008) online processing study, the processing burden for the scrambled Korean OSV sentences can be alleviated by a context where the first NP is a topic. In addition, the sense-semantically appropriate context minimizes the possibility that participants spend their processing resources to figure out why a certain situation is happening (e.g., a dog suddenly kicking a monkey; Song et al. 1997).

4.3.2 Case-marker elicitation production task.

I carried out a production task to investigate knowledge of case and the mapping of a Case feature to a case marker. Participants’ consistently poor performance on the comprehension task, regardless of prosody or context manipulations, might stem from a deficit in knowledge of scrambling, or a deficit in case, or a problem in mapping between morphological case and syntactic Case. To determine the cause, I examined whether case markers were produced and whether they were correct. If case markers are not produced at all, one reason would be a deficit in morphological case. If case markers are produced but they are often incorrect, this may indicate a mapping problem. If case markers are produced correctly, the poor performance can be attributed to other reasons (e.g., a deficit in scrambling).

To address this issue, a case-marker-elicitation task was created. This task made use of the fact that case markers in Korean cannot be dropped when the argument they mark is contrastively focused; replacing focus on a definite and animate referent is the strongest in terms of the contrastive strength (Lee, 2007, 2011). See Figure 4.3 for examples.
<table>
<thead>
<tr>
<th>Case-marker elicitation</th>
<th>Spoken stimuli and expected participant production</th>
<th>Visual stimuli</th>
</tr>
</thead>
</table>
| **Subject**-marker elicitation | **Researcher says:**
Big Bird-*nun* Elmo mile-yo.
Big Bird-TOP Elmo push-SES
‘Big Bird is pushing Elmo.’

**Participant says:**
Ani, Cookie Monster-*ka* Elmo-lul mile-yo.
no Cookie Monster-NOM Elmo-ACC push-SES
‘No, Cookie Monster is pushing Elmo.’ | ![Image](image1.jpg) |
| **Object**-marker elicitation | **Researcher says:**
Cookie Monster-*nun* Big Bird mile-yo.
Cookie Monster-TOP Big Bird push-SES
‘Cookie Monster is pushing Big Bird.’

**Participant says:**
Ani, Cookie Monster-*ka* Elmo-*lul* mile-yo.
no Cookie Monster-NOM Elmo-ACC push-SES
‘No, Cookie Monster is pushing Elmo.’ | ![Image](image2.jpg) |
| **Training session** | **Researcher says:**
Pakwuni an-ey koyangi twu mali isse-yo?
basket in-at cat two CL exist-SES
‘Are there two cats in the basket?’

**Participant says:**
Ani, han-mali iss-eyo.
no one-CL exist-SES
‘No, there is one (cat).’ | ![Image](image3.jpg) |

**Figure 4.3.** Examples of the case-marker-elicitation task.

For the subject-marker condition in the case-marker-elicitation task, I showed participants a picture depicting an event between 2 animate characters, then presented an incorrect description of the subject NP. For example, when the picture depicted *Cookie Monster* pushing *Elmo*, I incorrectly described the scene by saying (in Korean) “Big Bird is pushing Elmo,” using a different character name for the subject. Here, I used the topic marker -*nun* rather than the nominative -*ka* for the subject NP and dropped the accusative marker -*lul* on the object NP to avoid the use of nominative and accusative cases.\(^{13}\)

Participants were expected to respond with a correct description, such as (in Korean) “No,

\(^{13}\) Topic marker –*nun* is used on the subject NP because it sounds rather unnatural when both the subject NP and object NP lack case markers.
Cookie Monster is pushing Elmo.” Here, the nominative -ka (or the topic -nun) is necessary with the subject NP, Cookie Monster. Likewise, for the object-marker condition, I gave an incorrect object NP in my description of the picture. A training session was also included before the main task, during which, for example, a picture of one cat in a basket was shown to the participants. Participants were asked “Are there two cats in the basket?” and gave a response such as “No, there is only one cat.” Participants were asked to use full sentences when they answered. Care was taken not to use nominative or accusative case in any of the instructions.

4.4 Research Questions and Hypotheses

This research addresses two main research questions:

1. When learners of Korean (i.e., KHC and Korean cL1ers) show poor performance with scrambled OSV sentences, what is the source of the poor performance?
2. Do KHC show evidence of a knowledge deficit in case?¹⁴

Specific predictions for the experiments can be stated as follows:

1. If participants have knowledge of scrambling and case, the manipulation of prosody should help them perceive the case markers, and the manipulation of context should help them apply their knowledge of scrambling by reducing the processing burden; therefore, participants will show significantly-above-chance performance on scrambled OSV sentences in the picture-selection comprehension task. If they lack knowledge of either scrambling or case, manipulations of prosody and context will not increase their OSV comprehension accuracy.
2. If participants know case, they will show good performance on the case-marker-elicitation tasks (i.e., significantly above the inaccuracy rate).

4.5 The Current Study¹⁵

Section 4.5.1 is concerned with the picture-selection comprehension tasks, testing knowledge of scrambling and case in learner groups. Section 4.5.2 deals with the

¹⁴ Throughout this dissertation, knowledge of case generally means knowledge of the distinguishing function of case. See footnote 1 in Chapter 1.
¹⁵ Portions of this study were reported in K. Kim, O’Grady, and Deen (in press).
case-marker elicitation task, where knowledge of case is assessed in learner groups. Section 4.5.3 evaluates the source(s) of poor performance for each individual. Section 4.5.4 discusses the results, addressing the research questions.

The order of tasks and their timing are as follows: first the set of picture-selection tasks; next, the case-marker-elicitation task; last, a picture-narration task (for assessing TL proficiency). For KHC, the interval between the picture-selection tasks and the case-marker-elicitation task was about 3 months. About 1 month later, the picture-narration task was conducted.16 For Korean cL1ers, the picture-selection tasks were administered 1 or 2 weeks before the other 2 tasks, and these were completed on the same day.

4.5.1 Picture-selection comprehension tasks.

4.5.1.1. Participants.

Thirty-one KHC (ages 8–14), 21 young cL1ers (ages 3;6–6;0), and, as a control group, 23 older cL1ers (ages 10–11) participated in this study. The KHC, recruited from a Saturday Korean-language school in Hawai‘i,17 were mostly born in the United States18 and had Korean-speaking parents.19 All KHC reported being more comfortable speaking English than Korean. The young cL1ers were recruited from a kindergarten in South Korea, and the older cL1ers were recruited from an elementary school in South Korea; none of the cL1ers had lived in an English-speaking country for more than 3 months.

To measure Korean proficiency, a picture-narration task was used: The children were presented with 3 sets of 4 pictures depicting a series of events (see Appendix B), and were asked to tell stories based on the pictures (Song & Schwartz, 2009). During the task, the children were encouraged to speak as much as possible: For instance, the researcher provided prompts and asked questions such as “And what happened next?” (Unsworth,

---

16 It is unlikely that a KHC would have a substantial improvement in his/her Korean proficiency in 4 months. For very young native children, a substantial improvement is possible in 4 months. Notably, the correlation results shown in p. 50 shows that age is not a predictor for the TL proficiency of the KHC.

17 These children were the same as those in K. Kim et al. (in press). That study reported the comprehension data of 34 KHC. However, this dissertation excluded 3 KHC, because we could not collect proficiency data from them.

18 Four KHC were born in Korea and had immigrated to the U.S. before age 4.

19 One KHC had a non-Korean parent, a Japanese father. However, he spoke English to the child.
The production data from the picture-narration task were used to measure proficiency, by calculating the participants’ scores of complexity (total number of words divided by the total number of T-units) and accuracy (the number of error-free T-units divided by the total number of T-units). (See Appendix B for how the proficiency scores were calculated).

The KHC and young cL1ers were divided into 3 groups based on their proficiency scores, calculated using the protocol detailed in Chapter 3: Low (z-score below -0.5), Medium (z-score between -0.5 and 0.5) and High (z-score higher than 0.5). Table 4.1 shows detailed information for each proficiency group.

Table 4.1
Participant grouping by proficiency

<table>
<thead>
<tr>
<th></th>
<th>Low (z-score &lt; -0.5)</th>
<th>Medium (-0.5 &lt; z-score &lt; 0.5)</th>
<th>High (0.5 &lt; z-score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KHC</td>
<td>Number</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>z-score mean (SD)</td>
<td>-1.53 (0.75)</td>
<td>0.09 (0.32)</td>
</tr>
<tr>
<td></td>
<td>Mean age (Age range)</td>
<td>12;1 (9-14)</td>
<td>12;0 (8-13)</td>
</tr>
<tr>
<td>Young cL1ers</td>
<td>Number</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>z-score mean (SD)</td>
<td>-0.71 (0.03)</td>
<td>-0.14 (0.28)</td>
</tr>
<tr>
<td></td>
<td>Mean age (Age range)</td>
<td>3;11 (3;8-4;3)</td>
<td>4;9 (3;6-6;0)</td>
</tr>
</tbody>
</table>

A Pearson correlation showed that for the young cL1ers there was a strong correlation between age at testing and proficiency score (correlation coefficient = 0.66). However, there was no significant correlation between age and proficiency score for the KHC (correlation coefficient = -0.11). This indicates that age is a predictor for TL proficiency for the young cL1ers, but not for the KHC.

4.5.1.2. Materials.

The experiment included 3 picture-selection tasks: a baseline task with no manipulation, a task in which prosody was manipulated to make case markers prominent, and a task in which context was manipulated to increase the naturalness of the OSV sentences. In each task, participants were presented with 28 stimuli consisting of 7
canonical SOV sentences, 7 scrambled OSV sentences, and 14 non-transitive filler sentences. In addition, 3 practice items were used to train the children in the task, which involved listening to a sentence and then choosing one picture (from two options) that best illustrated the sentence’s meaning. All the items, including test items, fillers, and practice items, can be found in Appendix C. The presentation sequence was randomized; however, care was taken so that no 2 consecutive items included the same verb and no 3 consecutive items included the same word order. This was done to avoid carry-over from one trial to the next. The target pictures were counterbalanced between left and right for SOV items, OSV items, and fillers. Half of the participants were given the sentences in the initial randomized order, while the other half were given the sentences in reversed randomized order. The following are examples of the experimental sentences (see Figure 4.4 as well).

---

The following are examples of the filler sentences:

(iii) a. Yeki, Elmo isseyo.
    here, Elmo exist
    ‘Here is Elmo.’

(III) b. Na-nun Bert-ipnita.
    I-TOP Bert-is
    ‘I am Bert.’
(18) a. Canonical SOV

Big Bird-ka    Elmo-lul    ana-yo.
Big Bird-NOM   Elmo-ACC   hug-SES

‘Big Bird is hugging Elmo.’

b. Scrambled OSV

Big Bird-lul     Elmo-ka     ana-yo.
Big Bird-ACC     Elmo-NOM   hug-SES

‘Big Bird, Elmo is hugging.’

Figure 4.4. An example of the picture-selection task (used for the baseline and prosody tasks; for the context task, see Figure 4.2 in p. 45).

As they listened to the test items, the participants saw slides on a computer screen. On each slide, a black vertical line separated the 2 scenes. The scenes depicted on each side differed in a single crucial way. As seen in Figure 4.4, for example, both sides included Elmo and Big Bird. However, on the left side, Big Bird hugs Elmo; on the right side, Elmo hugs Big Bird. Participants were asked to indicate their choice of matching picture by marking either A (for the left side) or B (for the right) on a sheet of paper, or by pointing to the picture on the computer screen.

If participants neither apply knowledge of scrambling nor attend to case markers, Big Bird should be regarded as the subject in both (18a) and (18b), giving the canonical SOV word order. In this case, the children would select the picture (i.e., A in Figure 4.4) that depicts Big Bird hugging Elmo for both (18a) and (18b). However, if participants have knowledge of scrambling and case and apply that knowledge, they should choose the picture in which Elmo hugs Big Bird (i.e., B in Figure 4.4), for (18b), the scrambled OSV sentence.
The stimuli were created using a Text-To-Speech (TTS) program with the most updated version of the TTS voice source (Yumi’s voice). The use of the TTS program had two motivations. First, the TTS source comes from a professional speaker, which helps ensure that the pronunciation is clear and of high quality. Second, the use of the TTS program avoids inconsistent pronunciations and unintentional prosodic cues. The naturalness of the pronunciation of the TTS source was confirmed by acceptability judgments by native Korean-speaking adults (n = 94). (See Appendix A for more details.)

All participants completed the 3 tasks with a 1- or 2-week interval in between each task. The order of the tasks was randomized.

### 4.5.1.3 Results.

#### 4.5.1.3.1 Comprehension data.

The older cL1ers (n = 23), who were the control group, showed high accuracy across all 3 tasks, both in canonical SOV sentences [baseline (M = 93.8%; SD = 0.10); prosody (M = 95.7%; SD = 0.08); context (M = 96.9%; SD = 0.06)] and in scrambled OSV sentences [baseline (M = 82.6%; SD = 0.24); prosody (M = 87.6%; SD = 0.18); context (M = 88.2%; SD = 0.19)]. The differences in the mean scores between the two word orders (SOV vs. OSV) and between the 3 tasks (baseline vs. prosody vs. context) were analyzed in repeated-measures analyses of variance (ANOVA). Alpha was set at .05 for all statistical tests. For the older cL1ers, there was a main effect of word order [F1(1, 22) = 5.961, p = .023; F2(1, 6) = 18.134, p < .01], a marginal main effect of task type [F1(2, 44) = 3.003, p = .06; F2(2, 12) = 5.099, p = .025], and a lack of interaction between word order and task type [(F1(2, 44) = .346, p = .710; F2(2, 12) = .074, p = .929]. The main effect of word order shows that scrambled OSV was more difficult than canonical SOV for the control group. The marginal main effect of task type suggests that the facilitating conditions generally improved the control group’s performance on the test items.

To examine the performance of the 2 target learner groups, KHC and cL1ers, a 2 (word order) x 3 (task type) x 3 (proficiency group) x 2 (learner group) ANOVA was conducted (within group factors = word order and task type; between group factors = proficiency group and learner group). This analysis yielded main effects of word order
$F_1(1, 46) = 173.715, p < .001; F_2(1, 36) = 1301.315, p < .001$, task type $[F_1(2, 92) = 7.170, p < .01; F_2(2, 72) = 13.079, p < .001]$, and proficiency group $[F_1(2, 46) = 18.818, p < .001; F_2(2, 36) = 425.729, p < .001]$ as well as the following interaction effects: word order x proficiency group $[F_1(2, 46) = 11.859, p < .001; F_2(2, 36) = 66.694, p < .001]$, word order x proficiency group x learner group $[F_1(2, 46) = 6.694, p < .01; F_2(2, 36) = 121.434, p < .001]$, and word order x task type $[F_1(1, 46) = 4.872, p = .032; F_2(2, 72) = 14.553, p < .001]$. To investigate further specific patterns for the 2 groups, separate repeated-measures ANOVAs were conducted for each group.

The KHC ($n = 31$) showed high accuracy across all 3 tasks for the [not shown] SOV sentences [baseline ($M = 98.6\%; SD = 0.04$); prosody ($M = 97.7\%; SD = 0.05$); context ($M = 97.7\%; SD = 0.05$)], but low accuracy on the OSV sentences [baseline ($M = 28.6\%; SD = 0.38$); prosody ($M = 48.4\%; SD = 0.39$); context 42.9\% ($SD = 0.41$)]. Figure 4.5 show the accuracy rates of the 3 KHC proficiency groups on the OSV items in each task.

![Figure 4.5. Accuracy rates on scrambled OSV sentences by the 3 KHC proficiency groups. Error bars represent standard errors.](image)

Notably, the high proficiency KHC (henceforth, KHC-High; $n = 10$) showed accuracy rates on OSV higher than 70% across the 3 tasks [(baseline ($M = 72.9\%; SD = 0.33$); prosody ($M = 87.1\%; SD = 0.17$); context ($M = 84.3\%; SD = 0.27$)] and ceiling performance on the SOV [baseline ($M = 98.6\%; SD = 0.05$); prosody ($M = 95.7\%; SD = 0.07$); context ($M = 95.7\%; SD = 0.07$)] [not shown], while the low proficiency
KHC (henceforth, KHC-Low; n = 10) showed accuracy rates of less than 20% on OSV across the 3 tasks [baseline (M = 0.0%; SD = 0); prosody (M = 11.4%; SD = 0.19); context (M = 15.7%; SD = 0.30)] but ceiling performance on SOV [baseline (M = 100.0%; SD = 0); prosody (M = 98.6%; SD = 0.05); context (M = 98.6%; SD = 0.05)] [not shown]. For the mid-proficiency KHC (henceforth, KHC-Mid; n = 11), the accuracy rates on OSV are less than 50% across the 3 tasks [baseline (M = 14.3%; SD = 0.21); prosody (M = 46.8%; SD = 0.32); context (M = 29.9%; SD = 0.30)], but they showed a dramatic improvement, in particular, between the baseline (14.3%) and prosody tasks (46.8%). Again, this group also showed ceiling performance on SOV [baseline (M = 97.4%; SD = 0.06); prosody (M = 98.7%; SD = 0.04); context (M = 98.7%; SD = 0.04)] [not shown].

For inferential statistics, a 2 (word order) x 3 (task type) x 3 (proficiency group) ANOVA was conducted (within group factors = word order and task type; between group factor = proficiency group) with the KHC data, which found main effects of word order [$F_1(1, 28) = 276.409, p < .001; F_2(1, 18) = 1433.619, p < .001$], task type [$F_1(2, 56) = 8.385, p < .01; F_2(2, 36) = 15.263, p < .001$], and proficiency group [$F_1(2, 28) = 27.274, p < .001; F_2(2, 12) = 252.109, p < .001$], as well as interaction effects between word order and task type [$F_1(2, 56) = 8.493, p < .01; F_2(2, 36) = 17.967, p < .001$] and between word order and proficiency group [$F_1(1, 28) = 39.694, p < .001; F_2(2, 18) = 207.181, p < .001$]. The main effect of word order demonstrates that scrambled OSV sentences were more difficult than canonical SOV sentences for the KHC; the main effect of task type suggests that task manipulations significantly improved their comprehension of the test items. The interaction effect between word order and task type comes from the fact that the task manipulations affected OSV items more than SOV items. The interaction between word order and proficiency group comes from the fact that lower level proficiency groups (i.e., KHC-Low and KHC-Mid) had more difficulty than the KHC-High group in understanding OSV items. In addition, planned comparisons on the OSV items showed that the KHC performed significantly better on the prosodic task than the baseline task [$t_1(30) = -4.690, p < .001; t_2(6) = -4.374, p < .01$] and better on the context task than the baseline task [$t_1(30) = -2.908, p < .01; t_2(6) = -5.452, p < .01$]. These analyses demonstrate that both prosody and context manipulations significantly improved their
comprehension of the scrambled OSV sentences. The improvement occurred no matter the order of tasks. To test whether there is an effect of task order, a 2 (word order) x 3 (task type) x 3 (proficiency group) x 2 (learner group) x 3 (task order: baseline first vs. prosody first vs. context first) ANOVA was conducted. This analysis yielded no main effect of task order ($F = .893, p = .418$). In addition, the factor of task order did not interact with any other factors.

Figure 4.6 show the OSV accuracy rates of the 3 young cL1ers proficiency groups in each task.

![Figure 4.6](image)

Figure 4.6. Accuracy rates on scrambled OSV sentences by the 3 young cL1ers proficiency groups. Error bars represent standard errors.

The young cL1ers ($n = 21$) showed high accuracy across all 3 tests for the [not shown] SOV sentences [baseline ($M = 90.5\%; SD = 0.12$); prosody ($M = 89.8\%; SD = 0.14$); context ($M = 93.9\%; SD = 0.10$)], but low accuracy on the OSV sentences [baseline ($M = 44.9\%; SD = 0.31$); prosody ($M = 54.4\%; SD = 0.30$); context ($M = 55.8\%$);

---

21 Kamil Ud Deen (personal communication) raises the possibility that task manipulations might improve the participants’ use of strategies to arrive at a correct answer. For example, children may know the NOM-ACC order as canonical, but have no clue what ACC-NOM order means. With no prosodic manipulation, they just ignore ACC and NOM. With prosodic manipulation, they must attend to ACC-NOM order, and they reason as follows: I know NOM-ACC order, and this is not NOM-ACC order, so the correct answer must be the opposite of the NOM-ACC choice. Deen proposes that this situation could be avoided by having a third picture, which is a distractor. This is an insightful suggestion. However, the 2 pictures that I created are a minimal pair, and it would be much harder to design a set of 3 pictures that are minimally different from each other. However, I believe it would be worthwhile to try to do this in a future study.
For the young cL1ers, the differences between the proficiency groups were not as large as the differences between the KHC proficiency groups. The high proficiency cL1ers (henceforth, cL1ers-High; n = 9) showed accuracy rates on OSV of less than 70% across the 3 tasks [baseline (M = 46.0%; SD = 0.34); prosody (M = 68.3%; SD = 0.22); context (M = 66.7%; SD = 0.34) and ceiling performance on SOV [baseline (M = 96.8%; SD = 0.10); prosody (M = 96.8%; SD = 0.06); context (M = 96.8%; SD = 0.06)] [not shown], while the low-level cL1ers (henceforth, cL1ers-Low; n = 3) showed OSV accuracy rates over 25% but under 40% across the 3 tasks [baseline (M = 38.1%; SD = 0.22); prosody (M = 38.1%; SD = 0.22); context (M = 28.6%; SD = 0.14)] but over 85% on SOV [baseline (M = 85.7%; SD = 0.14); prosody (M = 90.5%; SD = 0.08); context (M = 85.7%; SD = 0.14) [not shown]. For the medium-proficiency cL1ers (henceforth, cL1ers-Mid; n = 9), the accuracy rates on OSV are near chance across the 3 tasks [baseline (M = 46.0%; SD = 0.32); prosody (M = 46.0%; SD = 0.36); context (M = 54.0%; SD = 0.23)] but over 80% on SOV [baseline (M = 85.7%; SD = 0.14); prosody (M = 82.5%; SD = 0.19); context (M = 93.7%; SD = 0.10)] [not shown].

For inferential statistics, a 2 (word order) x 3 (task type) x 3 (proficiency group) ANOVA was conducted (within group factors = word order and task type; between group factor = proficiency group) with the cL1ers data. A main effect of word order \[F_1(1, 20) = 29.524, p < .001; F_2(1, 18) = 226.293, p < .001\], a marginal main effect of task type only in the by-item analysis \[F_1(2, 36) = 1.605, p = .215; F_2(2, 36) = 2.801, p = .074\], and an interaction effect between task type and proficiency group \[F_1(4, 36) = 2.844, p = .038; F_2(4, 36) = 2.514, p = .059\] were found. The main effect of word order shows that scrambled OSV sentences were more difficult than canonical SOV sentences for the young Korean-speaking children to comprehend. However, the effect of task type was not robust, which suggests that the manipulations improved the children’s comprehension of the test items, but not robustly. The interaction effect between task type and proficiency group suggests that not all proficiency groups benefitted from the task manipulations. Planned comparisons on the OSV items showed that the cL1ers performed significantly better on the prosodic task than the baseline task, in the by-participant analysis only \[t_1(20) = -2.390, p = .027; t_2(6) = -1.050, p = .334\] and on the context task than the baseline task \[t_1(20) = -2.212, p = .039; t_2(6) = 10.843, p < .001\]. These analyses show
that the context manipulation significantly improved this group’s comprehension of scrambled OSV sentences, while the effect of the prosody manipulation was not robust (because the effect appeared only in the by-participant analysis.

The differences in the accuracy rates between prosody and baseline tasks and between context and baseline tasks are plotted in Figures 4.7, 4.8, 4.9 and 4.10 for each individual. Positive scores indicate gains in accuracy.

**Figure 4.7.** Prosody–baseline difference in the KHC.

**Figure 4.8.** Context–baseline difference in the KHC.
Figures 4.7 and 4.8 show the tendency of the KHC’s performance to improve in the facilitating tasks. The improvement is shown across the 3 proficiency groups. Out of the 10 KHC-Low participants, 5 showed improvement in the prosody task and 4 showed improvement in the context task. Out of the 11 KHC-Mid participants, 10 showed improvement in the prosody task and 7 showed improvement in the context task. Out of the 10 KHC-High participants, 5 showed improvement in the prosody task and 4 showed improvement in the context task.

Figure 4.9. Prosody–baseline difference in the cL1ers.

Figure 4.10. Context–baseline difference in the cL1ers.

Figures 4.9 and 4.10 show the cL1ers’ improvement in the facilitating tasks. The improvement was seen in the cL1ers-Mid and cL1ers-High participants. Out of the 9
cL1ers-Mid participants, 2 showed improvement in the prosody task and 5 showed improvement in the context task. Out of the 9 cL1ers-High participants, 7 showed improvement in the prosody task and 5 showed improvement in the context task. These improvements suggest that performance factors affect language comprehension.

4.5.1.3.2 Assessing competence from comprehension data.

Let us examine whether each learner group knows the relevant properties of the TL grammar. As discussed in Chapter 2, I rely on the comparison-against-chance diagnostic to claim presence of knowledge. For this, I combined the accuracy score of each participant on OSV on the 2 manipulation tasks \( (k = 14) \) because the 2 manipulation tasks were each designed to reveal the knowledge of the participants. To test for the presence of knowledge, for each group, I ran a one-way ANOVA between the OSV accuracy of the group and chance (see Appendix D for more details). The results are shown in Table 4.2.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy on OSV ((k = 14))</th>
<th>Statistical information</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>KHC-Low</td>
<td>13.6% ((1.9/14)) ((SD = 0.24))</td>
<td>(F(1, 18) = 23.200, p &lt; .001)</td>
<td>Significantly below chance: Strong dependence on word order for interpretation</td>
</tr>
<tr>
<td>KHC-Mid</td>
<td>38.3% ((5.4/14)) ((SD = 0.25))</td>
<td>(F(1, 20) = 2.328, p = .143)</td>
<td></td>
</tr>
<tr>
<td>KHC-High</td>
<td>85.7% ((12.0/14)) ((SD = 0.18))</td>
<td>(F(1, 18) = 40.179, p &lt; .001)</td>
<td>Presence of TL knowledge</td>
</tr>
<tr>
<td>cL1ers-Low</td>
<td>33.3% ((4.7/14)) ((SD = 0.16))</td>
<td>(F(1, 4) = 3.063, p = .155)</td>
<td></td>
</tr>
<tr>
<td>cL1ers-Mid</td>
<td>50.0% ((7.0/14)) ((SD = 0.28))</td>
<td>(F(1, 16) = 0.000, p = 1.000)</td>
<td></td>
</tr>
<tr>
<td>cL1ers-High</td>
<td>67.5% ((9.4/14)) ((SD = 0.25))</td>
<td>(F(1, 16) = 4.293, p = .055)</td>
<td>Presence of TL knowledge ((but, p &lt; .06))</td>
</tr>
</tbody>
</table>

The results displayed in Table 4.2 indicate that both the KHC-High and cL1ers-High groups have TL knowledge, because there is a significant difference between these groups’ scores on OSV items and chance. The KHC-Mid, cL1ers-Low, and cL1ers-Mid groups did not show significant differences from chance. As for the KHC-Low group, there was a significant difference, but this means that the actual score was significantly
below chance. It shows that the KHC-Low group frequently misinterpreted OSV sentences as SOV.

As discussed in Chapter 2, we must be cautious in concluding that there is a deficit in knowledge when a participant group (in this study, the KHC-Low, KHC-Mid, cL1ers-Low, and cL1ers-Mid groups) does not pass the comparison-against-chance diagnostic. Recall that I introduced another diagnostic, the comparison-against-baseline diagnostic. An absence of knowledge will be considered when the participant groups pass neither the comparison-against-chance diagnostic nor the comparison-against-baseline diagnostic.

To apply the comparison-against-baseline diagnostic, the following question must be addressed: What is the baseline condition for comparison with scrambled OSV? The answer I propose is this: The baseline condition is (in)accuracy on SOV items. This is because if participants do not have knowledge of OSV, they would treat OSV sentences as if they were SOV sentences. For example, let us revisit (1a–c), where this time I marked strikethroughs on the case markers as in (1a) and (1c).

(1) a. Wensungi-ka kay-lul cha.
   monkey-NOM dog-ACC kick
   ‘A monkey is kicking a dog.’

b. Wensungi kay cha.
   monkey dog kick
   ‘A monkey is kicking a dog.’

c. Wensungi-lul kay-ka cha.
   monkey-ACC dog-NOM kick
   ‘A monkey, a dog is kicking.’

With no knowledge of case, the three sentences in (1a–c) should all have the same interpretation. If a participant does not know case and shows 100% accuracy on the SOV items (as in (1a)), s/he should show 0% accuracy on the OSV items (as in (1c)). That is, the accuracy rate on the OSV items would not be different from the inaccuracy rate on
the SOV items. Based on this logic, I combined the accuracy score of each participant on the OSV items \( (k = 14) \) on the 2 manipulation tasks. Then I calculated the SOV inaccuracy score on the 2 manipulation tasks (i.e., by subtracting the SOV accuracy score from 14). Subsequently, for each group, I carried out a repeated-measures ANOVA between the OSV accuracy score and the SOV inaccuracy score, to examine whether there was a significant difference between the two numbers (see Appendix E for more details). The results are presented in Table 4.3.

Table 4.3

<table>
<thead>
<tr>
<th></th>
<th>Accuracy on OSV</th>
<th>Inaccuracy on SOV</th>
<th>Statistical information</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>KHC-Low</td>
<td>13.6% (1.9/14)</td>
<td>1.4% (0.2/14)</td>
<td>( F(1, 9) = 2.499, p = .148 )</td>
<td></td>
</tr>
<tr>
<td>KHC-Mid</td>
<td>38.3% (5.4/14)</td>
<td>1.3% (0.2/14)</td>
<td>( F(1, 10) = 21.460, p &lt; .01 )</td>
<td>Possibility of TL knowledge</td>
</tr>
<tr>
<td>KHC-High</td>
<td>85.7% (12.0/14)</td>
<td>4.3% (0.6/14)</td>
<td>( F(1, 9) = 161.552, p &lt; .001 )</td>
<td>Possibility of TL knowledge</td>
</tr>
<tr>
<td>cL1ers-Low</td>
<td>33.3% (4.7/14)</td>
<td>11.9% (1.7/14)</td>
<td>( F(1, 2) = 3.000, p = .225 )</td>
<td></td>
</tr>
<tr>
<td>cL1ers-Mid</td>
<td>50.0% (7.0/14)</td>
<td>11.9% (1.7/14)</td>
<td>( F(1, 8) = 17.067, p &lt; .01 )</td>
<td>Possibility of TL knowledge</td>
</tr>
<tr>
<td>cL1ers-High</td>
<td>67.5% (9.4/14)</td>
<td>3.1% (0.4/14)</td>
<td>( F(1, 8) = 57.176, p &lt; .001 )</td>
<td>Possibility of TL knowledge</td>
</tr>
</tbody>
</table>

The results suggest that in this less conservative diagnostic, not only the KHC-High and cL1ers-High groups but also the KHC-Mid and cL1ers-Mid groups can be seen to possibly have knowledge of scrambling and case. That is, we cannot simply claim an absence of knowledge in the KHC-Mid and cL1ers-Mid groups based on the results from the comparison-against-chance diagnostic. The groups for which we conclude there is an absence of knowledge are only KHC-Low and cL1ers-Low.

22 This logic can be viewed in the following way (William O’Grady, personal communication): Assuming that the learners have the grammatical knowledge needed to interpret the SOV pattern, we can take their error rate on that pattern to be a measure of the rate at which, for one reason or another, they do not use that knowledge. Now, if they are really interpreting the OSV pattern as the SOV pattern, they should depart from the (incorrect) SOV-interpretation for the OSV pattern at no greater a rate than is observed with the genuine SOV pattern. Any significant difference between the two would therefore indicate that something else is going on in the OSV pattern, and that something else might be the distinguishing function of case.
Now, let us look at the individual data. First, let us apply the comparison-against-chance diagnostic. There are 14 OSV items in the 2 manipulation tasks combined. The probability of having 11 or more correct out of the 14 OSV items is .029 ($p < .05$). Only when participants have knowledge of scrambling and case can they reasonably be expected to obtain such a score. Thus, at least 11 correct out of the 14 OSV items in the 2 manipulation tasks suggests the presence of knowledge in the individual participant. Next, let us apply the comparison-against-baseline diagnostic.

Under this second diagnostic, as discussed above, the baseline is the inaccuracy score on the SOV items. There are 14 OSV items and 14 SOV items in the 2 manipulation tasks combined. The average inaccuracy score of the 14 SOV items was 4.3% (0.6/14) across the KHC ($n = 31$) and cL1ers ($n = 21$). The probability of having 3 or more than 3 correct out of the 14 OSV items, distinguishing them from the SOV items, is .020 ($p < .05$).

See Figures 4.11 and 4.12.

---

23 This is based on a binomial distribution. The $p$-value is calculated as follows:
\[
\frac{14!}{11!*3!}*0.5^{11}*(1-0.5)^{3} + \frac{14!}{12!*2!}*0.5^{12}*(1-0.5)^{2} + \\
\frac{14!}{13!*1!}*0.5^{13}*(1-0.5)^{1} + \frac{14!}{14!*0!}*0.5^{14} = 0.029. 
\]
This is the same as the following situation: When the probability of a tossed coin coming up heads is .5, the probability of achieving 11 or more heads in 14 tosses is .029.

24 This is based on a binomial distribution. The $p$-value is calculated as follows:
\[
\frac{14!}{3!*11!}*0.043^{3}*(1-0.043)^{11} + \frac{14!}{4!*10!}*0.043^{4}*(1-0.043)^{10} + \frac{14!}{5!*9!}*0.043^{5}*(1-0.043)^{9} + \\
\frac{14!}{6!*8!}*0.043^{6}*(1-0.043)^{8} + \frac{14!}{7!*7!}*0.043^{7}*(1-0.043)^{7} + \\
\frac{14!}{8!*6!}*0.043^{8}*(1-0.043)^{6} + \frac{14!}{9!*5!}*0.043^{9}*(1-0.043)^{5} + \\
\frac{14!}{10!*4!}*0.043^{10}*(1-0.043)^{4} + \frac{14!}{11!*3!}*0.043^{11}*(1-0.043)^{3} + \\
\frac{14!}{12!*2!}*0.043^{12}*(1-0.043)^{2} + \frac{14!}{13!*1!}*0.043^{13}*(1-0.043)^{1} + \\
\frac{14!}{14!*0!}*0.043^{14} = 0.020. 
\]
This is the same as the following situation: When a biased coin comes up heads with probability 0.043 when tossed, the probability of achieving 3 or more heads after 14 tosses is .020.
Figures 4.11 and 4.12 show that 7 (out of 10) KHC-High participants, 25 1 (out of 11) KHC-Mid participant, 26 1 (out of 10) KHC-Low participant, 27 3 (out of 9) cL1ers-High participants, 28 2 (out of 9) cL1ers-Mid participants, 29 and 0 (out of 3) cL1ers-Low participants showed 11 or more correct on the OSV items of the 2 manipulation tasks. This suggests that under this comparison-against-chance diagnostic (i.e., 11 or more
correct responses)

25 The 7 KHC are P23, P26, P27, P28, P29, P30, and P31.
26 The 1 KHC is P14.
27 The 1 KHC is P9.
28 The 3 cL1ers are P18, P19 and P21.
29 The 2 cL1ers are P8 and P11.
correct), 9 (out of 31) KHC and 5 (out of 21) cL1ers have knowledge of scrambling and case.

Figures 4.11 and 4.12 also show that all 10 KHC-High participants, 8 (out of 11) KHC-Mid participants,\(^{30}\) 2 (out of 10) KHC-Low participants,\(^{31}\) all 9 cL1ers-High participants, 7 (out of 9) cL1ers-Mid participants,\(^{32}\) and 2 (out of 3) cL1ers-Low participants\(^{33}\) showed 3 or more correct on the OSV items of the 2 manipulation tasks. So, under the comparison-against-baseline diagnostic (i.e., 3 or more correct), 20 (out of 31) KHC and 18 (out of 21) cL1ers possibly have knowledge of scrambling and case. Finally, the participants who have a deficit in knowledge are 11 KHC\(^{34}\) and 3 cL1ers,\(^{35}\) who passed neither the comparison-against-chance diagnostic nor the comparison-against-baseline diagnostic. Figure 4.13 summarizes grouping of the participants based on their performance.

<table>
<thead>
<tr>
<th>Scope of performance showing knowledge deficit</th>
<th>Debatable area</th>
<th>Scope of good performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(knowledge: −)</td>
<td>(knowledge: ?)</td>
<td>(knowledge: +)</td>
</tr>
<tr>
<td>significantly-above-baseline</td>
<td></td>
<td>significantly-above-chance</td>
</tr>
</tbody>
</table>

**Group:**
- KHC-Low; cL1ers-Low
- KHC-Mid; cL1ers-Mid
- KHC-High; cL1ers-High

**Individual:**
- 11 KHC; 3 cL1ers
- 11 KHC; 13 cL1ers
- 9 KHC; 5 cL1ers

*Figure 4.13. Categorizing the participants based on the number of correct responses.*

---

\(^{30}\) The 8 KHC are all the KHC-Mid participants except P12, P13, and P16.

\(^{31}\) The 2 KHC are P5 and P9.

\(^{32}\) The 7 cL1ers are all the cL1ers-Mid participants except P4 and P9.

\(^{33}\) The 2 cL1ers are P1 and P2.

\(^{34}\) The 11 KHC are as follows: P1, P2, P3, P4, P6, P7, P8, P10, P12, P13, and P16.

\(^{35}\) The 3 cL1ers are P3, P4, and P9.
4.5.2 Case-marker-elicitation production task.

4.5.2.1. Participants.

All 31 KHC and 21 young cL1ers who participated in the picture-selection task took part in the case-marker-elicitation task. As a control group, 15 older native Korean-speaking children (old cL1ers; ages 10–11) also participated in this production task. None of the control group participant had lived in an English-speaking country for more than 3 months; these children were not the same as those who completed the picture-selection task.

4.5.2.2. Materials.

Three conditions of the experiment were administered in 3 sequential sessions: (a) Session 1 (base condition; \( k = 3 \)), in which participants described 3 given pictures; (b) Session 2 (subject-marker-elicitation condition; \( k = 3 \)), in which I elicited contrastive focus on the subject of the participants’ utterance by incorrectly naming the subject; and (c) Session 3 (object-marker-elicitation condition; \( k = 3 \)), in which I elicited contrastive focus on the direct object of the participants’ utterance by incorrectly naming the direct object. All the items can be found in Appendix F. The order of sessions was Session 1, Session 2, and Session 3. There was a training session between Session 1 and Session 2 (see Figure 4.3, p. 47).

4.5.2.3. Results.

4.5.2.3.1 Production data.

Most of the KHC attempted to produce case markers in their oral production, although some errors in use occurred. See Tables 4.4 and 4.5.
Table 4.4  
**Subject-markings in Sessions 1 and 2, case-marker elicitation task: KHC**

<table>
<thead>
<tr>
<th></th>
<th>Older cL1ers (n = 15)</th>
<th>KHC-Low (n = 10)</th>
<th>KHC-Mid (n = 11)</th>
<th>KHC-High (n = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Session 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(non-focused subject)</td>
<td>Out of 45 tokens,</td>
<td>Out of 30 tokens,</td>
<td>Out of 33 tokens,</td>
<td>Out of 30 tokens,</td>
</tr>
<tr>
<td></td>
<td>• NOM 43 (95.6%)</td>
<td>• NOM 12 (40.0%)</td>
<td>• NOM 23 (69.7%)</td>
<td>• NOM 28 (93.3%)</td>
</tr>
<tr>
<td></td>
<td>• TOP 2 (4.4%)</td>
<td>• TOP 11 (36.7%)</td>
<td>• TOP 6 (18.2%)</td>
<td>• TOP 2 (6.7%)</td>
</tr>
<tr>
<td></td>
<td>• Ø 7 (23.3%)</td>
<td>• Ø 4 (12.1%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Session 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(focused subject)</td>
<td>Out of 45 tokens,</td>
<td>Out of 30 tokens,</td>
<td>Out of 33 tokens,</td>
<td>Out of 30 tokens,</td>
</tr>
<tr>
<td></td>
<td>• NOM 36 (80.0%)</td>
<td>• NOM 14 (46.7%)</td>
<td>• NOM 23 (69.7%)</td>
<td>• NOM 25 (83.3%)</td>
</tr>
<tr>
<td></td>
<td>• TOP 9 (20.0%)</td>
<td>• TOP 7 (23.3%)</td>
<td>• TOP 7 (21.2%)</td>
<td>• TOP 5 (16.7%)</td>
</tr>
<tr>
<td></td>
<td>• Ø 9 (30.0%)</td>
<td>• Ø 2 (6.1%)</td>
<td>• Ø 2 (6.1%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The older cL1ers (n = 15) and the children in the KHC-High group (n = 10) used case correctly and productively in the case-marker elicitation task. In the subject-marking conditions, they correctly produced either nominative markers or topic markers 100% of the time; in the object-marking condition, they correctly produced accusative markers, more than 90% of the time. The KHC-Low group (n = 10) supplied case markers at high rates in the subject-marking conditions (76.7% in Session 1; 70.7% in Session 2) and at lower rates in the object-marking conditions (43.3% in Session 1; 43.3% in Session 3); their subject-markings were all correct, while their object-markings included some errors (e.g., dative marker -hanthey for accusative -lul). The KHC-Mid Group (n = 11) supplied case markers at high rates in both subject-marking and object-marking conditions. Their

Table 4.5  
**Object-markings in Sessions 1 and 3, case-marker elicitation task: KHC**

<table>
<thead>
<tr>
<th></th>
<th>Older cL1ers (n = 15)</th>
<th>KHC-Low (n = 10)</th>
<th>KHC-Mid (n = 11)</th>
<th>KHC-High (n = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Session 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(non-focused object)</td>
<td>Out of 45 tokens,</td>
<td>Out of 30 tokens,</td>
<td>Out of 33 tokens,</td>
<td>Out of 30 tokens,</td>
</tr>
<tr>
<td></td>
<td>• ACC 45 (100.0%)</td>
<td>• ACC 9 (30.0%)</td>
<td>• ACC 24 (72.7%)</td>
<td>• ACC 29 (96.7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Ø 1 (3.1%)</td>
<td>• Ø 1 (3.1%)</td>
<td>• *NOM 1 (3.3%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• *DAT 4 (13.3%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Session 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(focused object)</td>
<td>Out of 45 tokens,</td>
<td>Out of 30 tokens,</td>
<td>Out of 33 tokens,</td>
<td>Out of 30 tokens,</td>
</tr>
<tr>
<td></td>
<td>• ACC 45 (100.0%)</td>
<td>• ACC 9 (30.0%)</td>
<td>• ACC 23 (69.7%)</td>
<td>• ACC 30 (100.0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Ø 17 (56.7%)</td>
<td>• Ø 1 (3.1%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• *DAT 4 (13.3%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The older cL1ers (n = 15) and the children in the KHC-High group (n = 10) used case correctly and productively in the case-marker elicitation task. In the subject-marking conditions, they correctly produced either nominative markers or topic markers 100% of the time; in the object-marking condition, they correctly produced accusative markers, more than 90% of the time. The KHC-Low group (n = 10) supplied case markers at high rates in the subject-marking conditions (76.7% in Session 1; 70.7% in Session 2) and at lower rates in the object-marking conditions (43.3% in Session 1; 43.3% in Session 3); their subject-markings were all correct, while their object-markings included some errors (e.g., dative marker -hanthey for accusative -lul). The KHC-Mid Group (n = 11) supplied case markers at high rates in both subject-marking and object-marking conditions. Their
rates of subject-marking were 87.9% and 90.9%, for Session 1 and Session 2, respectively; their rates of object-marking were 96.9% and 97.0% for Session 1 and Session 3, respectively. However, again, these object-markings showed some incorrect case-markings (e.g., dative -hanthey for accusative -lul), while the subject-markings were all correct.

Now, let us look at the data from young cL1ers. See Tables 4.6 and 4.7.

Table 4.6
Subject-markings in Sessions 1 and 2, case-marker-elicitation task: cL1ers

<table>
<thead>
<tr>
<th>Session 1 (non-focused subject)</th>
<th>Older cL1ers (n = 15)</th>
<th>cL1ers-Low (n = 3)</th>
<th>cL1ers-Mid (n = 9)</th>
<th>cL1ers-High (n = 9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Out of 45 tokens,</td>
<td>Out of 9 tokens,</td>
<td>Out of 27 tokens,</td>
<td>Out of 27 tokens,</td>
<td></td>
</tr>
<tr>
<td>• NOM 43 (95.6%)</td>
<td>• NOM 3 (33.3%)</td>
<td>• NOM 22 (81.5%)</td>
<td>• NOM 24 (88.9%)</td>
<td></td>
</tr>
<tr>
<td>• TOP 2 (4.4%)</td>
<td>• Ø 6 (66.7%)</td>
<td>• Ø 5 (18.5%)</td>
<td>• TOP 3 (11.1%)</td>
<td></td>
</tr>
<tr>
<td>Session 2 (focused subject)</td>
<td>Out of 45 tokens,</td>
<td>Out of 27 tokens,</td>
<td>Out of 27 tokens,</td>
<td></td>
</tr>
<tr>
<td>• NOM 36 (80.0%)</td>
<td>• NOM 5 (55.5%)</td>
<td>• NOM 19 (70.4%)</td>
<td>• NOM 24 (88.9%)</td>
<td></td>
</tr>
<tr>
<td>• TOP 9 (20.0%)</td>
<td>• Ø 4 (44.5%)</td>
<td>• TOP 2 (7.4%)</td>
<td>• TOP 3 (11.1%)</td>
<td></td>
</tr>
</tbody>
</table>

Notes. NOM (nominative case; -i/-ka), TOP (topic marker; -un/-nun), ACC (accusative case; -ul/-lul), DAT (dative case; -hanthey, -eykey), Ø (no case marking)

Table 4.7
Object-markings in Sessions 1 and 3, case-marker-elicitation task: cL1ers

<table>
<thead>
<tr>
<th>Session 1 (non-focused object)</th>
<th>Older cL1ers (n = 15)</th>
<th>cL1ers-Low (n = 3)</th>
<th>cL1ers-Mid (n = 9)</th>
<th>cL1ers-High (n = 9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Out of 45 tokens,</td>
<td>Out of 9 tokens,</td>
<td>Out of 27 tokens,</td>
<td>Out of 27 tokens,</td>
<td></td>
</tr>
<tr>
<td>• ACC 45 (100.0%)</td>
<td>• ACC 1 (11.1%)</td>
<td>• ACC 3 (11.1%)</td>
<td>• ACC 23 (85.2%)</td>
<td></td>
</tr>
<tr>
<td>• Ø 6 (66.7%)</td>
<td>• Ø 17 (63.0%)</td>
<td>• Ø 17 (63.0%)</td>
<td>• *DAT 4 (14.8%)</td>
<td></td>
</tr>
<tr>
<td>• *DAT 2 (22.2%)</td>
<td></td>
<td>• *DAT 7 (25.9%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session 3 (focused object)</td>
<td>Out of 45 tokens,</td>
<td>Out of 27 tokens,</td>
<td>Out of 27 tokens,</td>
<td></td>
</tr>
<tr>
<td>• ACC 45 (100.0%)</td>
<td>Out of 9 tokens,</td>
<td>Out of 27 tokens,</td>
<td>Out of 27 tokens,</td>
<td></td>
</tr>
<tr>
<td>• Ø 7 (77.7%)</td>
<td>• ACC 5 (18.5%)</td>
<td>• ACC 5 (18.5%)</td>
<td>• ACC 24 (88.9%)</td>
<td></td>
</tr>
<tr>
<td>• *DAT 2 (22.3%)</td>
<td>• Ø 14 (51.9%)</td>
<td>• Ø 14 (51.9%)</td>
<td>• Ø 2 (7.4%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• *DAT 7 (25.9%)</td>
<td>• *DAT 7 (25.9%)</td>
<td>• *DAT 1 (3.7%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• *NOM 1 (3.7%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes. NOM (nominative case; -i/-ka), TOP (topic marker; -un/-nun), ACC (accusative case; -ul/-lul), DAT (dative case; -hanthey, -eykey), Ø (no case marking)

The cL1ers-High group (n = 9) showed productive and largely correct use of case. In the subject-marking conditions, they correctly produced either nominative markers or topic markers 100% of the time; in the object-marking conditions, they produced accusative
markers more than 85% of the time, but also showed a little erroneous use of dative case (i.e., -hanthey) for accusative -lul. The cL1ers-Mid group \( (n = 9) \) showed high rates of supplying markers for the subject-marking conditions, but lower rates for the object-marking conditions. The rates of acceptable subject-marking were 81.5% and 77.8%, for Session 1 and Session 2, respectively; the rates of object-marking were 37.0% and 48.1% for Session 1 and Session 3, respectively. Again, these object-markings showed some incorrect case-makings (e.g., dative marker -hanthey for accusative -lul), while the subject-markings were all correct. The cL1ers-Low group \( (n = 3) \) showed many omissions of case markers: Their rates of supplying markers in the subject-marking conditions were 33.3% \((3/9)\) in Session 1 and 55.5% \((5/9)\) in Session 2, and their rates in the object-marking conditions were 33.3% \((3/9)\) in Session 1 and 22.3% \((2/9)\) in Session 3. Their subject-markings were all correct, while their object-markings included the incorrect use of dative -hanthey for accusative -lul.

The notable errors that the participants produced were object-markings, in particular, the dative marker -hanthey for accusative -lul. In fact, this error pattern was already observed in a previous study by Choi (1997, p. 99), who reported that young native Korean-speaking children made “a number of errors” involving the use of the animate dative marker -hanthey for the accusative marker -lul. She presented 10 instances of errors as examples, in which the verbs used with dative -hanthey included ha- ‘do’, man- ‘meet’, mek- ‘eat’, mwul- ‘bite’, pis- ‘comb’, ttala- ‘follow’, and ttayli- ‘hit’. Interestingly, she stated that incorrect marking did not show up as other locative markers such as -ey, -eyse, -eyta which are used for inanimate NPs. Moreover, errors in object marking have been reported in studies with heritage speakers. For example, Montrul and Bowles (2009) reported poor performance by heritage Spanish speakers in both production and acceptability-judgment tasks testing knowledge that the preposition a must occur with an animate object, but not with an inanimate object. Both Choi and Montrul and Bowles indicate that mapping correct case morphology from Case features is not simple for learners: The learners should search case morphology for Case features as well as search morphology to carry semantic (animate/inanimate) information.
Now, let us look at the individual data. Figures 4.14 and 4.15 show the distribution of subject-markings in Session 1 and Session 2 by individual KHC.

**Figure 4.14.** Subject-markings by individual KHC in Session 1: Non-focused subject.

**Figure 4.15.** Subject-markings by individual KHC in Session 2: Focused subject.

The figures show that of the KHC, P2 and P10 produced no case markers for subjects in either session; the others produced either nominative -ka or topic -num with subjects more than 50% of the time (i.e., at least 2 out of 3 times) in at least one of the sessions. The only case marking error was produced by P16 in Session 2; the child used accusative -lul for subject marking.
Figures 4.16 and 4.17 show the distribution of object-markings in Session 1 and Session 3 by individual KHC.

**Figure 4.16.** Object-markings by individual KHC in Session 1: Non-focused object.

**Figure 4.17.** Object-markings by individual KHC in Session 3: Focused object.

The figures indicate that many participants in the KHC-Low group produced objects without markers. In particular, P2, P9, and P10 had no markers at all in either session. In addition, 8 KHC (P3, P4, P6, P11, P12, P15, P21, and P22) showed at least one instance of incorrect object-marking.
Figures 4.18 and 4.19 show the distribution of subject-markings in Session 1 and Session 2 by individual cL1ers.

**Figure 4.18.** Subject-markings by individual cL1ers in Session 1: Non-focused subject.

**Figure 4.19.** Subject-markings by individual cL1ers in Session 2: Focused subject.

The figures show that of the cL1ers, P8 produced no case markers for subjects in either session; P1, P2, and P7 produced nominative marker -ka, but less than 50% of the time, in both sessions (i.e., 1 out of 3 times in both sessions). All other cL1ers produced either nominative -ka or topic -nun with subjects more than 50% of the time in both sessions. No case errors appeared.
Figures 4.20 and 4.21 display the distribution of object-markings in Session 1 and Session 3 by individual cL1ers.

**Figure 4.20.** Object-markings by individual cL1ers in Session 1: Non-focused object.

**Figure 4.21.** Object-markings by individual cL1ers in Session 3: Focused object.

The figures indicate that many participants in the cL1er-Low and cL1er-Mid groups produced objects with no markers. For example, P1, P2, P6, P7, and P8 produced no case...
markers at all across the 2 sessions. In addition, 7 cL1ers had at least one instance of incorrect case in object marking: P3, P4, P5, P10, P12, P13, and P14.\textsuperscript{36}

\textbf{4.5.2.3.2 Assessing competence from production data.}

Let us examine whether the learners know the relevant properties of TL grammar. First, we assess whether each learner group has TL knowledge. As discussed in Chapter 2, I claim \textit{presence of knowledge} when the group accuracy in production tasks is significantly above their inaccuracy rate; I claim \textit{absence of knowledge} when the group accuracy in production tasks is significantly below their inaccuracy rate; and I claim \textit{possibility of knowledge} when the group accuracy is neither significantly above nor below the group inaccuracy. For this, I combined the number of correct case markings for each participant from the subject-marking conditions ($k = 6$; 3 from Session 1 and 3 from Session 2) and the object-marking conditions ($k = 6$; 3 from Session 1 and 3 from Session 3). Next, for each learner group, I ran two separate repeated measures ANOVAs to compare (a) the number of correct vs. incorrect subject-markings and (b) the number of correct vs. incorrect object-markings. See Tables 4.8 and 4.9 for the results.

\textsuperscript{36} In fact, the results from the case-marker-elicitation task (both for the KHC and the cL1ers) show no difference in case-marker production between the non-focused session and the focused session. A further study is required to explain this null effect of focus.
Using the criteria described above, we can state that 4 groups (KHC-Mid, KHC-High, cL1ers-Mid, and cL1ers-High) know the subject-markings because each group had correct subject-markings significantly above incorrect subject-markings. As for object-marking, 2 out of 6 groups (KHC-High and cL1ers-High) know it, showing accurate use of accusative -lul at rates significantly above the inaccuracy rates. In addition, we can state that 2 groups (KHC-Low, cL1ers-Low) possibly know the
subject-markings because each group showed correct subject-markings neither significantly above nor below incorrect subject-markings. As for object marking, 2 out of 6 groups (KHC-Low and KHC-Mid) possibly know it, having accurately used accusative -lul neither significantly above nor below their inaccuracy rates. Finally, we can state that 2 out of 6 groups (cL1ers-Low and cL1ers-Mid) do not know the object-markings because their accuracy rates on the object-markings are significantly below their inaccuracy rates.

Next, let us assess each individual’s knowledge of case. As discussed in Chapter 2, presence of knowledge is operationalized as 5 or 6 correct (out of 6; 83.3%–100%); absence of knowledge is operationalized as 0 or 1 correct (out of 6; 0%–16.7%); and possibility of knowledge is operationalized as between 2 and 4 correct (out of 6; 33.3%–66.7%). This is separately calculated for subjects and for objects. See Figures 4.22, 4.23, 4.24, and 4.25.

![Figure 4.22. Number of correct subject-markings: KHC.](image)

---

As discussed, these numbers are determined not based on statistics, but based on intuition. Intuitively, out of 6 tokens, 5 or 6 correct is well above 1 or 0 incorrect (i.e., the number of accurate responses is well above the number of inaccurate responses) and 0 or 1 correct is well below 5 or 6 incorrect (i.e., the number of accurate responses is well below the number of inaccurate responses). However, 2–4 correct is not very different from the number of incorrect responses (i.e., 4–2 incorrect).
Figure 4.23. Number of correct object-markings: KHC.

Figure 4.24. Number of correct subject-markings: cL1ers.

Figure 4.25. Number of correct object-markings: cL1ers.
Based on the criteria just presented, out of 31 KHC and 21 cL1ers, 27 KHC\textsuperscript{38} and 17 cL1ers\textsuperscript{39} know subject-marking, and 19 KHC\textsuperscript{40} and 8 cL1ers\textsuperscript{41} know object-marking. These participants marked case correctly 5 or 6 times out of 6 opportunities. In addition, 2 KHC\textsuperscript{42} and 1 cL1er\textsuperscript{43} might know subject-marking, and 4 KHC\textsuperscript{44} and 2 cL1ers\textsuperscript{45} might know object-marking. These participants marked case correctly between 2 and 4 times out of 6 items. As for the lack of case-marking knowledge, I consider 2 KHC\textsuperscript{46} and 3 cL1ers\textsuperscript{47} as not knowing subject-marking, and 8 KHC\textsuperscript{48} and 11 cL1ers\textsuperscript{49} as not knowing object-marking. These participants marked case correctly only at most 1 time out of 6 items. Figures 4.26 and 4.27 summarize the results.

![Diagram showing the scope of performance, debatable area, and scope of good performance for KHC and cL1ers.](image)

**Group:**
- KHC-Low; cL1ers-Low
- KHC-Mid; KHC-High; cL1ers-Mid; cL1ers-High

**Individual:**
- 2 KHC; 3 cL1ers
- 2 KHC; 1 cL1ers
- 27 KHC; 17 cL1ers

*Figure 4.26. Categorizing the participants: Knowledge of subject-marking.*

\textsuperscript{38} The 27 KHC are all the KHC except P1, P2, P10, and P14.
\textsuperscript{39} The 17 cL1ers are all the cL1ers except P1, P2, P7 and P8.
\textsuperscript{40} The 19 KHC are all the KHC except P1, P2, P3, P4, P6, P7, P8, P9, P10, P11, P15, and P21.
\textsuperscript{41} The 8 cL1ers are as follows: P14, P15, P16, P17, P18, P19, P20, and P21.
\textsuperscript{42} The 2 KHC are as follows: P1 and P14.
\textsuperscript{43} The 1 cL1er is P7.
\textsuperscript{44} The 4 KHC are P1, P3, P4, and P7.
\textsuperscript{45} The 2 cL1ers are P9 and P10.
\textsuperscript{46} The 2 KHC are P2 and P10.
\textsuperscript{47} The 3 cL1ers are P1, P2, and P8.
\textsuperscript{48} The 8 KHC are as follows: P2, P6, P8, P9, P10, P11, P15, and P21.
\textsuperscript{49} The 11 cL1ers are as follows: P1, P2, P3, P4, P5, P6, P7, P8, P11, P12, and P13.
4.5.3 Source of poor performance of each individual in the baseline comprehension task.

This section is concerned with individual evaluation, focusing on what might be the sources of poor performance in the baseline comprehension task for each individual. For this, I examine the results from the picture-selection comprehension tasks and the case-marker-elicitation production task, comprehensively. If a participant passes the comparison-against-chance diagnostic with the combined OSV items from the 2 manipulation picture-selection tasks (i.e., 11 or more correct out of the 14 OSV items), it suggests that the participant knows both scrambling and case. I use the plus sign (+) to mark the presence of knowledge. If a participant passes the comparison-against-baseline diagnostic but did not pass the comparison-against-chance diagnostic (i.e., less than 11 but 3 or more than 3 correct out of the 14 OSV items), I use the question mark (?). This is because I cannot claim definite presence (or absence) of knowledge based on this diagnostic. When a participant passes neither the comparison-against-chance diagnostic nor the comparison-against-baseline diagnostic, I use the minus sign (–), which indicates an absence of knowledge, either in scrambling or in case.

If a participant has 5–6 correct out of 6 for subjects and for objects in the case-marker-elicitation production task, it suggests that the participant knows each case marking, respectively. I use the plus sign (+) to indicate the presence of knowledge. However, if a participant has 2–4 correct out of 6, I claim neither presence nor absence of knowledge, and use the question mark (?). When a participant has 0–1 correct out of 6, I
consider this an absence of knowledge, and use the minus sign (–). In addition, if a
participant shows an error in the use of case, I attribute it to showing a problem in
mapping from Case feature to case morphology.

The sources of poor performance in the baseline task of the picture-selection
comprehension test can be the following: If a participant is assigned either the minus
mark (–) on the 7 OSV items in the baseline picture-selection task (i.e., 0–1 correct out of
7 OSV items; numbers calculated based on the comparison-against-baseline diagnostic\textsuperscript{50})
or the question mark (?) on the 7 OSV items in the baseline task (i.e., 2–5 correct out of 7
OSV items; based on the comparison-against-chance diagnostic\textsuperscript{51}), but is assigned the
plus mark (+) (i.e., 11 or more correct out of the 14 OSV items) in the manipulation tasks,
this suggests that the source of poor performance in the baseline task is a performance
problem (i.e., perception failure or processing burden). Similarly, if a participant is
assigned the minus mark (–) on the 7 OSV items in the baseline task but is assigned the
question mark (?) on the 14 OSV items in the manipulation tasks, this suggests that the
source of poor performance in the baseline task could be a performance problem (i.e.,
perception failure or processing burden).\textsuperscript{52} If a participant shows an absence of
knowledge in case (as attested in the case-marker-elicitation task), the poor performance
in the baseline picture-selection task can be attributed to a deficit in knowledge of case. If
a participant shows an error in the use of case (as attested in the case-marker-elicitation
task), the poor performance in the baseline picture-selection task could be due to a
problem of mapping between case morphology and Case feature. See Tables 4.10 and
4.11 for the individual evaluations.

\textsuperscript{50} This is based on a binomial distribution: \{7!/(2!*5!)}^*0.043^2*(1-0.043)^5 + 
\{7!/(3!*4!)}^*0.043^3*(1-0.043)^4 + \{7!/(4!*3!)}^*0.043^4*(1-0.043)^3 + 
\{7!/(5!*2!)}^*0.043^5*(1-0.043)^2 + \{7!/(6!*1!)}^*0.043^6*(1-0.043)^1 + 
\{7!/(7!*0!)}^*0.043^7*(1-0.043)^0 = 0.034 (p < .05). This is the same as the following situation:
When a coin comes up heads with probability 0.043 when tossed, the probability of achieving 2
or more heads after 7 tosses is .034.

\textsuperscript{51} This is based on a binomial distribution. The p-value is calculated as follows:
\{7!/(6!*1!)}^*0.5^6*(1-0.5)^1 + \{7!/(7!*0!)}^*0.5^7*(1-0.5)^0 = 0.063. This p-value falls slightly
short of regular standards in the social sciences (p < .05). This is the same as the following
situation: When a coin comes up heads with probability 0.5 when tossed, the probability of
achieving 6 or more heads after 7 tosses is .063.

\textsuperscript{52} However, William O’Grady (personal communication) points out that an improvement from –
to ? in the manipulation tasks does not suggest a performance problem on the baseline pattern to
the same degree as an improvement from – to +.
### Table 4.10
**Individual evaluation: KHC**

<table>
<thead>
<tr>
<th></th>
<th>Scrambling + Case (Picture-selection comprehension task)</th>
<th>Case (Case-marker-elicitation production task)</th>
<th>Source(s) of chance or below chance performance in the baseline task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base Eval (max=7) Mamp. Eval Subj. Obj. Subj. error Obj. error</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KHC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>0 – 0 – ? ?</td>
<td>Other(s)</td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>0 – 0 – –</td>
<td>case</td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>0 – 0 – + ?</td>
<td>DAT Mapping</td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>0 – 1 – + ?</td>
<td>DAT Mapping</td>
<td></td>
</tr>
<tr>
<td>P5</td>
<td>0 – 3 – ?</td>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>P6</td>
<td>0 – 2 – + –</td>
<td>Mapping</td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td>0 – 0 – + ?</td>
<td>Other(s)</td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>0 – 1 – + –</td>
<td>Other(s)</td>
<td></td>
</tr>
<tr>
<td>P9</td>
<td>0 – 11 – +</td>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>P10</td>
<td>0 – 1 – –</td>
<td>case</td>
<td></td>
</tr>
<tr>
<td>Mid</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P11</td>
<td>0 – 5 – ?</td>
<td>DAT Performance</td>
<td></td>
</tr>
<tr>
<td>P12</td>
<td>1 – 0 – + +</td>
<td>Mapping</td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>1 – 2 – + +</td>
<td>Other(s)</td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>5 – 11 + ?</td>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>P15</td>
<td>0 – 5 – ?</td>
<td>DAT Performance</td>
<td></td>
</tr>
<tr>
<td>P16</td>
<td>0 – 1 – + +</td>
<td>Mapping</td>
<td></td>
</tr>
<tr>
<td>P17</td>
<td>1 – 8 – ?</td>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>P18</td>
<td>1 – 9 – ?</td>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>P19</td>
<td>0 – 5 – +</td>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>P20</td>
<td>0 – 4 – +</td>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>P21</td>
<td>2 – 9 – ?</td>
<td>DAT Mapping</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P22</td>
<td>2 – 7 – ?</td>
<td>NOM Mapping</td>
<td></td>
</tr>
<tr>
<td>P23</td>
<td>7 + 14 + +</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>P24</td>
<td>1 – 10 – ?</td>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>P25</td>
<td>3 – 9 – ?</td>
<td>Other(s)</td>
<td></td>
</tr>
<tr>
<td>P26</td>
<td>5 – 12 +</td>
<td>Performance</td>
<td></td>
</tr>
<tr>
<td>P27</td>
<td>6 + 13 + +</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>P28</td>
<td>7 + 14 + +</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>P29</td>
<td>7 + 14 + +</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>P30</td>
<td>6 + 14 + +</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>P31</td>
<td>7 + 13 + +</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

**Notes.** 
**Comprehension task:** + = presence of knowledge (11–14 correct out of 14, or 6–7 correct out of 7); ? = possibility of knowledge (3–10 correct out of 14, or 2–5 correct out of 7); − = absence of knowledge (0–2 correct out of 14, or 0–1 correct out of 7).  
**Production task:** + = presence of knowledge (5–6 correct out of 6); ? = possibility of knowledge (2–4 correct out of 6); − = absence of knowledge (0–1 correct out of 6).  
**Sources:** Performance = − in the baseline but + or ? in the manipulation tasks, or ? in the baseline but + in the manipulation tasks; Mapping = error in case; case = − both in subject-markings and object-markings; Other(s) = poor performance in the baseline task due to a reason that is not identified; N/A = not applicable.
**Table 4.11**

*Individual evaluation: cL1ers*

<table>
<thead>
<tr>
<th>cL1er</th>
<th>Scrambling + Case (Picture-selection comprehension task)</th>
<th>Case (Case-marker-elicitation production task)</th>
<th>Source(s) of poor performance in the baseline task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base (max=7) Eval</td>
<td>Manip. (max=14) Eval</td>
<td>Subj.</td>
</tr>
<tr>
<td>Low</td>
<td>P1</td>
<td>4</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>3</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>P3</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Mid</td>
<td>P4</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>P5</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>P6</td>
<td>4</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>P7</td>
<td>5</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>P8</td>
<td>7</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>P9</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>P10</td>
<td>4</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>P11</td>
<td>4</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>P12</td>
<td>3</td>
<td>?</td>
</tr>
<tr>
<td>High</td>
<td>P13</td>
<td>2</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>P14</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>P15</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>P16</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>P17</td>
<td>3</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>P18</td>
<td>7</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>P19</td>
<td>4</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>P20</td>
<td>5</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>P21</td>
<td>6</td>
<td>+</td>
</tr>
</tbody>
</table>

*Notes.* **Comprehension task:** + = presence of knowledge (11–14 correct out of 14, or 6–7 correct out of 7); ? = possibility of knowledge (3–10 correct out of 14, or 2–5 correct out of 7); – = absence of knowledge (0–2 correct out of 14, or 0–1 correct out of 7). **Production task:** + = presence of knowledge (5–6 correct out of 6); ? = possibility of knowledge (2–4 correct out of 6); – = absence of knowledge (0–1 correct out of 6). **Sources:** Performance = – in the baseline but + or ? in the manipulation tasks, or ? in the baseline but + in the manipulation tasks; Mapping = error in case; case = – both in subject-markings and object-markings; Other(s) = poor performance in the baseline task due to a reason that is not identified; N/A = not applicable.

Tables 4.10 and 4.11 indicate that out of the 31 KHC and 21 cL1ers, 25 KHC\(^{53}\) and 18 cL1ers\(^{54}\) did not pass the comparison-against-chance criterion in the baseline task (i.e., they are assigned either – or ?). Among these children, 3 KHC (P9, P14, and P26) and 2

\(^{53}\) The 25 KHC are all the KHC except P23, P27, P28, P29, P30, and P31.

\(^{54}\) The 18 cL1ers are all the cL1ers except P8, P18, and P21.
cL1ers (P11 and P19) passed the comparison-against-chance criterion in the facilitating tasks (i.e., they are assigned +). This suggests that their poor performance in the baseline task is due to a performance-related problem (i.e., perception failure or high processing load). That is, they have TL knowledge, but the knowledge was not revealed in the baseline task for performance-related reasons. In addition, performance on OSV in the baseline task by 8 KHC (P5, P11, P15, P17, P18, P19, P20, and P24) and 4 cL1ers (P5, P14, P15, and P16) is within the range of a deficit in knowledge (i.e., the range of −) but reaches the range of a debatable area in the manipulation tasks (i.e., the rage of ?). This suggests that although their OSV performance on the manipulation tasks was not sufficient to claim the presence of knowledge, their performance on OSV items in the facilitating tasks (vs. baseline task) nevertheless did improve.

Among the 25 KHC and 18 cL1ers who showed chance or below chance performance in the baseline task, 2 of the KHC (P2 and P10) and 2 of the cL1ers (P1 and P2) showed a knowledge deficit both in subject-markings and object-markings. This suggests that their poor performance might be due to a lack of knowledge of case. In addition, 9 KHC (P3, P4, P6, P11, P12, P15, P16, P21, and P22) and 7 cL1ers (P3, P4, P5, P10, P12, P13, and P14) made at least one error in the case-marker production tasks. This suggests that their poor performance in the baseline comprehension task might come from a problem in mapping case morphology onto Case features.

Poor performance can have various causes. For example, I attributed the poor performance of 2 KHC (P11 and P15) and 2 cL1ers (P5 and P14) to a performance problem due to a perception failure or high processing load. However, they also showed errors in the case-marker-elicitation task, which suggests that they might also suffer from a mapping problem. This suggests that poor performance can come from a combination of reasons.

4.5.4 Discussion.

The first research question is the following: When learners of Korean (i.e., KHC and Korean cL1ers) show poor performance with scrambled OSV sentences, what is the source of the poor performance? The results show that poor performance may have various sources: First, perception problems (such as a failure in case marker perception)
or heavy processing load (due to a lack of context) may be the reason: Three KHC (P9, P14, and P26) and 2 cL1ers (P11 and P19) showed chance or below chance performance in the baseline picture-selection task but showed significantly-above-chance performance in the facilitating tasks. In addition, performance on OSV in the baseline task by 8 KHC (P5, P11, P15, P17, P18, P19, P20, and P24) and 4 cL1ers (P5, P14, P15, and P16) is within the range of a deficit in knowledge (i.e., the range of –) but reaches the range of a debatable area in the manipulation tasks (i.e., the rage of ?). I suggested that their poorer performance in the baseline task (vs. the manipulation tasks) is due to a performance factor. Second, a deficit in knowledge of case may be the reason: Two KHC (P2 and P10) and 2 cL1ers (P1 and P2) showed a deficit in case in the case-marker-elicitation task. Third, a mapping problem between case morphology and Case feature may be the reason. Among the 25 KHC and 18 cL1ers showing chance or below chance performance in the baseline task, 9 KHC (P3, P4, P6, P11, P12, P15, P16, P21, and P22) and 7 cL1ers (P3, P4, P5, P10, P12, P13, and P14) fall into this category since they showed an error in case-marker production.

The second research question is the following: Do KHC show evidence of knowledge deficit in case? The results show that 2 KHC (P2, P10) do not know either subject or object case markings and that 6 additional KHC (P6, P8, P9, P11, P15, and P21) do not know object case marking. These children were born in Hawai‘i where the majority language is English, but their two parents spoke Korean to them at home. The children learned Korean from birth, which means that they learned Korean within the so-called critical period for language learning. The implication of these results will be discussed in Chapter 6.
CHAPTER V
THE ENGLISH TOUGH CONSTRUCTION

This chapter is concerned with the acquisition and processing of the English tough construction (henceforth, TC). First the grammar and processing of the English TC will be discussed, followed by discussion of the grammar and processing of the Korean translation equivalents. Then I will lay out a learnability problem facing Korean-speaking learners acquiring the English TC. Subsequently, I will discuss the performance-related sources of poor performance in the previous studies—agent-biased context and intransitively-biased optionally transitive verbs. I will then introduce the innovative methods that this study employed to test knowledge of the English TC. Finally, I will report the results of an experiment with native Korean-speaking L2ers of English, both L2 children and L2 adults, along with a native English-speaking adult control group.

5.1 Competence and Performance: The English TC

Consider the following English TC sentence (N. Chomsky, 1964, pp. 61-65).

(1) The boy is easy to please.

The boy is understood in (1) as a “pleasee,” not a “pleaser.” That is, to understand the English TC, it is necessary to have knowledge of tough movement, which means knowing that the matrix subject to be linked to the object gap in the subordinate clause. The next section discusses accounts of the English TC in terms of grammar and processing.

5.1.1 Grammar and Processing: Tough Movement.

The English TC is a construction in which a tough adjective (e.g., difficult, easy, hard, impossible, simple, tough) syntactically selects a subordinate infinitive clause missing an object and semantically selects for two arguments, one an entity and the other a proposition (Cormack, 1998). See (2a-b).
In (2a), the *tough* adjective, *easy*, subcategorizes a subordinate infinitive clause missing an object, and the missing object is connected to the matrix subject for interpretation. The *tough* adjective takes two arguments, an entity and a proposition. This can perhaps be more clearly seen from (2b): an entity (*the mother*) and a proposition (*PRO to please the boy*). The type of thematic role that is assigned to the entity argument of the *tough* adjective is *benefactive* (Wilder, 1991) or *experiencer* (Landau, 1999).

There have been debates over the syntactic operations that are involved in the English TC. On one view, it is taken as a case of A-movement (object-to-subject movement), with the missing object moved to the subject position of a higher clause (Rosenbaum, 1967). There is a problem with this approach, however, in that A-movement is motivated by Case assignment, but the base-generated position can be assigned accusative Case from the transitive verb. In opposition to this movement analysis, Lasnik and Fiengo (1974) proposed a deletion account in which the English TC is derived from a base structure such as *The boy is easy to please the boy*. In this sentence, the object argument of the embedded infinitive is deleted at PF under identity with the matrix subject NP. Later still, an A’-movement account was proposed, where a phonologically null *wh*-operator is postulated, as in (3) (Browning, 1987; N. Chomsky, 1977).

(3) The boy is easy [CP Op [TP PROarb to please ei]].

In (3), the TC subject, *the boy*, is base-generated in situ, and what moves from the object position of the subordinate clause is the operator (Op). This A’-movement view was put forward based on the fact that extraction from the TC infinitival violates the subadjacency principle, as is typically observed in overt *wh*-movement environments.
Hicks (2009) explained that (4) is ungrammatical because the Spec,CP position in the embedded infinitival clause is filled by the moved null operator, and this blocks use of the intermediate landing site for movement of the overt wh-phrase.

The proposal that the TC subject is base-generated in situ, however, is also problematic: It violates theta theory in that a single theta-role assigned by the embedded verb is shared between two arguments, the null operator in the infinitival and the TC subject. In this sense, within the Principles and Parameters model, TCs have remained “unexplained and in principle unexplainable” (Holmberg, 2000, p. 839).

More recent studies have posited that the derivation of the English TC involves two movement steps: A’-movement, and then A-movement (Brody, 1993; Hartman, 2011a, 2011b; Hicks, 2009; Hornstein, 2001). An operator moves from the object of the subordinate infinitival to the Spec,CP of the infinitival (A’-movement), “smuggling” the DP (which becomes the TC subject later in the derivation). Subsequently, the smuggled DP (the TC subject) itself undergoes A-movement independently of the null operator from the Spec,CP of the infinitival to the matrix subject position.

Evidence for A-movement was presented from the effect of the defective intervention constraint (N. Chomsky, 2000). The Italian A-movement sentence in (5b), for example, is ungrammatical when an intervening experiencer (a Maria) is present because it blocks movement to the subject position of the higher clause.

---

1 The defective intervention constraint is defined as “[α > β > γ] where > is c-command, β and γ match the probe α, but β is inactive so that the effects of matching are blocked” (N. Chomsky, 2000, p. 123) or “a phenomenon where some head (usually finite T) seeks a matching NP to agree with or attract, but some other NP intervenes” (Bruening, 2012, p. 1).
(5) a. Sembra (a Maria) che Gianni é stanco.
   Seems to Maria that Gianni is tired
   ‘It seems to Maria that Gianni is tired.’
b. Gianni sembra (*a Maria) essere stanco.
   Gianni seems to Maria to be tired.
   ‘Gianni seems to Maria to be tired.’

(Hartman, 2011b, p. 122, (2a-b))

Hartman (2011a) proposed that defective intervention is also observed in the English TC, as seen in (6b), which he argued shows that the matrix subject is derived by A-movement.

(6) a. It is important (to Mary) to avoid cholesterol.
b. Cholesterol is important (*to Mary) to avoid.

(Hartman, 2011a, p. 390, (10a-b))

As already mentioned, in the English TC, the object gap is connected to the matrix subject for the interpretation of the whole construction. This characteristic, however, is not unique to the English TC. There are other, similar constructions that include an object gap, as illustrated in (7).

(7) a. The boysi are [Degp too young [to teach ei]].
b. Butterfliesi are [Ap pretty [to look at ei]].

(Anderson, 2005, p. 33, (24c-d))

In (7a-b), the matrix subject NPs are connected to the object gap in the subordinate clause, as is the case in the English TC. The difference between the sentences in (7a-b) and the English TC is this: The grammatical matrix subject in (7a-b) is the thematic subject of the matrix predicate, but this is not the case in the English TC. In the TC sentence The boy is easy to please, the matrix predicate easy does not assign a thematic role to the grammatical matrix subject the boy. As mentioned, a tough adjective takes a proposition
as one of its arguments (and an entity as the other argument). Note that the boy is merely a component of the proposition, please the boy.

It has been claimed that the English TC is hard to process. Filipović and Hawkins (2013) proposed that “processing effort is minimized when structurally and semantically simple properties are used” (p. 164). They presented the English TC as an example involving complex properties, specifically displacement of an argument from a lower clause into a higher clause, where the displaced argument does not engage in a semantic relation with the matrix adjective. In addition, Filipović and Hawkins speculated that surface syntactic ambiguity between the TC construction (e.g., Elmo is easy to see) and the more common control pattern (e.g., Elmo is eager to see) can interfere with processing.

In this section, we discussed a variety of proposals on the English TC. What the various proposals uniformly indicate is that to understand the English TC, it is necessary to have knowledge of tough movement, which means knowing that the matrix subject must be linked to the object gap in the subordinate clause. The next section discusses the Korean translation equivalents of the English TC, which do not show tough movement.

5.1.2 Tough predicates in Korean.

It is generally agreed that Korean does not have a genuine English-like tough movement pattern (Chae, 1998; I. Kim, 2010; Y.-H. Kim, 2012; S.-H. Lee, 2003; Yeo, 2005). However, for every English pattern of the type X is easy to V, there is a corresponding pattern in Korean that looks similar.

(8) English
The professor is difficult to meet.

(9) Korean
Kyoswunim-i manna-ki-ka elyewu-si-ta.
professor-NOM meet-NML-NOM difficult-HON-DECL
‘It is difficult for the professor to meet someone.’
We know that the sentence-initial NP in (9) is the subject of the matrix predicate, just as its English counterpart is, because it is able to trigger the honorific agreement suffix -si in the matrix predicate—a reliable test for subjecthood in Korean (Hong, 1994). Notice, though, that the Korean pattern differs from its English counterpart in its interpretation: (9) means that it is difficult for the professor to meet someone (i.e., subject interpretation), not (as in English (8)) that it is difficult for someone to meet the professor (i.e., object interpretation).

Matters are further complicated by the fact that a less frequent sentence type, which is superficially very similar to (9), can have the interpretation ‘it is difficult for someone to meet the professor’. This is shown in (10).

(10) Kyoswunim-i manna-ki-ka elyep-ta.
        professor-NOM  meet-NML-NOM  difficult-DECL

‘The professor is difficult to meet.’

Crucially, however, the sentence-initial NP kyoswunim ‘professor’ in the pattern in (10) does not function as the subject, which is why there is no honorific agreement. (As I suggest in Appendix G, the real subject in this pattern is a null pronoun.)

When Korean learners of English encounter the English TC sentence in (8), the position of the NP the professor and the fact that it triggers agreement in the copula be suggest that the professor is the subject of the predicate difficult to meet. The closest Korean equivalent to the English TC is therefore the pattern in (9), with the subject interpretation. If transfer plays a role when Korean learners of English process the English TC, they will therefore incorrectly end up with a subject interpretation for the English pattern.

As for processing the Korean equivalents of the English TC, no study has examined whether there is a processing preference between the subject interpretation and the object interpretation. Table 5.1 summarizes the properties of the English TC and the Korean surface equivalents.
Table 5.1
Comparing the English TC and the Korean surface equivalents

<table>
<thead>
<tr>
<th></th>
<th>English TC</th>
<th>Korean surface equivalents</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grammar</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Tough movement</strong></td>
<td>Exists</td>
<td>Does not exist</td>
</tr>
<tr>
<td>Representation for the object interpretation</td>
<td>Possible</td>
<td>Possible</td>
</tr>
<tr>
<td>Representation for the subject interpretation</td>
<td>Impossible</td>
<td>Possible</td>
</tr>
<tr>
<td><strong>Processing</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Object interpretation processing allowed only</td>
<td>Both subject interpretation processing and object interpretation processing possible</td>
</tr>
</tbody>
</table>

### 5.1.3 Acquisition of the English TC by Korean-speaking learners.

For Korean learners of English, there is a learnability problem when acquiring the English TC. Acquisition of the English TC is underdetermined by both their L1 Korean and the TL input, and the English TC is not taught in the L2 English classroom.

First, as discussed in the previous section, the Korean language does not have *tough* movement. In addition, the learners must delearn the subject interpretation, which is a possible interpretation for the Korean surface equivalents.

Second, drawing an analogy from L2 input seems difficult. Crucially, input does not tell the learner what is *not* possible (Schwartz & Sprouse, 2013). Consider (11a-b).

(11)  
   a. The boy is eager to please.
   b. The boy is easy to please.

On the surface, (11a) and (11b) look similar to each other. However, native English speakers know that (11a) allows only the subject interpretation, while (11b), which is a *tough* construction, allows only the object interpretation. Now consider (12a-b).

(12)  
   a. The boy is $\text{[DegP too young [to teach]]}$.  
   b. The boy is $\text{[DegP old enough [to teach]]}$.  

91
On the surface, (12a) and (12b) look similar to the English TC. However, native English speakers know that both subject and object interpretations are possible in (12a-b). Now consider (13a-b).

\begin{align*}
\text{(13) a. The boy}_1 & \text{ is } [\text{DegP too easy [to teach e]}]. \\
\text{b. The boy}_1 & \text{ is } [\text{DegP easy enough [to teach e]}].
\end{align*}

(13a-b) look like (12a-b). However, native English speakers know that (13a-b) allow only the object interpretation because they are *tough* constructions. It is interesting to note that native English-speaking children eventually come to know all this, despite being exposed to input comprising these constructions that all look very similar.

Crucially, there is no input that tells the L2ers that the subject interpretation, which is possible in L1 Korean, is not possible for the English TC. What the grammar tells us is what is *possible* vs. *not possible* rather than what is more preferable vs. less preferable. Learners may come to know that the English TC allows the object interpretation based on the L2 input; however, it is hard to understand how the input could teach a learner to reject the subject interpretation.\footnote{I admit that input tells us that the English TC allows the object interpretation. For example, the placement of the object of a preposition in the matrix subject position (e.g., *Elmo is easy to talk to*) can indicate that the English TC has the object interpretation (Bonnie D. Schwartz and William O’Grady, personal communication). In addition, a construction such as *it is easy to learn French* can indicate that the English TC has an object interpretation (William O’Grady, personal communication). According to William O’Grady, more than one billion examples that start with *it is easy to* can be found in a Google search, and these examples indicate that PROarb serves as the subject of the verb, as in the English TC. Following his suggestion, I analyzed 200 such cases (chosen at random from the results of a Google search), and I found that about 90% of the cases started with the expletive *it* (e.g., *It is easy to learn French, It is easy to be green*), while the other 10% were the *tough* construction (e.g., *It is easy to assemble*). That is, the pattern *it is easy to* is frequently used in English, and in most cases, it is the pattern with the expletive *it*, where PROarb serves as the subject of the verb, which contributes to knowing that the English TC has PROarb. However, Bonnie D. Schwartz points out that an explanation is still required for how *It is easy to verb NP* tells learners that the TC has the object interpretation (i.e., *NP is easy to verb*).}

Anderson proposes (see below) that L1 English-acquiring children can have two possible grammars for the English TC—one for the subject interpretation and the other for the object interpretation. Anderson, citing Pinker (1984) and Baker (1979), considers the possibility that in order to rid the grammar of the subject interpretation, children make
use of *indirect negative evidence*. For example, when children notice that the meaning assigned by an adult (i.e., object interpretation of the English TC) is not the same as their own (i.e., subject interpretation of the English TC), their grammar for the subject interpretation could be marked for elimination. However, Anderson acknowledges that “it remains a subject of debate whether children actually make use of indirect negative evidence” (p. 432), stressing the need for further research.

Third, the English TC is not explicitly taught in the classroom in Korea. The Ministry of Education in Korea lists what secondary school students should learn in their English classes, and the English TC is not on the list (Ministry of Education, 2007, pp. 54–61) (see Appendix H for the list of constructions that are taught in Korean secondary schools). At the college level, English classes tend to focus more on promoting communicative skills than teaching specific English forms or constructions, to better meet the students’ needs (K. Kim, Yang, & Sung, 2013). In sum, the English TC must (most likely) be acquired through natural spoken or written English input rather than from instruction.

An L2 learnability problem arises: The Korean surface equivalent of the English TC allows the subject interpretation. In the absence of instruction and negative evidence, can L1-Korean L2ers of English come to know the English TC (i.e., abandon what gives rise to the subject interpretation)? It has been proposed that children have access to all possible grammars for the English TC that UG licenses, and converge on the adult grammar by eliminating the grammar(s) that are unlike the adult grammar (Anderson, 2005). How about L1-Korean L2 adults? Can they come to know the English TC, delearning the subject interpretation that is possible in Korean?

5.2 Previous Acquisition Studies

5.2.1 L1 Acquisition.


---

3 Every English textbook must pass a Ministry of Education evaluation to be used as a main text in Korean secondary schools. If a textbook does not follow the Ministry’s guidelines, it cannot pass the evaluation.
That is, the children erroneously interpret the sentence, linking the matrix subject with the subject of the subordinate infinitive clause, as in (14a); this will be called the “subject interpretation.” The correct interpretation, as in (14b), will be called the “object interpretation.”

(14) a. *The doll_i is hard [PRO_i to see PROarb].
   (subject interpretation: The doll = subject of the infinitival)
   b. The doll_i is hard [PROarb to see ei].
   (object interpretation: The doll = object of the infinitival)

Previous studies (e.g., Anderson, 2005; Cromer, 1970) have identified three stages in children’s acquisition of the English TC: (a) an initial subject-interpretation stage, as in (14a), (b) an intermediate stage showing fluctuation between subject interpretation and correct object interpretation, and (c) a final stage showing adult-like competence. One example of an experimental study whose findings support the three stages was done by Cromer (1970), who conducted an act-out task with 41 child speakers of British English (age range = 5;3 to 7;5). The child’s task was to show which animal did the biting in a series of sentences, such as *The duck is easy to bite vs. The duck is glad to bite. All of the test items featured the single verb to bite, and included one of two nouns, duck or wolf. Cromer claimed that the children could be divided into 3 groups on the basis of their performance: Primitive-rule Users (i.e., dominant subject interpretation); Intermediates; and Passers (i.e., dominant object interpretation). Cromer noted that at about age 6, children became aware that a non-named animal in a test sentence could be the actor, but they had not yet learned which adjectives (e.g., easy, glad) require which underlying structure.

C. Chomsky (1969) was the first to report that native English-speaking children performed poorly on the English TC. In her study, the children (n = 40; age range = 5;0–10;0) were shown a blindfolded doll and asked “Is the doll easy to see or hard to see?” The target response was “The doll is easy to see” because the doll was...
placed in a position where the children could easily see it.\textsuperscript{4} The results showed that 14 out of the 40 children incorrectly responded with “The doll is hard to see.” That is, they regarded the doll as a “see-er,” which is the incorrect subject interpretation. Almost all 5-year-olds answered incorrectly, and all the 9- to 10-year-olds answered correctly. The 6- to 8-year-olds’ responses were mixed.

Such results, however, do not clearly show whether the poor performance comes from lack of grammatical knowledge of the English TC, or whether the experimental situation itself misled children to the incorrect response. Fabian-Kraus and Ammon (1980) pointed to a serious methodological flaw in C. Chomsky’s experiment: Everything in the situation suggests that the question is about the doll’s seeing, which leads to the incorrect subject interpretation. They stated that “the blindfold functions as a \textit{misleading element} [emphasis KK] which children must overcome in order to arrive at the correct answer” (p. 402).

Fabian-Kraus and Ammon designed another experiment that attempted to avoid potential misleading factors. Children listened to a short story illustrated with cartoons and were then asked a series of questions that tested their knowledge of both the story and the English TC (e.g., “Who was hard to find?”). Among the 45 children (ages 4–6), 14% of the 4-year-olds, 91% of the 5-year olds, and 85% of the 6-year-olds responded correctly for all the TC items. This suggests that mastery of this construction takes place by age 5, somewhat earlier than the age implied by the results of C. Chomsky’s study.

In addition, Fabian-Kraus and Ammon noticed that children tended to perform better on items with obligatorily transitive verbs (e.g., 100% accuracy with the verb \textit{find}, in “Who was hard to find?”) than on those with optionally transitive verbs (e.g., 25% accuracy with the verb \textit{hear}, in “Who was hard to hear?”). Based on the finding that young children seldom make errors in their classification of transitive vs. intransitive verbs (Bowerman, 1974), Fabian-Kraus and Ammon assumed that 4-year-olds know the transitivity characteristic of verbs and speculated that the subject interpretation of \textit{who} is “blocked” if the embedded verb is obligatorily transitive (p. 409). That is, if a child

\footnotesize{
\begin{itemize}
  \item C. Chomsky used only this one TC item. Depending on the child’s answer, she asked further questions. Children who incorrectly answered “hard to see” were asked the following: “Would you make her easy to see?” “Why was she hard to see in the beginning?” “What did you do to make her easier to see?” Children who correctly answered “easy to see” were asked this: “Would you make her hard to see?”
\end{itemize}
}
knows that a particular verb is necessarily transitive, then in the question “Who was hard to verb?,” who has to link to the object because the object is required. In addition, based on the reported degree of transitivity (Gahl, Jurafsky, & Roland, 2004) of the 6 verbs in this study, I found a strong correlation between the degree of transitivity and the rate of response accuracy (correlation coefficient = .81).⁵

The strong correlation between transitivity and accuracy suggests that (intransitively-biased) optionally transitive verbs may be misleading. A similar example can be drawn from Macaruso, Shankweiler and Crain’s (1993) experiment. In a picture-selection task with 90 native English-speaking children (aged about 8), the children showed poor performance on only one item, The frog is impossible to jump. Note that jump is an optionally transitive verb with a strong intransitive bias: In the corpus analyses of both Connine, Ferreira, Jones, Clifton, and Frazier (1984) and Gahl et al. (2004), the transitive use of jump is less than 10%.

In sum, I propose that (at least) two sources of performance difficulty exist in the interpretation of the English TC. First, an infelicitous situation, as with C. Chomsky’s blindfolded doll, might be one misleading factor. Second, the use of (intransitively-biased) optionally transitive verbs such as jump might be another misleading element.

5.2.2 L2 Acquisition.

Acquisition of the English TC has also been studied in the L2 context, focusing exclusively on adult L2ers (Chiang & Costello, 1983; Cook, 1973; Cooper, Olshtain, Tucker, & Waterbury, 1979; d’Anglejan & Tucker, 1975; Yip, 1995). However, those studies exhibit some methodological shortcomings. First, most studies did not administer a reliable proficiency test. Second, some studies did not consider the L1 of their participants. This section discusses Cook (1973) and Cooper, Olshtain, Tucker, and Waterbury (1979), for examples.

⁵ The accuracy rates by verb are as follows: find, 100%; catch, 93%; save, 69%; draw, 53%; watch, 33%; and hear, 25%. The degree of transitivity (Gahl et al., 2004) for each verb is as follows: find, 97% (83/86); save, 86% (115/134); hear 81% (100/124); draw, 68% (47/69); and watch, 63% (50/79). Gahl et al. (2004) did not provide data on the item catch. Following the intuition of several native English speakers, I regard catch as an obligatorily transitive verb (except in baseball, where it means to take the position of catcher).
Cook (1973) was the first study to test knowledge of the English TC with adult L2ers. Using the same task that Cromer (1970) used, Cook conducted an experiment with 67 adult L2ers of English with diverse L1 backgrounds.\(^6\) He found that the L2ers could be divided into 3 groups, just like the children in Cromer’s study. Based on these results, Cook identified similar developmental stages as those of native English-speaking children: (a) an initial subject-interpretation stage, (b) an intermediate stage showing fluctuation between subject interpretation and correct object interpretation, and (c) a final stage showing native-like competence. From this Cook concluded that “there seems to be no certain evidence to show that adults are different from children in language learning, once the other attributes of the adult such as increased memory span have been cancelled out” (p. 28). However, this argument seems not compelling because the study did not consider L1 transfer and L2 proficiency at all. In particular, it is hard to discuss language development unless proficiency data are given.

Cooper, Olshtain, Tucker, and Waterbury (1979) claimed that learners find the English TC difficult regardless of their L1s. They conducted a comprehension task with 66 Arabic speakers and 50 Hebrew speakers who were learning English.\(^7\) The participants were asked to identify the agent in English TC sentences (e.g., The president is difficult to see. Who will see?) and in other types of sentences (e.g., Fred promised Harry to leave quickly. Who will leave?); the task was adapted from one used by d’Anglejan and Tucker (1975), who studied the comprehension of the English TC by native speakers of Canadian French. Cooper et al. reported that both the Arabic speakers and the Hebrew speakers showed far more errors on the English TC items than on the other types of sentences. However, they did not provide statistical analyses with their results. Without statistical analyses, the fact that L2ers do worse on the English TC items

\(^6\) The 67 L2ers’ L1s were as follows: Spanish (15), German (13), Persian (9), French (7), Polish (7), Italian (6), Arabic (2), Armenian (1), Chinese (1), Danish (1), Greek (1), Hebrew (1), Indonesian (1), Portuguese (1), and Serbo-Croat (1) (p. 24).

\(^7\) In fact, the authors divided the participants into 7 groups based on the places (e.g., English Language Institute in a university, high school, etc.) in which the participants were recruited: Arabic group 1 \((n = 18)\), Arabic group 2 \((n = 18)\), Arabic group 3 \((n = 15)\), Arabic group 4 \((n = 15)\), Hebrew group 5 \((n = 19)\), Hebrew group 6 \((n = 17)\), and Hebrew group 7 \((n = 14)\). The participants in the first 2 Arabic groups had proficiency test scores: Arabic group 1 (Michigan Test of English Language Proficiency [MTELFL] score range = 63–69) and Arabic group 2 (MTELFL score range = 74–81).
than other types of sentences does not necessarily mean that they do poorly on the English TC.

Cooper et al. further argued that the poor performances of their participants, as well as that of the Canadian French-speaking learners in d’Anglejan and Tucker’s study, suggest that the learners “grappled with the English material directly, without recourse to their first language” (p. 272). According to Cooper et al., if there had been an L1 effect, the Canadian French speakers should have shown better performance on the English TC than the Arabic/Hebrew-speaking learners, because French has a structure similar to the English TC, while Arabic and Hebrew do not. For example, *The president is difficult to see* is translated into French as *Le président est difficile à voir*, while the Arabic translation of *The pyramids are easy to see* would be the equivalent of *The pyramids are easy to be seen* (p. 271). However, because the studies did not present reliable L2 proficiency data, this argument does not seem compelling.

5.3 Methodological Innovations

To alleviate performance difficulty, both sentence-internal and sentence-external factors can be manipulated. The intention of the sentence-internal facilitation is to make the object gap in the subordinate clause more salient. This can be accomplished through the use of obligatorily transitive verbs in the subordinate clause. This should help the learners identify the object gap and, thus, help in their application of knowledge of the English TC. See (15a-b).

(15) a. Elmo is easy to race ___.
    b. Elmo is easy to find ___.

   The obligatorily transitive verb *find* signals this gap.

In (15a), the verb in the subordinate clause, *race*, belongs to an intransitively-biased, optionally transitive verb (Connine et al., 1984; Gahl et al., 2004). This verb does not contribute to the hearer’s ability to analyze the infinitival as having an object gap. In
contrast, the verb *find* in (15b) is an obligatorily transitive verb, much more obviously indicating an object gap.\(^8\)

The intention of the sentence-external facilitation is to introduce a property that is relevant to the correct object interpretation. This can be tested by making a comparison between a neutral context and a misleading context—what I called the “agent-based” context. The nature of the misleading contexts are similar to the one used in C. Chomsky’s (1969) study. More details are presented in Section 5.3.1.

5.3.1 TVJT.

I designed an experiment employing a Truth-Value Judgment task (TVJT) (Crain & Thornton, 1998) in which verb transitivity and context were both manipulated. The first manipulation was of verb transitivity, contrasting the participants’ interpretations of the English TC with obligatorily transitive verbs (i.e., *beat, catch, find, touch*) vs. optionally transitive verbs (i.e., *fight, hunt, move, race*). The optionally transitive verbs were carefully chosen from those with a transitivity bias of less than 50% (Connine et al., 1984; Gahl et al., 2004).

Second, context was manipulated. In a neutral context, an introduced property was relevant to—and thus contributed to—the correct interpretation of an English TC test sentence (TC subject = object of the subordinate clause). However, the context must be carefully designed to avoid having the relevant context information itself create a bias toward an object interpretation. To make participants judge test sentences based on their English TC knowledge, not an inference from the context information, the task depicted neutral situations where both a subject interpretation and an object interpretation are in principle possible. For example, in one set of stimuli in a neutral context, there are two referents, Bert and Elmo. Bert wears roller skates, but Elmo doesn’t. The test sentence is *Bert is easy to catch today*. In this situation, Bert can easily get away from Elmo due to the roller skates (Bert = the “catchee”), and at the same time, Bert can easily catch Elmo thanks to the roller skates (Bert = the “catcher”). That is, the roller skates give an

---

\(^8\) Both William O’Grady and Bonnie D. Schwartz noted that in a sentence such as *Elmo is easy to give advice to*, the preposition *to* strongly signals the gap. I agree with the point. However, this pattern was not used in this experiment’s materials, as the experiment required minimal pairs of the 2 conditions (obligatorily transitive vs. intransitively-biased, optionally transitive) as seen in (15a) and (15b).
advantageous property to the wearer no matter whether the wearer is a “catchee” or a “catcher.” In the correct object interpretation of Bert is easy to catch today, the response is NO,9 whereas in the incorrect subject interpretation, the response would be YES. See the example story in Figure 5.1.

<table>
<thead>
<tr>
<th>Context</th>
<th>Elmo is the fastest in the class. He is better than Bert at playing tag because he can run fast. He can chase Bert and run away from him. But today, Bert is wearing roller skates. He is very good at roller skating. When Bert and Elmo play tag, Bert is much faster than Elmo.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test sentence</td>
<td>The bear says, “Bert is easy to catch today.” (Target: NO) The bear asks “Why?” (Target: Bert has skates.)</td>
</tr>
</tbody>
</table>

*Figure 5.1. An example of a neutral context TVJT story.*

In the misleading agent-biased context, the depictions of situations included a property that created a strong bias toward a subject interpretation (TC subject = subject of the subordinate clause). For example, in one set of stimuli in an agent-biased context, there are two referents, Bert and Elmo. Elmo wears spider-man gloves and is good at catching Bert. However, on this particular day, Bert borrows the magic gloves from Elmo and is good at catching Elmo. The test sentence is Bert is easy to catch today. In this situation, the spider-man gloves contribute an advantage only when the wearer is a “catcher”; the spider-man gloves do not contribute an advantage for a “catchee” running away. In the correct object interpretation of the test sentence Bert is easy to catch today, the response is NO because compared with the past when Elmo wore the spider-man

---

9 For Bert is hard to catch today, the correct response is YES. In this way, target answers (True vs. False) were counterbalanced.
gloves, he would now have a hard time catching Bert.\(^{10}\) In the incorrect subject interpretation, the response would be YES because now Bert has the spider-man gloves and will be good at catching Elmo. See the sample story in Figure 5.2.

<table>
<thead>
<tr>
<th>Context</th>
<th>Spoken stimuli</th>
<th>Visual stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bert and Elmo like to play tag. Elmo has spider-man gloves. Elmo can hold anything easily when he is wearing the magic gloves. He is better than Bert at playing tag because his gloves make him good at grabbing hold of Bert. Today, Bert has borrowed the magic gloves from Elmo. Bert and Elmo are playing tag. Today, Bert is better than Elmo at playing tag.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test sentence</td>
<td>The bear says, “Bert is easy to catch today.”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Target: NO)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The bear asks, “Why?”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Target: Before, Elmo used the gloves. Today, he doesn’t.)</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 5.2. An example of an agent-biased context TVJT story.*

### 5.4 Research Questions and Hypotheses

I ask the following research questions:

1. Do adult English L2ers and child English L2ers (L1 = Korean) come to know that the English TC matrix subject is linked to the object gap in the embedded clause? In particular, are they able to delearn the subject interpretation that is possible in Korean?
2. Are learners of English able to apply their knowledge of the English TC effectively and systematically in comprehension?

\(^{10}\) For *Bert is hard to catch today*, the correct response is YES.
The predictions with respect to the TVJT are as follows:

1. If participants know the English TC, they will show good performance (i.e., significantly-above-chance performance) on the English TVJT.

2. If participants know the English TC and their use of the knowledge is target-like, they will show the correct object interpretation consistently, even in the items where intransitively-biased optionally transitive verbs or agent-biased contexts are used. If participants know the English TC but have an imperfect command of the knowledge, they will show good performance in the items in which verb transitivity and/or context have been manipulated not to lead to the incorrect subject interpretation.

5.5 The Current Study

5.5.1 Participants.

Forty-nine adult native speakers of Korean (ages = 19–26) and 30 child native speakers of Korean (ages = 9–12; mean 11) who were learning English as a second language in Korea participated. All the adult L2ers were university students in Korea at the time of testing, none majoring in English or linguistics. Six of the adults had lived in an English-speaking country for 1 year as exchange students after they entered university. Except for those 6, no other adult participant had learned English in an immersion setting. All the adult L2ers had learned English in the classroom from age 9. The total amount of instruction time for English in school before they entered university was about 730.5 hours (B. Lee, 2003, p. 119). In addition to school instruction, they had learned English by reading additional English materials or joining private English-education institutes. Their English education at school or/and private institutes aimed to prepare them for the college entrance English exam that requires test takers to understand 50 English passages (100–150 words in each passage) in 70 minutes. Korean secondary students train themselves to understand English passages as quickly and accurately as possible. At university, the English requirement for all students is to pass at least one English course.

11 Lee calculated the time based on the following information: English instruction time for 3rd grade and 4th grade in elementary school is one hour a week and for 5th grade and 6th grade is 2 hours a week. English instruction time for first- and second-year middle school students is 3 hours a week. Third-year middle school students and all high school students study English about 4 hours a week.
The course focuses on promoting English communicative skills. Private English-education institutes for university students are intended to either prepare them for standardized tests (e.g., TOEFL, TOEIC) or improve their communicative skills, which may be important for their future jobs.

All the 30 L2 children started learning English from the ages of 4 or 5 (mean = 5.2). The mean exposure to English for the child L2ers was about 5.7 years. The child L2ers learned English from private English-immersion institutes in Korea where most of the teachers were English native speakers. At the time of testing, all the participants were enrolled in the highest proficiency class. In the institute, children aged 4–6 learned English for 5 hours a day from Monday to Friday (mean exposure to English = about 1,000 hours a year). Children aged over 7 went to the institute after their classes at elementary school. They learned English for about 2 or 3 hours a day from Monday to Friday in the institute (mean exposure to English = about 500 hours a year). Five of the children had stayed in the United States: 3 children for 2 years at ages 5 and 6; 1 child for 2 years at ages 7 and 8; and 1 child for 1 year at age 5. In the interview that was conducted after participants completed the set of experimental tasks, no participant said that s/he had been instructed in the grammar of the English TC. I gave examples of English TC sentences (e.g., *Elmo is easy to see, Big Bird is hard to see*) to the participants and then asked them to talk about their experiences of the English instruction they had had before, with respect to the grammar of the given English sentence type.

A picture-narration task (Song & Schwartz, 2009) was administered to measure the participants’ English proficiency. Both adult L2er and child L2er groups were divided into 3 subgroups based on their proficiency scores, calculated using the protocol detailed in Chapter 3: Low (z-score below -0.5), Medium (z-score between -0.5 and 0.5) and High (z-score higher than 0.5) (see Appendix I for the proficiency data). Table 5.2 shows the information for each proficiency group.
Table 5.2
Participant grouping by proficiency

<table>
<thead>
<tr>
<th></th>
<th>Adult L2ers (n = 49)</th>
<th></th>
<th>Child L2ers (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>z-score mean (SD)</td>
<td>Number</td>
</tr>
<tr>
<td>Low (z-score &lt; -0.5)</td>
<td>13</td>
<td>-1.21 (0.51)</td>
<td>11</td>
</tr>
<tr>
<td>Medium (-0.5 &lt; z-score &lt; 0.5)</td>
<td>18</td>
<td>-0.06 (0.32)</td>
<td>10</td>
</tr>
<tr>
<td>High (0.5 &lt; z-score)</td>
<td>18</td>
<td>1.15 (0.58)</td>
<td>9</td>
</tr>
</tbody>
</table>

The adult and child L2ers completed the English TVJT and then the Korean TVJT, with a short break of about 5-10 minutes between the 2 tasks, and thus their performance on the Korean TVJT could have been affected by their experience on the English TVJT. Ten adult native speakers of Korean (ages = 20–28) participated in the Korean version of the TVJT, the purpose of which was to test for preferences in the Korean equivalents to the English TC (not affected by the English TVJT). These 10 participants did not take the English TVJT. In addition, 10 native English-speaking adults also participated in the English TVJT experiment as a control group.

5.5.2 Materials.

In the TVJT, participants judged whether a test sentence matched a situation or not. The TVJT was a 2x2 design, varying context (neutral vs. agent-biased) and transitivity of the verb in the subordinate clause (obligatorily transitive vs. intransitively-biased optionally transitive). Two tasks were administered in a sequence: English TVJT and then Korean TVJT (i.e., using the translation equivalents of the English TVJT). Eight verbs were used in each task: 4 obligatorily transitive verbs (beat, catch, find, touch) and 4 optionally transitive verbs (fight, hunt, move, race). The transitivity bias of the optionally transitive verbs was less than 50%: fight (23.5%, 32/136), hunt (45.2%, 71/157), move (34.0%, 32/94), and race (36.3%, 56/154) (Gahl et al., 2004). Each verb was used twice, in 2 different items, where the context types were different (neutral vs.

---


---

104
agent-biased), resulting in 16 test sentences in each task. Target answers (True vs. False) were counterbalanced. Each task also included 8 fillers, which were 4 passive sentences and 4 active sentences, as well as 2 practice items (see Appendix J for all these items). Two lists were created so that the target answers were different from each other. This was accomplished by using the adjectives easy and hard. For instance, if the target response is YES for the test sentence Bert is easy to see today, it is NO for Bert is hard to see today. Half of the participants completed one list; the other half completed the other list. For each list, the presentation sequence was randomized; however, care was taken so that no 2 consecutive items included the same verb and no 3 consecutive items included the same condition. This was to avoid carry-over from one trial to the next. For each list, half of the participants were given the sentences in the initial randomized order, while the other half were given the sentences in reversed randomized order.

To investigate whether the participants distinguished between the 2 verb types in English and in Korean, an English Acceptability-Judgment task (AJT) and a Korean AJT were conducted on the verbs in the TVJT's after the participants completed the main TVJT experiments. There was a break of about 5–10 minutes between the TVJT and the AJT. In the English AJT, they read 8 sentences missing a direct object NP (e.g., Kim is fighting) and 8 sentences including a direct object NP (e.g., Kim is fighting Lee), and rated the sentences’ acceptability on a scale of 1 (sounds completely unnatural) to 6 (sounds completely natural). The Korean AJT, which used the exact translation equivalents of the English AJT items (see footnote 12 in p. 104), was administered right after the English AJT (see Appendix K for the AJT items).

All participants also completed an English proficiency test, which was a picture-narration task (Song & Schwartz, 2009), as well as a language background questionnaire. For adult L2 participants, a cloze test (Brown, 1980) was included as an additional language proficiency measurement (see Appendix L for the cloze-test items). A Pearson correlation showed that there was a strong correlation between the cloze-test scores (mean = 28.0, range = 9–44) and the picture-narration task scores (correlation coefficient = 0.61) for the adult L2ers.
5.5.3 Results.

5.5.3.1 AJT.

Figure 5.3 shows the acceptability ratings of the native English-speaking control group ($n = 10$) and the native Korean-speaking control group ($n = 10$).

![Acceptability ratings for verbs in transitive vs. intransitive sentences: English native speakers and Korean native speakers.](image)

**English AJT**

**Korean AJT**

*Figure 5.3. Acceptability ratings for verbs in transitive vs. intransitive sentences: English native speakers and Korean native speakers.*

**Notes.** Obl-Tran = Obligatorily transitive verbs used in transitive sentences; Obl-Intran = Obligatorily transitive verbs used in intransitive sentences; Intopt-Tran = Intransitively-biased, optionally transitive verbs used in transitive sentences; Intopt-Intran = Intransitively-biased, optionally transitive verbs used in intransitive sentences; error bars represent standard errors.

Figure 5.3 shows that the English native speakers ($n = 10$) know that the obligatorily transitive English verbs cannot be used in intransitive sentences. Note that the average acceptability rating for the intransitive sentences including obligatorily transitive verbs is far below the midpoint of the 6-point Likert scale, i.e., below 3.5. However, the Korean native speakers ($n = 10$) tended to allow the translation equivalents of the obligatorily transitive English verbs to be used in intransitive sentences, where the average is around 3.5. This is probably due to the fact that Korean is a language allowing null arguments. In addition, the Korean speakers showed a preference for transitive sentences over intransitive sentences. The averages for intransitive sentences are around 3.5, while the averages for transitive sentences are over 5.

In the English AJT, a repeated measures ANOVA shows main effects of verb type [$F_1(1, 9) = 449.672, p < .001; F_2(1, 6) = 892.826, p < .001$] and sentence type [$F_1(1, 9) = 174.400, p < .001; F_2(1, 6) = 525.287, p < .001$] as well as an interaction effect [$F_1(1,9)$]
= 231.669, \( p < .001 \); \( F_1(1, 6) = 601.599, p < .001 \). The main effects of verb type and sentence type come from the fact that the English native speakers gave significantly lower ratings to the ungrammatical sentences (i.e., intransitive sentences with obligatorily transitive verbs) than to the grammatical sentences. The interaction effect indicates that the participants distinguish between the 2 verb types: A planned comparison reveals that in transitive sentences, there is no difference between the 2 verb types \( t_1(9) = -.124, p = .904; t_2(3) = -.129, p = .905 \), while in intransitive sentences, the obligatorily transitive English verbs sound worse than the intransitively-biased optionally transitive verbs \( t_1(9) = -27.463, p < .001; t_2(3) = -65.054, p < .001 \).

In the Korean AJT, a repeated measures ANOVA shows a main effect of sentence type \( [F_1(1, 9) = 42.971, p < .001; F_2(1, 6) = 34.002, p < .01] \), an effect of interaction in the by-participant analysis only \( [F_1(1, 9) = 9.344, p = .014; F_2(1, 6) = 2.425, p = .170] \), but no main effect of verb type \( [F_1(1, 9) = .414, p = .536; F_2(1, 6) = .319, p = .593] \). The main effect of sentence type shows that transitive sentences sounded more natural than intransitive sentences. The null effect of verb type suggests that the selected translation equivalents of the English verbs were not similar to the English verbs in that every transitive verb in Korean, a null-argument language, can have a phonetically null object. The interaction effect that appeared in only the by-participant analysis suggests that it was not a robust effect. However, it is a sign of difference between the 2 verb types in terms of their tendency to occur in the sentence types (i.e., transitive vs. intransitive). A planned comparison revealed that in transitive sentences, the Korean equivalents of the obligatorily transitive English verbs sounded more natural than the Korean equivalents of the intransitively-biased optionally transitive English verbs \( t_1(9) = 3.948, p < .01; t_2(3) = 2.000, p = .139 \), while in intransitive sentences, the Korean equivalents of the intransitively-biased optionally transitive English verbs sounded more natural only numerically than the Korean equivalents of the obligatorily transitive English verbs \( t_1(9) = -1.136, p = .285; t_2(3) = -1.033, p = .378 \).

Now let us look at the results from the target L2 adult and L2 child groups. To examine the performance of the 2 target groups in the English AJT, a 2 (verb type) x 2 (sentence type) x 3 (proficiency group) x 2 (learner group) ANOVA was conducted (within group factors = verb type and sentence type; between group factors = proficiency.
This analysis yielded main effects of verb type \(F_1(1, 73) = 7.901, p < .01; F_2(1, 36) = 4.062, p = .051\), sentence type \(F_1(1, 73) = 31.540, p < .001; F_2(1, 36) = 36.587, p < .001\) and learner group \(F_1(1, 73) = 19.745, p < .001; F_2(1, 36) = 30.100, p < .001\], as well as the following interaction effects: verb type x sentence type \(F_1(1, 73) = 189.708, p < .001; F_2(1, 36) = 79.562, p < .001\], verb type x learner group \(F_1(1, 73) = 12.936, p < .01; F_2(1, 36) = 6.651, p = .014\], and verb type x sentence type x learner group \(F_1(1, 73) = 21.873, p < .001; F_2(1, 36) = 9.174, p < .01\). To examine the performance of the 2 target groups on the Korean AJT, again a 2 (verb type) x 2 (sentence type) x 3 (proficiency group) x 2 (learner group) ANOVA was conducted (within group factors = verb type and sentence type; between group factors = proficiency group and learner group). This analysis yielded a main effect of sentence type \(F_1(1, 73) = 98.766, p < .001; F_2(1, 36) = 643.772, p < .001\] as well as an interaction effect between verb type and sentence type \(F_1(1, 73) = 92.271, p < .001; F_2(1, 36) = 110.016, p < .001\]. To investigate further specific patterns for the 2 L2 groups (adult L2ers vs. child L2ers), separate repeated-measures ANOVAs were conducted for each group. A 2 (verb type) x 2 (sentence type) x 3 (proficiency group) ANOVA was conducted (within group factors = verb type and sentence type; between group factor = proficiency group). For the English AJT of the adult L2ers, there were main effects of verb type \(F_1(1, 46) = 23.171, p < .001; F_2(1, 18) = 8.883, p < .01\] and sentence type \(F_1(1, 46) = 49.915, p < .001; F_2(1, 18) = 16.396, p < .01\], as well as an interaction effect between verb type and sentence type \(F_1(1, 46) = 168.612, p < .001; F_2(1, 18) = 54.250, p < .001\]. No main effect of proficiency group \(F_1(1, 46) = 1.369, p = .264; F_2(2, 18) = 1.092, p = .357\] was found. For the Korean AJT of the adult L2ers, the results showed a main effect of sentence type \(F_1(1, 46) = 95.235, p < .001; F_2(1, 18) = 252.753, p < .001\] and an interaction effect between verb type and sentence type \(F_1(1, 46) = 36.673, p < .001; F_2(1, 18) = 18.419, p < .001\] were found. However, there was no main effect of verb type \(F_1(1, 46) = 2.807, p = .101; F_2(1, 18) = .781, p = .388\] or proficiency group \(F_1(1, 46) = .954, p = .393; F_2(2, 18) = .594, p = .562\]. For the English AJT of the child L2ers, the results showed a main effect of sentence type \(F_1(1, 27) = 5.696, p = .024; F_2(1, 18) = 22.337, p < .001\] and an interaction effect between verb type and sentence type \(F_1(1, 27) = 72.559, p < .001; F_2(1, 18) = 25.361,
There was no main effect of verb type \( F_1(1, 27) = .340, p = .565; F_2(1, 18) = .196, p = .664 \) or proficiency group \( F_1(1, 27) = .366, p = .697; F_2(2, 18) = 1.118, p = .349 \). For the Korean AJT of the child L2ers, the results showed a main effect of sentence type \( F_1(1, 27) = 24.472, p < .001; F_2(1, 18) = 871.194, p < .001 \) and an interaction effect between verb type and sentence type \( F_1(1, 27) = 53.343, p < .001; F_2(1, 18) = 330.551, p < .001 \). However, there was no main effect of verb type \( F_1(1, 27) = .302, p = .587; F_2(1, 18) = .118, p = .735 \) or proficiency group \( F_1(1, 27) = .349, p = .708; F_2(2, 18) = .552, p = .602 \).

These analyses suggest that in the English AJT and Korean AJT, both child and adult groups showed the same patterns regardless of proficiency. However, in order to compare the results from the TVJT by each proficiency group, I present the data for each proficiency group below.

Figure 5.4 shows the acceptability ratings for the English AJT and Korean AJT by the lowest proficiency adult L2er group (henceforth, aL2ers-Low).

![Figure 5.4. Acceptability ratings for verbs in transitive vs. intransitive sentences: aL2ers-Low.](image)

Notes. Obl-Tran = Obligatory transitive verbs used in transitive sentences; Obl-Intran = Obligatory transitive verbs used in intransitive sentences; Intopt-Tran = Intransitively-biased, optionally transitive verbs used in transitive sentences; Intopt-Intran = Intransitively-biased, optionally transitive verbs used in intransitive sentences; error bars represent standard errors.

Figure 5.4 shows that the aL2ers-Low group knows that the obligatorily transitive English verbs cannot be used in intransitive sentences; their average acceptability rating
for the intransitive sentences including obligatorily transitive verbs is below 3.5. However, they tended to allow the Korean translation equivalents of the obligatorily transitive English verbs to be used in intransitive sentences, where the average is around 3.5.

In the English AJT of the aL2ers-Low, a repeated measures ANOVA shows main effects of verb type in the by-item analysis only \( F_1(1, 12) = 2.930, p = .113; F_2(1, 6) = 6.347, p = .045 \) and of sentence type in the by-participant analysis only \( F_1(1, 12) = 9.397, p = .010; F_2(1, 6) = 2.477, p = .167 \) as well as an effect of interaction \( F_1(1, 12) = 27.184, p < .001; F_2(1, 6) = 11.929, p = .014 \). The less robust main effects of verb type and sentence type are due to the fact that the difference between the rating for ungrammatical intransitive sentences with obligatorily transitive verbs and the ratings for the other grammatical sentences was not large enough. The interaction effect indicates that the participants distinguish the 2 verb types: A planned comparison revealed that in transitive sentences, obligatorily transitive verbs sounded more natural than intransitively-biased optionally transitive verbs \( t_1(12) = 1.972, p = .072; t_2(3) = 2.000, p = .139 \), while in intransitive sentences, the intransitively-biased optionally transitive verbs sounded significantly better than the obligatorily transitive verbs \( t_1(12) = -7.459, p < .001; t_2(3) = -8.720, p < .01 \).

In the Korean AJT of the aL2ers-Low, a repeated measures ANOVA shows a main effect of sentence type \( F_1(1, 12) = 35.116, p < .001; F_2(1, 6) = 111.468, p < .001 \), a marginal effect of interaction in the by-participant analysis only \( F_1(1, 12) = 3.984, p = .069; F_2(1, 6) = 3.523, p = .110 \), but no main effect of verb type \( F_1(1, 12) = .728, p = .410; F_2(1, 6) = .342, p = .580 \). The main effect of sentence type shows that transitive sentences sounded more natural than intransitive sentences. The marginal interaction effect suggests that the 2 verbs types are different in terms of their tendency to appear in a sentence type, although the effect is not robust. A planned comparison revealed that in transitive sentences, the Korean equivalents of the obligatorily transitive English verbs sounded more natural than the Korean equivalents of the intransitively-biased optionally transitive verbs \( t_1(12) = 3.077, p = .010; t_2(3) = 2.574, p = .082 \), while in intransitive sentences, the Korean equivalents of the intransitively-biased optionally transitive verbs sounded more natural only numerically \( t_1(12) = -.415, p = .686; t_2(3) = -.251, p = .818 \).
Figure 5.5 shows the acceptability ratings for the English AJT and Korean AJT by the medium proficiency adult L2er group (henceforth, aL2ers-Mid).

\[ F_1(1, 17) = 7.969, \quad p = .012; \quad F_2(1, 6) = 3.044, \quad p = .132 \]

In the English AJT of the aL2ers-Mid group, a repeated measures ANOVA shows a main effect of sentence type in the by-participant analysis only \([F_1(1, 17) = 7.969, \quad p = .012; \quad F_2(1, 6) = 3.044, \quad p = .132]\), a marginal main effect of verb type in the by-participant analysis only \([F_1(1, 17) = 3.680, \quad p = .072; \quad F_2(1, 6) = .753, \quad p = .419]\), and an effect of interaction \([F_1(1, 17) = 113.971, \quad p < .001; \quad F_2(1, 6) = 18.112, \quad p < .01]\). The less robust main effects of verb type and sentence type suggest that the difference between the rating for ungrammatical intransitive sentences with obligatorily transitive verbs and the ratings for the other grammatical sentences was not large enough. The interaction effect indicates that the participants distinguish between the 2 verb types: A planned comparison revealed that in transitive sentences, obligatorily transitive verbs sounded more natural.
than intransitively-biased optionally transitive verbs \([t_1(17) = 7.092, p < .001; t_2(3) = 1.950, p = .146]\), while in intransitive sentences, the intransitively-biased optionally transitive verbs sounded significantly better than the obligatorily transitive verbs \([t_1(17) = -7.277, p < .001; t_2(3) = -8.001, p < .01]\).

In the Korean AJT of the aL2ers-Mid, a repeated measures ANOVA shows a main effect of sentence type \([F_1(1, 17) = 20.279, p < .001; F_2(1, 6) = 65.722, p < .001]\) and an effect of interaction \([F_1(1, 17) = 23.202, p < .001; F_2(1, 6) = 12.538, p = .012]\), but no main effect of verb type \([F_1(1, 12) = 1.645, p = .217; F_2(1, 6) = .708, p = .432]\). The main effect of sentence type shows that transitive sentences sounded more natural than intransitive sentences. The interaction effect suggests that the 2 verb types are different in terms of their tendency to occur in a sentence type. A planned comparison revealed that in transitive sentences, the Korean equivalents of the obligatorily transitive English verbs sounded more natural than the Korean equivalents of the intransitively-biased optionally transitive verbs \([t_1(17) = 3.638, p < .01; t_2(3) = 2.361, p = .099]\), while in intransitive sentences, the Korean equivalents of the intransitively-biased optionally transitive verbs sounded more natural only numerically \([t_1(17) = -1.192, p = .250; t_2(3) = -.692, p = .539]\).

Figure 5.6 shows the acceptability ratings for the English AJT and Korean AJT by the highest proficiency adult L2er group (henceforth, aL2ers-High).

![Chart of acceptability ratings for English AJT and Korean AJT](image)

**Figure 5.6.** Acceptability ratings for verbs in transitive vs. intransitive sentences: aL2ers-High.

**Notes.** Obl-trans = Obligatorily transitive verbs used in transitive sentences; Obl-Int = Obligatorily transitive verbs used in intransitive sentences; Intopt-trans = Intransitively-biased, optionally transitive verbs used in transitive sentences; Intopt-Int = Intransitively-biased, optionally transitive verbs used in intransitive sentences; error bars represent standard errors.
The figure shows that the aL2ers-High group knows that the obligatorily transitive English verbs cannot be used in intransitive sentences; their average acceptability rating for the intransitive sentences with obligatorily transitive verbs is below 3.5. However, they tended to allow the Korean translation equivalents of the obligatorily transitive English verbs to be used in intransitive sentences, where the average is over 3.5.

In the aL2ers-High English AJT, a repeated measures ANOVA shows main effects of verb type \([F_1(1, 17) = 33.863, p < .001; F_2(1, 6) = 6.412, p = .045]\) and sentence type \([F_1(1, 17) = 59.420, p < .001; F_2(1, 6) = 17.281, p < .01]\) as well as an interaction effect \([F_1(1, 17) = 61.481, p < .001; F_2(1, 6) = 29.205, p < .01]\), which is the native pattern (see Figure 5.3). The main effect of verb type shows that intransitively-biased optionally transitive verbs sounded more natural than obligatorily transitive verbs. In fact, this is due to the presence of the obligatorily transitive verbs appearing in the (ungrammatical) intransitive sentences. The main effect of sentence type shows that transitive sentences sounded more natural than intransitive sentences. The interaction effect indicates that the participants distinguish between the 2 verb types: A planned comparison revealed that in transitive sentences, obligatorily transitive verbs sounded more natural than intransitively-biased optionally transitive verbs \([t_1(17) = 3.881, p < .01; t_2(3) = 1.984, p = .141]\), while in intransitive sentences, the intransitively-biased optionally transitive verbs sounded significantly better than the obligatorily transitive verbs \([t_1(17) = -8.969, p < .001; t_2(3) = -9.475, p < .01]\).

In the aL2ers-High Korean AJT, a repeated measures ANOVA shows a main effect of sentence type \([F_1(1, 17) = 44.371, p < .001; F_2(1, 6) = 79.818, p < .001]\) and an effect of interaction in only the by-participant analysis \([F_1(1, 17) = 17.339, p < .01; F_2(1, 6) = 4.784, p = .071]\), but no main effect of verb type \([F_1(1, 17) = .484, p = .496; F_2(1, 6) = .041, p = .847]\). The main effect of sentence type shows that transitive sentences sounded more natural than intransitive sentences. The interaction effect suggests that the 2 verb types are different in terms of their tendency to appear in a sentence type. A planned comparison revealed that in transitive sentences, the Korean equivalents of the obligatorily transitive English verbs sounded more natural than the Korean equivalents of the intransitively-biased optionally transitive verbs \([t_1(17) = 3.455, p < .01; t_2(3) = 1.870, p = .158]\), while in intransitive sentences, the Korean equivalents of the
intransitively-biased optionally transitive verbs sounded more natural \(t_1(17) = -2.166, p = .045; t_2(3) = -0.439, p = .690\).

Now let us look at the AJT results from the child L2ers. Figure 5.7 shows the acceptability ratings for the English AJT and the Korean AJT by the lowest proficiency child L2er group (henceforth, cL2ers-Low).

![Figure 5.7. Acceptability ratings for verbs in transitive vs. intransitive sentences: cL2ers-Low.](image)

**Notes.** Obl-trans = Obligatorily transitive verbs used in transitive sentences; Obl-Int = Obligatorily transitive verbs used in intransitive sentences; Intopt-trans = Intransitively-biased, optionally transitive verbs used in transitive sentences; Intopt-Int = Intransitively-biased, optionally transitive verbs used in intransitive sentences; error bars represent standard errors.

Figure 5.7 shows that the cL2ers-Low group did not know that the obligatorily transitive English verbs cannot be used in intransitive sentences; their average acceptability rating for the intransitive sentences with obligatorily transitive verbs is over 3.5. Likewise, in the Korean AJT, they tended to allow the translation equivalents of the obligatorily transitive English verbs to be used in intransitive sentences, where again the average is over 3.5.

In the cL2ers-Low English AJT, a repeated measures ANOVA shows no main effect of either verb type \(F_1(1, 10) = 2.034, p = .184; F_2(1, 6) = 1.141, p = .326\) or sentence type \(F_1(1, 10) = .317, p = .586; F_2(1, 6) = .973, p = .362\), but there is an interaction effect \(F_1(1, 10) = 20.729, p < .01; F_2(1, 6) = 4.430, p = .080\). The interaction effect indicates that the participants distinguish between the 2 verb types: A planned comparison revealed that in transitive sentences, obligatorily transitive verbs sounded more natural.
than intransitively-biased optionally transitive verbs \( t_1(10) = 3.048, p = .012; t_2(3) = 2.061, p = .129 \), while in intransitive sentences, the intransitively-biased optionally transitive verbs sounded better than the obligatorily transitive verbs \( t_1(10) = -1.875, p = .090; t_2(3) = -1.681, p = .191 \).

In the cL2ers-Low Korean AJT, a repeated measures ANOVA shows a (marginal) main effect of sentence type \( F_1(1, 10) = 4.554, p = .059; F_2(1, 6) = 157.538, p < .001 \), but not of verb type \( F_1(1, 10) = .079, p = .785; F_2(1, 6) = .047, p = .836 \), and an effect of interaction \( F_1(1, 10) = 15.277, p < .01; F_2(1, 6) = 129.385, p < .001 \). The (marginal) main effect of sentence type shows that transitive sentences sounded more natural than intransitive sentences. The interaction effect suggests that the 2 verbs types are different in terms of their tendency to occur in a sentence type. A planned comparison revealed that in transitive sentences, the Korean equivalents of the obligatorily transitive English verbs sounded more natural than the Korean equivalents of the intransitively-biased optionally transitive verbs \( t_1(10) = 6.094, p < .001; t_2(3) = 1.857, p = .160 \), while in intransitive sentences, the Korean equivalents of the intransitively-biased optionally transitive verbs sounded more natural \( t_1(10) = -1.789, p = .104; t_2(3) = -2.509, p = .087 \).

Figure 5.8 shows the acceptability ratings for the English AJT and the Korean AJT by the medium proficiency child L2er group (henceforth, cL2ers-Mid).

---

**Notes.** Obl-trans = Obligatorily transitive verbs used in transitive sentences; Obl-Int = Obligatorily transitive verbs used in intransitive sentences; Intopt-trans = Intransitively-biased, optionally transitive verbs used in transitive sentences; Intopt-Int = Intransitively-biased, optionally transitive verbs used in intransitive sentences; error bars represent standard errors.
Figure 5.8 shows that the cL2ers-Mid group did not know that the obligatorily transitive English verbs cannot be used in intransitive sentences; their average acceptability rating for the intransitive sentences with obligatorily transitive verbs is over 3.5. Likewise, in the Korean AJT, they tended to allow the translation equivalents of the obligatorily transitive English verbs to be used in intransitive sentences, where the average is over 3.5.

In the cL2ers-Mid English AJT, a repeated measures ANOVA shows no main effect of either verb type \(F_1(1, 9) = .793, p = .396; F_2(1, 6) = .511, p = .501\) or sentence type \(F_1(1, 9) = 1.684, p = .227; F_2(1, 6) = 4.024, p = .092\), but there is an interaction effect \(F_1(1, 9) = 26.422, p < .01; F_2(1, 6) = 8.832, p = .025\). The interaction effect indicates that the participants distinguish between the 2 verb types: A planned comparison revealed that in transitive sentences, obligatorily transitive verbs sounded more natural than intransitively-biased optionally transitive verbs \(t_1(9) = 3.689, p < .01; t_2(3) = 2.602, p = .080\), while in intransitive sentences, the intransitively-biased optionally transitive verbs sounded better than the obligatorily transitive verbs \(t_1(9) = -3.637, p < .01; t_2(3) = -3.306, p = .046\).

In the Korean AJT of the cL2ers-Mid group, a repeated measures ANOVA shows a main effect of sentence type \(F_1(1, 9) = 8.660, p = .016; F_2(1, 6) = 409.548, p < .001\) and an effect of interaction \(F_1(1, 9) = 12.334, p < .01; F_2(1, 6) = 102.387, p < .001\), but not of verb type \(F_1(1, 9) = .336, p = .577; F_2(1, 6) = .089, p = .775\). The main effect of sentence type shows that transitive sentences sounded more natural than intransitive sentences. The interaction effect suggests that the 2 verbs types are different in terms of their tendency to appear in a sentence type. A planned comparison revealed that in transitive sentences, the Korean equivalents of the obligatorily transitive English verbs sounded significantly more natural than the Korean equivalents of the intransitively-biased optionally transitive verbs \(t_1(9) = 4.521, p < .01; t_2(3) = 2.128, p = .123\), while in intransitive sentences, the Korean equivalents of the English intransitively-biased optionally transitive verbs sounded more natural numerically \(t_1(9) = -1.575, p = .150; t_2(3) = -1.282, p = .290\).

Figure 5.9 shows the acceptability ratings for the English AJT and Korean AJT by the highest proficiency child L2er group (henceforth, cL2ers-High).
Figure 5.9 shows that it is hard to conclude whether the cL2ers-High group knows the ungrammaticality of intransitive sentences with obligatorily transitive English verbs, based on the raw acceptability ratings: Their average acceptability rating for those sentences is around 3.5. In the Korean AJT, they tended to allow the translation equivalents of the obligatorily transitive English verbs used in intransitive sentences, where the average is over 3.5.

In the cL2ers-High English AJT, a repeated measures ANOVA shows a main effect of sentence type [$F_1(1, 8) = 5.547, p = .046$; $F_2(1, 6) = 91.177, p < .001$], a marginal main effect of verb type in the by-participant analysis only [$F_1(1, 8) = 3.579, p = .095$; $F_2(1, 6) = 1.446, p = .274$], and an interaction effect [$F_1(1, 8) = 27.630, p < .01$; $F_2(1, 6) = 31.937, p < .01$]. The (marginal) main effects of sentence type and verb type suggest that these participants were sensitive to the ungrammatical intransitive sentences with obligatorily transitive verbs. The interaction effect indicates that the participants distinguish between the 2 verb types: A planned comparison revealed that in transitive sentences, obligatorily transitive verbs sounded more natural than intransitively-biased optionally transitive verbs [$t_1(8) = 3.207, p = .012$; $t_2(3) = 1.877, p = .157$], while in intransitive sentences, the intransitively-biased optionally transitive verbs sounded better than the obligatorily transitive verbs [$t_1(8) = -4.061, p < .01$; $t_2(3) = -5.108, p = .015$].

Notes. Obl-trans = Obligorily transitive verbs used in transitive sentences;
Obl-Int = Obligorily transitive verbs used in intransitive sentences;
Intopt-trans = Intransitively-biased, optionally transitive verbs used in transitive sentences; Intopt-Int = Intransitively-biased, optionally transitive verbs used in intransitive sentences; error bars represent standard errors.
In the cL2ers-High Korean AJT, a repeated measures ANOVA shows a main effect of sentence type \( [F_1(1, 8) = 11.154, p = .010; F_2(1, 6) = 340.909, p < .001] \) and an effect of interaction \( [F_1(1, 8) = 35.039, p < .001; F_2(1, 6) = 106.909, p < .001] \), but no main effect of verb type \( [F_1(1, 8) = 2.286, p = .169; F_2(1, 6) = .519, p = .498] \). The main effect of sentence type shows that transitive sentences sounded more natural than intransitive sentences. The interaction effect suggests that the 2 verb types are different in terms of their tendency to occur in a sentence type. A planned comparison revealed that in transitive sentences, the Korean equivalents of the obligatorily transitive English verbs sounded more natural than the Korean equivalents of the intransitively-biased optionally transitive verbs \( [t_1(8) = 3.592, p < .01; t_2(3) = 2.013, p = .138] \), while in intransitive sentences, the Korean equivalents of the intransitively-biased optionally transitive verbs sounded more natural \( [t_1(8) = -4.311, p < .01; t_2(3) = -2.350, p = .100] \).

All in all, the English AJT results showed that the 6 participant groups distinguish the 2 English verb types in terms of their tendency to occur in the 2 sentence types, as seen in the significant interaction effects. In addition, all 3 adult L2 groups know that obligatorily transitive verbs cannot be used in intransitive sentences, where mean acceptability ratings are lower than 3.5. However, only the aL2ers-High group showed robust main effects of verb type and sentence type, which is native-like. Among the child L2 groups, the cL2ers-High group seems to know that obligatorily transitive verbs cannot be used in intransitive sentences: The main effects of verb type and sentence type seem to indicate this knowledge, although their acceptability rating on the ungrammatical intransitive sentences with obligatorily transitive verbs was around 3.5. For the other 2 child L2 groups, there is no evidence of knowledge of the ungrammaticality of the intransitive sentences using obligatorily transitive verbs.

The Korean AJT results showed interaction effects across the 6 groups, which indicate that the 2 Korean verb groups are distinct in terms of their tendency to occur in the 2 sentence types. Notably, all 6 groups tended to allow the Korean translation equivalents of the obligatorily transitive English verbs used in Korean intransitive sentences, where the average is around or over 3.5.
5.5.3.2 TVJT.

This section presents the results of the English TVJT and the Korean TVJT. Knowledge of the English TC was assessed with the English TVJT, in which a sentence-internal factor (verb type: obligatorily transitive verbs vs. intransitively-biased optionally transitive verb) and a sentence-external factor (context type: neutral context vs. agent-biased context) were manipulated. The Korean TVJT reveals how the Korean participants interpret the Korean translation equivalents of the English TC. Recall that the closest Korean translation equivalents allow both subject interpretation and object interpretation (see pp. 89-90).

5.5.3.2.1 TVJT data.

See Figure 5.10 for the TVJT results of the control groups.

![TVJT data graph](image)

**Figure 5.10.** Object interpretation (%) in the English/Korean TVJT: control groups.

*Notes.* O+N = Obligorotarily transitive verb + Neutral context; O+A = Obligorotarily transitive verb + Agent-biased context; I+N = Intransitively-biased, optionally transitive verb + Neutral context; I+A = Intransitively-biased, optionally transitive verb + Agent-biased context; error bars represent standard errors.

The native English-speaking adult control group (n = 10) showed 100% accuracy on all 16 test items in the English TVJT. That is, even though the agent-biased context and the intransitively-biased transitive verbs were misleading, the adult native speaker participants relied exclusively on knowledge of the English TC for the interpretation.

The native Korean-speaking adult control group (n = 10) showed a very strong subject-interpretation bias in the Korean TVJT. The proportion of the object
interpretation does not exceed 15% across the 4 conditions: obligatorily transitive verbs with neutral context (henceforth, O+N) ($M = 12.5\%$; $SD = 0.18$), obligatorily transitive verbs with agent-biased context (henceforth, O+A) ($M = 0\%$; $SD = 0$), intransitively-biased optionally transitive verbs with neutral context (henceforth, I+N) ($M = 10.0\%$; $SD = 0.24$), and intransitively-biased optionally transitive verbs with agent-biased context (henceforth, I+A) ($M = 2.5\%$; $SD = 0.08$). The results show that when processing an ambiguous surface realization of the Korean translation equivalents of the English TC, the parser strongly prefers the subject-interpretation representation over the object-interpretation representation.\(^{13}\)

In sum, the TVJT results for the control groups clearly show that adult native English speakers know the English TC and use the knowledge perfectly, while native Korean speakers have a clear subject-interpretation bias in the translation equivalents of the English TC.

Now let us look at the results from the target L2 adult and L2 child groups. To examine the performance of the 2 target groups in the English TVJT, a 2 (context type) x 2 (verb type) x 3 (proficiency group) x 2 (learner group) ANOVA was conducted (within group factors = context type and verb type; between group factors = proficiency group and learner group). This analysis yielded main effects of context type [$F_1(1, 73) = 44.871$, $p < .001$; $F_2(1, 36) = 43.474$, $p < .001$], verb type [$F_1(1, 73) = 63.064$, $p < .001$; $F_2(1, 36) = 98.026$, $p < .001$], proficiency group [$F_1(2, 73) = 11.982$, $p < .001$; $F_2(2, 36) = 78.255$, $p < .001$], and learner group [$F_1(1, 73) = 10.825$, $p < .01$; $F_2(1, 36) = 144.754$, $p < .001$], as well as the following interaction effects: context type x verb type [$F_1(1, 73) = 24.143$, $p < .001$; $F_2(1, 36) = 26.483$, $p < .001$], verb type x proficiency group [$F_1(2, 73) = 5.065$, $p < .01$; $F_2(2, 36) = 10.606$, $p < .001$], and proficiency group x learner group [$F_1(1, 73) = 5.748$, $p < .01$; $F_2(2, 36) = 61.645$, $p < .001$]. To examine the performance of the 2 target groups in the Korean TVJT, again a 2 (context type) x 2 (verb type) x 3 (proficiency group) x 2 (learner group) ANOVA was conducted (within group factors = context type and verb type; between group factors = proficiency group and learner group). This analysis yielded main effects of context type [$F_1(1, 73) = 14.451$, $p < .001$;...
$F_2(1, 36) = 29.793, p < .001$] and verb type [$F_1(1, 73) = 6.250, p = .015; F_2(1, 36) = 3.303, p = .077]$ as well as the following interaction effect: context type x verb type [$F_1(1, 73) = 5.125, p = .027; F_2(1, 36) = 8.741, p < .01$].

To investigate further specific patterns for the 2 learner groups, separate repeated-measures ANOVAs were conducted for each group. A 2 (context type) x 2 (verb type) x 3 (proficiency group) ANOVA was conducted (within group factors = context type and verb type; between group factor = proficiency group). For the English TVJT of the adult L2ers, the results showed main effects of context type [$F_1(1, 46) = 23.512, p < .001; F_2(1, 18) = 21.020, p < .001$], verb type [$F_1(1, 46) = 49.628, p < .001; F_2(1, 18) = 52.227, p < .001$], and proficiency group [$F_1(1, 46) = 20.202, p < .001; F_2(1, 18) = 79.482, p < .001$], as well as interaction effects between context type and verb type [$F_1(1, 46) = 18.410, p < .001; F_2(1, 18) = 15.894, p < .01$] and between verb type and proficiency group [$F_1(2, 46) = 6.484, p < .01; F_2(1, 18) = 9.443, p < .01$]. For the Korean TVJT of the adult L2ers, the results showed a main effect of context type [$F_1(1, 46) = 15.149, p < .001; F_2(1, 18) = 20.463, p < .001$], and a (marginal) main effect of proficiency group [$F_1(1, 46) = 2.602, p = .085; F_2(2, 18) = 13.939, p < .001$], but not of verb type [$F_1(1, 46) = 2.441, p = .125; F_2(1, 18) = 1.024, p = .325$].

For the English TVJT of the child L2ers, the results showed main effects of context type [$F_1(1, 27) = 19.992, p < .001; F_2(1, 18) = 24.000, p < .001$] and verb type [$F_1(1, 27) = 21.058, p < .001; F_2(1, 18) = 79.787, p < .001$], and an interaction effect between context type and verb type [$F_1(1, 27) = 7.807, p < .01; F_2(1, 18) = 10.667, p < .01$]. There was no main effect of proficiency group [$F_1(1, 27) = .593, p = .560; F_2(1, 18) = 2.011, p = .163$]. For the Korean TVJT of the child L2ers, the results showed main effects of context type [$F_1(1, 27) = 3.171, p = .086; F_2(1, 18) = 10.286, p < .01$] and verb type [$F_1(1, 27) = 4.546, p = .042; F_2(1, 18) = 4.261, p = .054$]. However, there was no main effect of proficiency group [$F_1(1, 27) = .221, p = .803; F_2(2, 18) = .283, p = .757$].

These ANOVA results show that both context type and verb type affected the performance of the L2 adult and L2 child groups on the English TVJT, while only context type affected their performance on the Korean TVJT. In addition, a proficiency effect appeared only in the adult L2er group. To gain a clearer picture of the results, I examine each learner group.
Figure 5.11 depicts the TVJT results from the aL2ers-Low group.

**Figure 5.11** Object interpretation (%) in the English/Korean TVJT: aL2ers-Low.

Notes. O+N = Obligatorily transitive verb + Neutral context; O+A = Obligatorily transitive verb + Agent-biased context; I+N = Intransitively-biased, optionally transitive verb + Neutral context; I+A = Intransitively-biased, optionally transitive verb + Agent-biased context; error bars represent standard errors.

In the English TVJT, the aL2ers-Low participants showed a strong preference for the incorrect subject interpretation. All conditions except the O+N condition (proportion of object interpretation = 26.9%; SD = 0.35) show a floor effect, where the proportion of object interpretation is less than 10% (3.8% (SD = 0.09), 5.8% (SD = 0.15), and 7.7% (SD = 0.16) for the O+A, I+N, and I+A conditions, respectively). A repeated measures ANOVA shows main effects of context type \( F_1(1, 12) = 5.149, p = .043; F_2(1, 6) = 8.442, p = .027 \)), a marginal main effect of verb type in the by-item analysis only \( F_1(1, 12) = 1.834, p = .201; F_2(1, 6) = 5.651, p = .055 \), and an interaction effect \( F_1(1, 12) = 8.667, p = .012; F_2(1, 6) = 11.791, p = .014 \). The main effect of context type indicates that the contextual information affected the interpretation of the English TC, leading to the erroneous subject interpretation in the agent-biased context. The robust effect of context type together with the less robust effect of verb type (marginal in only the participant analysis) suggests that the verb type information did not affect the interpretation of the English TC as strongly as the contextual information did. The significant interaction effect seems to come from a floor effect. The participants already had a subject-interpretation bias, even in the most facilitating condition (i.e., O+N) for the object interpretation. For the items with intransitively-biased optionally transitive verbs
(i.e., I+N, I+A), the participants were more inclined toward the subject interpretation. The neutral context does not promote the object interpretation, while the agent-biased context instead promotes the incorrect subject interpretation. This is to say, in the I+N and I+A conditions where the object-interpretation proportion is almost at floor (5.8% and 7.7%), the agent-biased context does not have room to play a role in leading to the subject interpretation bias. Regarding this issue, Breakwell, Smith, and Wright (2012) note that when ceiling performance or floor performance appears in one or more conditions of an experiment, the interaction effect is uninterpretable. Therés Grüter (personal communication) further points out that ANOVAs are highly unstable with data distributed in this way, as the assumption of normality is badly violated.

In the Korean TVJT, the aL2ers-Low participants showed a strong preference for the subject interpretation. The proportion of object interpretation is less than 20% across the 4 conditions: O+N ($M = 17.3\%; \text{SD} = 0.24$), O+A ($M = 11.5\%; \text{SD} = 0.30$), I+N ($M = 11.5\%; \text{SD} = 0.22$), and I+A ($M = 3.8\%; \text{SD} = 0.14$). A repeated measures ANOVA shows marginal main effects of context type in the by-participant analysis only [$F_1(1, 12) = 3.419, p = .089; F_2(1, 6) = 3.419, p = .114$] and of verb type in the by-participant analysis only [$F_1(1, 12) = 3.419, p = .089; F_2(1, 6) = 2.882, p = .140$], but no effect of interaction [$F_1(1, 12) = .037, p = .851; F_2(1, 6) = .070, p = .801$]. The effects of context and verb type are not robust, but the effects indicate that context and verb type play a role in the interpretation of the Korean equivalents of the English TC, to some extent.

Figure 5.12 depicts the TVJT results from the aL2ers-Mid group.
Figure 5.12. Object interpretation (%) in the English/Korean TVJT: aL2ers-Mid.

Notes. O+N = Obligatorily transitive verb + Neutral context; O+A = Obligatorily transitive verb + Agent-biased context; I+N = Intransitively-biased, optionally transitive verb + Neutral context; I+A = Intransitively-biased, optionally transitive verb + Agent-biased context; error bars represent standard errors.

In the English TVJT, the aL2ers-Mid participants showed a subject-interpretation bias. Only the O+N condition showed a proportion of object interpretation over 50% ($M = 54.2\%$; $SD = 0.37$). The other conditions showed less than 30% object interpretation: O+A ($M = 27.8\%$; $SD = 0.33$), I+N ($M = 13.9\%$; $SD = 0.18$), and I+A ($M = 11.1\%$; $SD = 0.21$). A repeated measures ANOVA shows main effects of context type [$F_1(1, 17) = 17.000, p < .01$; $F_2(1, 6) = 19.737, p < .01$] and verb type [$F_1(1, 17) = 19.841, p < .01$; $F_2(1, 6) = 19.841, p < .01$] as well as an effect of interaction [$F_1(1, 17) = 9.430, p < .01$; $F_2(1, 6) = 13.926, p = .010$]. The main effect of context type indicates that the contextual information affected the interpretation of the English TC. That is, the participants showed more subject interpretation in the agent-biased context. The main effect of verb type indicates that the verb type information affected the interpretation of the English TC. The interaction effect comes from a floor effect in the I+N (13.9%) and I+A (11.1%) conditions, as already discussed.

In the Korean TVJT, the aL2ers-Mid participants showed a subject-interpretation bias. The proportion of object interpretation is less than 30% across the 4 conditions: O+N ($M = 25.0\%$; $SD = 0.28$), O+A ($M = 6.9\%$; $SD = 0.14$), I+N ($M = 19.4\%$; $SD = 0.26$), and I+A ($M = 11.1\%$; $SD = 0.20$). A repeated measures ANOVA shows a main effect of context type [$F_1(1, 17) = 14.860, p < .01$; $F_2(1, 6) = 11.400, p = .015$], but not verb type [$F_1(1, 17) = .035, p = .854$; $F_2(1, 6) = .042, p = .844$], and an effect of interaction.
$F_1(1, 17) = 1.637, p = .218; F_2(1,6) = 1.547, p = .260$. The main effect of context type indicates that the contextual information affected the interpretation of the Korean translation equivalents of the English TC. That is, the participants showed more subject interpretation in the agent-biased context.

Figure 5.13 depicts the TVJT results from the aL2ers-High group.

![Figure 5.13. Object interpretation (%) in the English/Korean TVJT: aL2ers-High.](image)

Notes. O+N = Obligatorily transitive verb + Neutral context; O+A = Obligatorily transitive verb + Agent-biased context; I+N = Intransitively-biased, optionally transitive verb + Neutral context; I+A = Intransitively-biased, optionally transitive verb + Agent-biased context; error bars represent standard errors.

In the English TVJT, the aL2ers-High participants showed an object-interpretation bias in the O+N and O+A conditions (87.5% ($SD = 0.29$) and 76.4% ($SD = 0.30$), respectively), but not in the I+N and I+A conditions (40.3% ($SD = 0.33$) and 37.5% ($SD = 0.36$), respectively). A repeated measures ANOVA shows a main effect of verb type $[F_1(1, 17) = 48.767, p < .001; F_2(1, 6) = 27.457, p < .01]$ and a marginal main effect of context type in the by-participant analysis only $[F_1(1, 17) = 4.208, p = .056; F_2(1, 6) = 1.829, p = .225]$, but no effect of interaction $[F_1(1, 17) = 1.545, p = .231; F_2(1, 6) = .659, p = .448]$. The effect of verb type was robust, while the effect of context type was not. This suggests that the contextual information did not affect the interpretation of the English TC as strongly as the verb information, which is the opposite of the pattern in the aL2ers-Low group.\(^\text{14}\)

\(^{14}\) This suggests that for comprehension, low proficiency adult L2ers rely more heavily on context information than high proficiency adult L2ers do. That is, the extent to which adult L2ers rely on context information for comprehension seems to hinge on TL proficiency.
In the Korean TVJT, the aL2ers-High participants showed a subject-interpretation bias. The proportion of object interpretation was less than 40% across the 4 conditions: O+N ($M = 37.5\%$; $SD = 0.32$), O+A ($M = 23.6\%$; $SD = 0.31$), I+N ($M = 26.4\%$; $SD = 0.28$), and I+A ($M = 26.4\%$; $SD = 0.36$). However, the data show more object interpretation than the data from the native Korean-speaking control group (see Figure 5.10) and the other aL2ers groups. These participants’ performance on the Korean TVJT may have been affected by their experience with the English TVJT. A repeated measure ANOVA shows a marginal main effect of context in the by-item analysis only [$F_1(1, 17) = 2.457, p = .135; F_2(1, 6) = 5.769, p = .053$] and an interaction effect [$F_1(1, 17) = 5.120, p = .037; F_2(1, 6) = 5.769, p = .053$], but no main effect of verb type [$F_1(1, 17) = .739, p = .402; F_2(1, 6) = .239, p = .642$].

Now let us look at the child L2ers’ results. Like the adult L2ers, the child L2ers tended toward the subject interpretation. Figure 5.14 depicts the TVJT results from the cL2ers-Low group.

In the English TVJT, the cL2ers-Low participants strongly favored the subject interpretation. All conditions except the O+N condition ($M = 31.8\%$; $SD = 0.32$) showed a floor effect, where the proportion of object interpretation is less than 10% (9.1% ($SD = 0.17$), 4.5% ($SD = 0.16$), and 4.5% ($SD = 0.16$) for the O+A, I+N, and I+A

---

**Notes.** O+N = Obligatorily transitive verb + Neutral context; O+A = Obligatorily transitive verb + Agent-biased context; I+N = Intransitively-biased, optionally transitive verb + Neutral context; I+A = Intransitively-biased, optionally transitive verb + Agent-biased context; error bars represent standard errors.
conditions, respectively). A repeated measures ANOVA shows a main effect of context type \(F_1(1, 10) = 6.908, p = .033\); \(F_2(1, 6) = 3.947, p = .094\) and verb type \(F_1(1, 10) = 8.033, p = .018\); \(F_2(1, 6) = 13.364, p = .011\), but no effect of interaction \(F_1(1, 10) = 2.941, p = .117\); \(F_2(1, 6) = 3.947, p = .094\). The main effect of context indicates that the participants tended to be misled by the agent-biased contextual information. The effect of verb type suggests that the participants tended to be misled by the intransitively-biased optionally transitive verbs.

In the Korean TVJT, the cL2ers-Low participants showed a strong preference for the subject interpretation. The proportion of object interpretation was less than 20% across the 4 conditions: O+N (\(M = 15.9\%\); \(SD = 0.20\)), O+A (\(M = 6.8\%\); \(SD = 0.16\)), I+N (\(M = 6.8\%\); \(SD = 0.12\)), and I+A (\(M = 4.5\%\); \(SD = 0.10\)). The results showed a main effect of context type in the by-item analysis only \(F_1(1, 10) = 2.119, p = .176\); \(F_2(1, 6) = 25.000, p < .01\) and an effect of interaction in the by-item analysis only \(F_1(1, 10) = .672, p = .432\); \(F_2(1, 6) = 9.000, p = .024\), but no main effect of verb type \(F_1(1, 10) = 2.119, p = .176\); \(F_2(1, 6) = 1.923, p = .215\). The effect of context type is not robust, but it suggests that the children used contextual information in the interpretation of the Korean equivalents of the English TCs. The interaction effect comes from the floor effect in the I+N (4.5%) and I+A (4.5%) conditions.

Figure 5.15 depicts the TVJT results from the cL2ers-Mid group.

![English TVJT](image)

![Korean TVJT](image)

*Figure 5.15. Object interpretation (%) in the English/Korean TVJT: cL2ers-Mid.*

**Notes.** O+N = Obligatorily transitive verb + Neutral context; O+A = Obligatorily transitive verb + Agent-biased context; I+N = Intransitively-biased, optionally transitive verb + Neutral context; I+A = Intransitively-biased, optionally transitive verb + Agent-biased context; error bars represent standard errors.
In the English TVJT, the cL2ers-Mid participants showed a subject-interpretation bias. The proportion of object interpretation is as follows across the 4 conditions: O+N, $M = 40.0\%$ ($SD = 0.39$); O+A, $M = 10.0\%$ ($SD = 0.24$); I+N, $M = 10.0\%$ ($SD = 0.24$); and I+A, $M = 2.5\%$ ($SD = 0.08$). A repeated measures ANOVA shows main effects of context type [$F_1(1, 9) = 9.878, p = .012$; $F_2(1, 6) = 9.000, p = .024$], verb type [$F_1(1, 9) = 9.878, p = .012$; $F_2(1, 6) = 25.000, p < .01$], and a marginal effect of interaction in the by-participant analysis only [$F_1(1, 9) = 4.893, p = .054$; $F_2(1, 6) = 3.240, p = .122$]. The main effect of context indicates that the contextual information affected the interpretation of the English TC. Specifically, the participants showed more subject interpretations in the agent-biased context. The effect of verb type suggests that the verb information affected the interpretation of the English TC; these children were misled by the intransitively-biased optionally transitive verbs. The interaction effect comes from a floor effect.

In the Korean TVJT, the cL2ers-Mid participants strongly favored the subject interpretation. The proportion of the object interpretation is as follows across the 4 conditions: O+N, $M = 20.0\%$ ($SD = 0.31$); O+A, $M = 5.0\%$ ($SD = 0.11$); I+N, $M = 7.5\%$ ($SD = 0.17$); and I+A, $M = 12.5\%$ ($SD = 0.24$). The results showed a main effect of context in the by-item analysis only [$F_1(1, 9) = 2.250, p = .168$; $F_2(1, 6) = 6.000, p = .050$] and an interaction effect in the by-item analysis only [$F_1(1, 9) = 3.273, p = .104$; $F_2(1, 6) = 24.000, p < .01$], but no main effect of verb type ($F_1(1, 9) = .474, p = .509$; $F_2(1, 6) = .200, p = .670$). The effect of context type was not robust but it indicates that the contextual information affected the interpretation of the English TC to some extent. The interaction effect again comes from a floor effect.

Figure 5.16 depicts the TVJT results from the cL2ers-High group.
In the English TVJT, the cL2ers-High participants showed a subject-interpretation bias. The proportion of the object interpretation is as follows across the 4 conditions: O+N, $M = 47.2\%$ ($SD = 0.46$); O+A, $M = 25.0\%$ ($SD = 0.35$); I+N, $M = 11.1\%$ ($SD = 0.21$); and I+A, $M = 2.8\%$ ($SD = 0.08$). A repeated measures ANOVA shows a main effect of verb type [$F_1(1, 8) = 5.765, p = .043$; $F_2(1, 6) = 57.522, p < .001$] and a marginal main effect of context type [$F_1(1, 8) = 4.566, p = .065$; $F_2(1, 6) = 33.000, p < .01$], as well as an effect of interaction in the by-item analysis only [$F_1(1, 8) = 1.220, p = .302$; $F_2(1, 6) = 6.818, p = .040$]. The effect of verb type suggests that the verb type information affected the interpretation of the English TC. The marginal main effect of context type indicates that the children were misled by the agent-biased context in their interpretation of the English TC.

In the Korean TVJT, the cL2ers-High participants strongly favored the subject interpretation. The proportion of the object interpretation is less than 20% across the 4 conditions: O+N ($M = 19.4\%; SD = 0.33$), O+A ($M = 16.7\%; SD = 0.35$), I+N ($M = 11.1\%; SD = 0.25$), and I+A ($M = 8.3\%; SD = 0.13$). There was no main effect of verb type [$F_1(1, 8) = 2.000, p = .195$; $F_2(1, 6) = 3.769, p = .100$], context type [$F_1(1, 8) = .229, p = .645$; $F_2(1, 6) = .871, p = .387$], nor was there an interaction effect [$F_1(1, 8) = .000, p = .999$; $F_2(1, 6) = .097, p = .766$].

Figure 5.17 summarize the English TVJT results.
Figure 5.17. Object interpretation (%) in the English TVJT.

Notes. O+N = Obligatory transitive verb + Neutral context; O+A = Obligatory transitive verb + Agent-biased context; I+N = Intransitively-biased, optionally transitive verb + Neutral context; I+A = Intransitively-biased, optionally transitive verb + Agent-biased context; error bars represent standard errors.

Figure 5.17 shows that among the 6 L2er groups, significantly-above-chance performance appeared only in the O+N and O+A conditions by the aL2ers-High group alone. No group showed above chance performance in the I+N and I+A conditions. The figure also indicates that in the interpretation of the English TC, all the L2er groups were affected by verb and context information, whereas English native speakers were not influenced by such information, always going with the object interpretation.

5.5.3.2.2 Assessing competence from comprehension data.

Now, let us examine whether any L2ers know the relevant aspects of the TL grammar. As discussed in Chapter 2, I rely on the comparison-against-chance diagnostic to claim presence of knowledge. To test for the presence of knowledge in each participant group, I ran one-way ANOVAs to compare between the group accuracy rates for the English TVJT and chance scores (i.e., 50% (8/16)). The results are shown in Table 5.3.
Table 5.3
*Testing the presence of TL knowledge (with the comparison-against-chance diagnostic)*

<table>
<thead>
<tr>
<th>Group</th>
<th>Accuracy on the English TVJT</th>
<th>Statistical Information</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>aL2ers-Low</td>
<td>11.1% (SD = 0.13)</td>
<td>F(1, 24) = 111.518, p &lt; .001</td>
<td>Significantly lower than 50%: Strong subject-interpretation preference</td>
</tr>
<tr>
<td>aL2ers-Mid</td>
<td>26.8% (SD = 0.22)</td>
<td>F(1, 34) = 20.421, p &lt; .001</td>
<td>Significantly lower than 50%: Strong subject-interpretation preference</td>
</tr>
<tr>
<td>aL2ers-High</td>
<td>60.4% (SD = 0.28)</td>
<td>F(1, 34) = 2.576, p = .118</td>
<td></td>
</tr>
<tr>
<td>cL2ers-Low</td>
<td>12.5% (SD = 0.13)</td>
<td>F(1, 20) = 90.000, p &lt; .001</td>
<td>Significantly lower than 50%: Strong subject-interpretation preference</td>
</tr>
<tr>
<td>cL2ers-Mid</td>
<td>15.6% (SD = 0.21)</td>
<td>F(1, 18) = 26.053, p &lt; .001</td>
<td>Significantly lower than 50%: Strong subject-interpretation preference</td>
</tr>
<tr>
<td>cL2ers-High</td>
<td>21.5% (SD = 0.21)</td>
<td>F(1, 16) = 16.562, p &lt; .01</td>
<td>Significantly lower than 50%: Strong subject-interpretation preference</td>
</tr>
</tbody>
</table>

The results displayed in Table 5.3 indicate that none of the L2 proficiency groups has TL knowledge, because no group showed significantly-above-chance performance. Notably, all the groups except the aL2ers-High showed a significantly-below-chance proportion of object interpretations, which means that these groups had a strong subject-interpretation preference for the English TC.

Now, let us apply the comparison-against-baseline diagnostic. To apply this diagnostic, the following question must be addressed: What is the baseline condition? The answer that I propose is this: Participants’ performance on the intransitively-biased optionally transitive verbs + agent-biased context (the I+A condition), which is the most negatively biased condition, can be the baseline condition. If participants do not know the key properties of the English TC, there would be no significant difference in their performance on the I+A condition and the other conditions. That is, when the group accuracy on at least 1 of the 3 other conditions is significantly above the accuracy on the I+A condition, a possibility of knowledge can be considered.
However, a question arises: What if a significant difference in 1 of these 3 comparisons occurs in the Korean TVJT of the Korean control group ($n = 10$), which was not influenced by the English TVJT? If this occurs, a significant difference in the *same* comparison from the English TC would tell us nothing. To check this possibility, I ran paired sample $t$-tests between the I+A and the I+N conditions, between the I+A and the O+A conditions, and between the I+A and the O+N conditions with the Korean TVJT results from the Korean control group. These analyses showed a significant difference between I+A (2.5%) and O+N (12.5%) [$t(9) = 2.449$, $p = .037$], but not between I+A (2.5%) and I+N (10.0%) [$t(9) = 1.406$, $p = .193$] or between I+A (2.5%) and O+A (0%) [$t(9) = -1.000$, $p = .343$]. This means that we need to exclude the comparison between the I+A condition and the O+N condition in the English TVJT, when we apply the comparison-against-baseline diagnostic. See Table 5.4 for the results.
Table 5.4

<table>
<thead>
<tr>
<th>Group</th>
<th>I+A</th>
<th>I+N</th>
<th>O+A</th>
<th>O+N</th>
<th>Statistical information</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>aL2ers-Low</td>
<td>7.7%</td>
<td>5.8%</td>
<td>3.8%</td>
<td>26.9%</td>
<td>• I+A vs. I+N: $t(12) = .562, p = .584$</td>
<td>Possibility of TL knowledge</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.15)</td>
<td>(0.09)</td>
<td>(0.35)</td>
<td>• I+A vs. O+A: $t(12) = .805, p = .436$</td>
<td></td>
</tr>
<tr>
<td>aL2ers-Mid</td>
<td>11.1%</td>
<td>13.9%</td>
<td>27.8%</td>
<td>54.2%</td>
<td>• I+A vs. I+N: $t(17) = -.809, p = .430$</td>
<td>Possibility of TL knowledge</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.18)</td>
<td>(0.33)</td>
<td>(0.37)</td>
<td>• I+A vs. O+A: $t(17) = -2.287, p = .035$</td>
<td></td>
</tr>
<tr>
<td>aL2ers-High</td>
<td>37.5%</td>
<td>40.3%</td>
<td>76.4%</td>
<td>87.5%</td>
<td>• I+A vs. I+N: $t(17) = -.489, p = .631$</td>
<td>Possibility of TL knowledge</td>
</tr>
<tr>
<td></td>
<td>(0.36)</td>
<td>(0.33)</td>
<td>(0.30)</td>
<td>(0.29)</td>
<td>• I+A vs. O+A: $t(17) = -5.504, p &lt; .001$</td>
<td></td>
</tr>
<tr>
<td>cL2ers-Low</td>
<td>4.5%</td>
<td>4.5%</td>
<td>9.1%</td>
<td>31.8%</td>
<td>• I+A vs. I+N: $t(10) = .000, p = 1.000$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.17)</td>
<td>(0.32)</td>
<td>• I+A vs. O+A: $t(10) = .614, p = .553$</td>
<td></td>
</tr>
<tr>
<td>cL2ers-Mid</td>
<td>2.5%</td>
<td>10.0%</td>
<td>10.0%</td>
<td>40.0%</td>
<td>• I+A vs. I+N: $t(9) = -1.406, p = .193$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.24)</td>
<td>(0.24)</td>
<td>(0.39)</td>
<td>• I+A vs. O+A: $t(9) = -1.406, p = .193$</td>
<td></td>
</tr>
<tr>
<td>cL2ers-High</td>
<td>2.8%</td>
<td>11.1%</td>
<td>25.0%</td>
<td>47.2%</td>
<td>• I+A vs. I+N: $t(8) = -1.414, p = .195$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.22)</td>
<td>(0.35)</td>
<td>(0.46)</td>
<td>• I+A vs. O+A: $t(8) = -1.955, p = .086$</td>
<td></td>
</tr>
</tbody>
</table>

Note. Number in parenthesis = standard deviation

Table 5.4 shows that the aL2ers-Mid group and the aL2ers-High group had a significant difference between the I+A condition and the O+A condition. Recall that I do not claim the presence of TL knowledge with this comparison-against-baseline diagnostic. Instead, I claim that we cannot simply conclude that the aL2ers-Mid and the aL2ers-High do not know the English TC just because they did not pass the comparison-against-chance criterion. These results suggest, then, these 2 groups can be taken to possibly have TL knowledge of the English TC.

There is another way of applying the comparison-against-baseline diagnostic: setting the baseline as the participants’ performance on the Korean TVJT. If participants do not know the English TC, they would interpret the English TC by using the grammar that they already do know, i.e., the Korean grammar. If this occurs, the expected proportion of
object interpretation in the English TVJT would not be significantly different from the proportion of object interpretation in the Korean TVJT. For this, I carried out repeated-measures ANOVA on the proportion of participants’ object interpretations between the English TVJT \((k = 16)\) and the Korean TVJT \((k = 16)\) to examine whether there is a significant difference between the two. See Table 5.5 for the results.

**Table 5.5**  
*Testing the presence of TL knowledge (with the comparison-against-baseline diagnostic)*  

<table>
<thead>
<tr>
<th></th>
<th>English TVJT</th>
<th>Korean TVJT</th>
<th>Statistical information</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>aL2ers-Low</td>
<td>11.1% (0.13)</td>
<td>11.1% (0.19)</td>
<td>(F(1, 12) = .000, p = 1.000)</td>
<td></td>
</tr>
<tr>
<td>aL2ers-Mid</td>
<td>26.8% (0.22)</td>
<td>15.6% (0.18)</td>
<td>(F(1, 17) = 7.608, p = .013)</td>
<td><strong>Possibility of TL knowledge</strong></td>
</tr>
<tr>
<td>aL2ers-High</td>
<td>60.4% (0.28)</td>
<td>28.5% (0.28)</td>
<td>(F(1, 17) = 25.805, p &lt; .01)</td>
<td><strong>Possibility of TL knowledge</strong></td>
</tr>
<tr>
<td>cL2ers-Low</td>
<td>12.5% (0.13)</td>
<td>8.5% (0.10)</td>
<td>(F(1, 10) = 1.156, p = .308)</td>
<td></td>
</tr>
<tr>
<td>cL2ers-Mid</td>
<td>15.6% (0.21)</td>
<td>11.3% (0.19)</td>
<td>(F(1, 9) = 1.465, p = .257)</td>
<td></td>
</tr>
<tr>
<td>cL2ers-High</td>
<td>21.5% (0.21)</td>
<td>13.9% (0.24)</td>
<td>(F(1, 8) = 1.806, p = .216)</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Number in parenthesis = standard deviation

Only the aL2ers-Mid group and the aL2ers-High group had a significant difference in their interpretations between the English TVJT and the Korean TVJT. These results suggest that with this diagnostic, the aL2ers-Mid and aL2ers-High groups can be seen as possibly having knowledge of the English TC (see Appendix M for the raw data). The two comparison-against-baseline diagnostics show that we cannot claim either presence or absence of knowledge with these 2 groups\(^\text{15}\); on the other hand, we can conclude *absence of knowledge* with the other 4 groups (aL2ers-Low, cL2ers-Low, cL2ers-Mid, cL2ers-High).

---

\(^{15}\) As mentioned earlier, however, performance on the Korean TVJT could have been influenced by the English TVJT. That is, the object interpretation in the Korean TVJT could come from the object interpretation in the English TVJT. Thus to have a sense of how strong this learning effect may have been, I conducted a one way ANOVA between the proportion of each participant group’s object interpretations in the English TVJT and the average proportion of the control group’s object interpretation in the Korean TVJT. The results of this comparison show significantly higher object interpretation in the English TVJT than in the Korean control group’s Korean TVJT for aL2ers-High, aL2ers-Mid, and cL2ers-High groups \((p < .05)\).
and cL2ers-High), since they did not pass either the comparison-against-chance diagnostic or the comparison-against-baseline diagnostic.

Of course, group data can be misleading in that it can mask the presence or absence of knowledge in individual participants in the group. To test for the presence of TL knowledge in individual participants, individual data analysis is necessary.

First, let us apply the comparison-against-chance diagnostic. In a binomial distribution, the probability of 12 or more correct (out of 16) is .038 ($p < .05$). Thus, I claim presence of knowledge in an individual participant who had at least 12 correct responses out of the 16 items. See Figures 5.18 and 5.19 which illustrate the proportion of the object interpretation in the English TVJT by the adult participants and the child participants. In addition, the figures also show the proportions of no-response and yes-response for their object interpretations. The no-response indicates that they rejected the subject interpretation for the English TC, exhibiting knowledge of ungrammaticality.

![Figure 5.18. Number of correct object interpretations in the English TVJT by aL2ers.](image)

$\frac{16!}{(12!*4!)}*0.5^{12}*(1-0.5)^4 + \frac{16!}{(13!*3!)}*0.5^{13}*(1-0.5)^3 + \frac{16!}{(14!*2!)}*0.5^{14}*(1-0.5)^2 + \frac{16!}{(15!*1!)}*0.5^{15}*(1-0.5)^1 + \frac{16!}{(16!*0!)}*0.5^{16}*(1-0.5)^0 = 0.038$. 

---

16
The figures illustrate that 7 aL2ers (P33, P37, P44, P45, P46, P47, and P49) (but no cL2ers) showed evidence of knowledge of the English TC, with performance on the English TVJT significantly above chance (i.e., at least 12 correct out of 16). All these 7 L2 adult participants come from the high proficiency group. Notably, these participants showed knowledge of ungrammaticality, because some of their object interpretations came from a no-response. In particular, 1 aL2er, P46, showed 100% object interpretation, which is native-like.

Next, let us apply the comparison-against-baseline diagnostic. The baseline we identified was the I+A condition. However, this condition only had 4 items, which makes it hard to apply statistical analysis to the individual participant data. For this reason, I made 2 groupings: (a) by the factor of context (i.e. agent-biased vs. neutral) and (b) by the factor of verb transitivity (intransitively-biased optionally transitive verbs vs. obligatorily transitive verbs), where each grouping has 8 tokens. That is, for one analysis, the baseline condition is the 8 items with agent-biased context while the comparison condition is the 8 items with neutral context; for the other analysis, the baseline condition is the 8 items with intransitively-biased optionally transitive verbs while the comparison condition is the 8 items with obligatorily transitive verbs. We can consider a possibility of knowledge if a participant shows 7 or more correct out of the 8 items ($p < .05$), in at least 1 of these 2 groupings. See Figures 5.20, 5.21, 5.22, and 5.23.

Figure 5.19. Number of correct object interpretations in the English TVJT by cL2ers.
Figure 5.20. Number of correct object interpretations in the English TVJT by aL2ers: Grouping by context.

Figure 5.21. Number of correct object interpretations in the English TVJT by aL2ers: Grouping by verb transitivity.

Figure 5.22. Number of correct object interpretations in the English TVJT by cL2ers: Grouping by context.
The figures illustrate that 16 aL2ers\textsuperscript{18} and 2 cL2ers\textsuperscript{19} passed this comparison-against-baseline criterion, as their performance in at least 1 of the 2 groupings was 7 or more correct out of 8 ($p < .05$). None of these participants came from the lowest proficiency groups. The results suggest that we cannot claim a deficit in knowledge for the 9 aL2ers\textsuperscript{20} and 2 cL2ers\textsuperscript{21} who did not pass the comparison-against-chance criterion but did pass the comparison-against-baseline criterion.\textsuperscript{22} Figure 5.24 summarizes the results.

\textbf{Figure 5.23.} Number of correct object interpretations in the English TVJT by cL2ers: Grouping by verb transitivity.

\textsuperscript{18} These 16 aL2ers are P14, P30, P32, P33, P35, P37, P38, P39, P40, P41, P43, P44, P45, P46, P47, and P49.

\textsuperscript{19} These 2 cL2ers are P18 and P26.

\textsuperscript{20} These 9 aL2ers are P14, P30, P32, P35, P38, P39, P40, P41, and P43.

\textsuperscript{21} These 2 cL2ers are P18 and P26.

\textsuperscript{22} It is noteworthy that the criterion is passed more in the verb transitivity grouping than in the context grouping. This is because the obligatorily transitive verbs \textit{facilitate} target performance (i.e., object interpretation), whereas the neutral contexts are designed merely not to bias to the erroneous subject interpretation.
5.6 Discussion

The first research question was: Do adult English L2ers and child English L2ers (L1 = Korean) come to know that the English TC matrix subject is linked to the object gap in the embedded clause? In particular, are they able to delearn the subject interpretation that is possible (and even preferred) in Korean? The results showed that some Korean learners of English are able to come to know the English TC, delearning the subject interpretation. As discussed, 7 aL2ers showed evidence of knowledge of the English TC, passing the comparison-against-chance criterion.

However, in the L2 child groups, no child passed the comparison-against-chance criterion. Does this indicate that cL2ers cannot come to know the English TC? I do not believe so. In a pilot experiment that I conducted with child L2ers in Hawai‘i before, I found that 1 child L2er showed 100% object interpretation out of 16 test items. At the time of testing, he was 9 years old, and had lived in Hawai‘i for 3 years. Still, why did the child L2ers show poorer performance than the proficiency-matching adult L2ers? This question will be discussed in the next chapter (in particular, see Section 6.4).

The second research question was: Are L2ers of English able to apply the knowledge of the English TC effectively and systematically in comprehension? I found that verb information and contextual information significantly affected some of the L2ers’ interpretation of the English TC, while such information did not affect the English native...
speakers in their interpretation of the English TC. That is, the L2ers showed more incorrect subject interpretation when the verb (i.e., intransitively-biased optionally transitive verbs) and/or context information (i.e., agent-biased context) pointed toward the incorrect subject interpretation. This suggests that for the L2ers who showed knowledge of the English TC, their application of it is, at least, unstable, and their knowledge does not win over other misleading information in comprehension. If this is so, then how was their interpretation of the English TC vulnerable to being misled by agent-biased contexts or intransitively-biased optionally transitive verbs? This question will be discussed in more details in the next chapter.

---

23 Except 1 aL2er (P46), who showed 100% object interpretations.
CHAPTER VI
GENERAL DISCUSSION AND CONCLUSION

This chapter presents a brief summary of the main findings. It then discusses how linguistic knowledge can be obscured due to learners’ processing limitations and low automaticity or to researchers’ data treatment and interpretation. Theoretical implications of the results in this dissertation are then discussed, evaluating the claims which attribute adult L2ers’ poor performance in a TL to a deficit in knowledge. Finally, it discusses learnability issues shown in the Korean OSV study and in the English TC study.

6.1 Summary of the Main Findings

This dissertation investigated whether target knowledge can be revealed through the manipulation of non-grammatical factors so as to facilitate learners’ application of the linguistic knowledge. To this end, first, knowledge of the Korean OSV pattern was assessed with 2 types of picture-selection comprehension tasks—one manipulating a sentence-internal factor (i.e., augmenting the prosodic salience of the case morphemes in scrambled Korean OSV sentences, as in Elmo-lul Big Bird-ka anayo ‘Elmo, Big Bird hugs’), the other manipulating a sentence-external factor (i.e., a sense-semantically compatible discourse to make the first NP of a scrambled sentence topical, which reduces the processing burden)—as well as with a cognitively less demanding production task eliciting case markers. This study found that learners’ comprehension failures can be due to various causes, including perception problems, processing load, mapping problems from case morphology to Case feature, and a deficit in knowledge of (the distinguishing function of) case.

Second, knowledge of the English TC was assessed with a Truth-Value Judgment task—again, manipulating a sentence-internal factor (i.e., highlighting the object gap with the use of obligatorily transitive verbs, as in Elmo is easy to catch ei vs. concealing the object gap with the use of intransitively-biased optionally transitive verbs, as in Elmo is easy to race ei) and a sentence-external factor (e.g., for Elmo is easy to catch, a less misleading context showing that Elmo can be naturally interpreted either as a “catcher” or a “catchee” vs. a misleading context, strongly biasing Elmo to be interpreted erroneously
as a “catcher”)—along with an Acceptability-Judgment task. This study found that most learners’ interpretation of the English TC was vulnerable to being misled by intransitively-biased optionally transitive verbs and/or by agent-biased contexts.

6.2 What Obscures Linguistic Knowledge?

6.2.1 Learners’ processing limitations and low automaticity.

Proponents of the Missing Surface Inflection Hypothesis (Haznedar & Schwartz, 1997; Prévost & White, 2000) focus on the processing limitations of learners to account for erroneous language behavior. Perhaps language learners do not have sufficient processing resources to coordinate the variety of information needed for language processing. For example, it has been reported that heritage learners are slow at lexical processing in the heritage language (O’Grady, Schafer, Perla, Lee, & Wieting, 2009; Tang, 2010), that young native English-speaking children are poor at incorporating relevant discourse/pragmatic information when resolving temporary syntactic ambiguities, and thus have little or no ability to revise initial parsing commitments (Trueswell, Sekerina, Hill, & Logrip, 1999), and that L2ers generally show lower automaticity and slower speed in processing the TL than adult native speakers do (Hopp, 2009).

In addition, when competition appears between syntactic and non-syntactic systems, the non-syntactic systems can win over the syntactic system for language learners (but not child L1ers; see Trueswell et al., 1999). According to Reuland’s (2001) hierarchy hypothesis, the hierarchy among the systems responsible for language comprehension depends on the degree of automatization. Typically, syntax outcompetes strategies based on semantics and discourse because syntax is highly automatized and thus the most economical option. However, Avrutin (2006) argues that when “the narrow syntax is weakened and is no longer the most economical option, the hierarchy proposed by Reuland changes” (p. 58). Avrutin asserts that aphasics, whose syntactic system is weakened, occasionally rely on non-syntactic systems, in conditions where competition between two systems (syntactic and non-syntactic) occurs.

If the hierarchy solely depends on the degree of automatization, the following question arises: Is syntax the most economical system for processing the target language when it is the less dominant language? It seems possible that syntax is not as economical
if L1 knowledge and L2 knowledge compete in the course of syntactic derivation. In regard to L2 acquisition, Truscott and Sharwood-Smith (2004) speculate that when a syntactic representation for an L2 sentence is being built, competition for inclusion in the representation between L1 knowledge and L2 knowledge can occur. However, L1 knowledge has an “enormous advantage over its L2 counterpart” for its “extremely high resting level due to constant use over a long period” and thus “routinely wins the competition” (p. 14). In addition, processing the L2 knowledge itself can be cognitively more demanding. In Sharwood Smith’s competence/control model (1986), “some rule or principle may be acquired (in the competence sense) but suffer a long delay before full control is established” (p. 12). In a similar vein, Sorace (1985) made a distinction between L2 knowledge and the procedural knowledge that is necessary for accessing it, and claimed that “the development of procedural knowledge lags behind that of internalized knowledge” (p. 240). All this indicates that syntax may not be the most economical option for L2ers and heritage speakers who are not proficient in the target language. For those speakers, either semantic or pragmatic processing can be more automatic and more economical, which may result in poor performance in comprehension, when the obtained information from semantic or pragmatic processing competes with the information from the syntactic representation.

The comprehension of the target constructions in this dissertation involves a variety of information types and competition among them. Understanding of the non-canonical order of Korean scrambled OSV sentences must outcompete dependence on canonical word-order information, which is a basic strategy for finding syntactic relations (Bever, 1970); non-canonical sentences typically require additional syntactic operations compared to canonical sentences, and previous studies have reported erroneous performance in the processing of non-canonical sentences (Dabrowska & Street, 2005; Ferreira, 2003; Grüter & Crago, 2011). In a similar vein, the Isomorphic Mapping Hypothesis (O’Grady & Lee, 2005; O’Grady, Yamashita, & Lee, 2005) predicts difficulty in comprehending sentences in which the order of NPs is not aligned with the structure of the corresponding event. For the English TC, O’Grady (1997, p. 121) explained the difficulty of the English TC thus: “all other things being equal, gaps seem to be treated with increasing disfavor the more distant they are from the element that determines their
interpretation” and thus “a subject gap is preferred to a direct object gap since the former is closer to the element in the higher clause that determines its interpretation.” In addition, there is misleading information from verbs (i.e., intransitively-biased optionally transitive verbs) and context (i.e., agent-biased context), which competes with the information from the syntax. If syntax is not ‘strong enough’ to use the TL knowledge effectively or to suppress the involvement of the knowledge from the other language, or it does not outcompete the non-syntactic systems, poor performance may appear in spite of the presence of TL knowledge.

6.2.2 Data treatment and interpretation.

6.2.2.1. Group data.

Quantitative studies examine groups of participants and generalize the tendency of group behavior via statistical techniques. The basic assumption of quantitative research is that the participants in a group are homogenous.

However, the results of both of the studies in this dissertation research indicate a great deal of variation on the part of the learners. First, proficiency matters. For instance, had we not grouped the KHC based on their proficiency and merely looked at those participants as a single group, it would have been concluded that the KHC (as a single group; \( n = 31 \)) had no knowledge of the Korean OSV pattern, because their accuracy rate on the OSV items was not significantly above 50% in any of the 3 tasks (baseline 28.6%; prosody 48.4%; context 42.9%). That is, if the KHC had not been grouped based on their Korean proficiency, knowledge of the Korean OSV in the KHC-High group (\( n = 10 \); baseline 72.9%; prosody 87.1%; context 84.3%) would have remained hidden. Likewise, if the L2ers (\( n = 49 \)) in the English TC study had not been grouped based on their English proficiency, the good performance on some facilitating conditions in the aL2ers-High group (\( n = 18 \); O+N 87.5%; O+A 76.4%) would have been missed.

Second, even in the same proficiency group, there are individual differences. For example, in the same KHC-Mid group of the Korean OSV study, some participants passed the comparison-against-chance criterion for the picture-selection comprehension test, while others did not. Likewise, in the aL2ers-High group of the English TC study, 1 participant showed correct object interpretation 100% of the time, while some
participants showed 0% or near 0% accuracy, strongly preferring the subject interpretation. In fact, based on the comparison-against-chance diagnostic for group performance, one would have to conclude that even the highest proficiency L2er groups did not have knowledge of the English TC. However, we found that some individual learners in the group do have TL knowledge, passing the comparison-against-chance criterion for individual performance. This suggests that the picture of group performance can thus mask the picture of the individual performances in the group, obscuring presence (or absence) of the linguistic knowledge of individuals. That is, individual analysis, in addition to group analysis, is important. Knowledge is in the head of an individual, not in the head of a group (Bonnie D. Schwartz, personal communication).

6.2.2.2 Task effects.

White, Bruhn-Garavito, Kawasaki, Pater, and Prévost’s (1997) L2 anaphor resolution study found that learners’ performance varied considerably between a story-based variant and a picture-based variant of their TVJT, which suggests that differences in task demands may affect anaphor resolution in L2ers.

Task effects arose in this dissertation research as well. One cL1ers-Mid participant of the Korean study showed ceiling performance on Korean OSV in the picture-selection comprehension tasks while exhibiting floor performance in the case-marker-elicitation task. This is consistent with the observation that language development characteristically shows comprehension preceding production (e.g. Clark, 1993, p. 246). However, some KHC-Low participants exhibited the reverse pattern, that is, poorer OSV comprehension (in the picture-selection task) than production (in the case-marker-elicitation task). If the comprehension task had been the only test, the KHC-Low groups’ knowledge of case might have remained hidden. At the same time, if only the production task had been administered, the 1 cL1ers’ knowledge of case might not have been revealed.

Although the experimental manipulations intending to heighten performance were more successful than not, many of the learners still displayed patterns that were far from target-like. Moreover, the participants were not equally affected by the manipulations. For instance, some of the KHC responded well to the sentence-internal prosodic manipulation while others did not, some KHC responded well to the sentence-external
context manipulation while others did not, and some responded well to both. If only one of the facilitating tasks had been employed in the study, knowledge of the Korean scrambled OSV pattern that some KHC might have remained hidden. Likewise, some adult L2ers of the English TC study responded well to the facilitating context conditions while others did not, and some responded well to both facilitating verb transitivity conditions and context conditions, which shows that the participants were not equally affected by the manipulations.

6.2.2.3 Data interpretation.

In Chapter 2, I discussed two diagnostics to reject the null hypothesis: the comparison-against-chance diagnostic and the comparison-against-baseline diagnostic. I observed that data could be interpreted quite differently according to the diagnostic that is adopted. For example, in the Korean OSV study, under the comparison-against-chance diagnostic, KHC-High and cL1ers-High appear to have knowledge of Korean OSV. Under the comparison-against-baseline diagnostic, we can consider that KHC-Mid and cL1ers-Mid also possibly have the target knowledge. If a researcher sets the cut-off point as being ceiling or native-like performance (i.e., 100% accuracy on Korean OSV), one could argue that no learner group has the target knowledge. Likewise, in the English TC study, under the comparison-against-chance diagnostic, no adult (or child) learner group has knowledge of the English TC; however, under the comparison-against-baseline diagnostic, one can maintain that 2 adult learner groups (i.e., aL2ers-Mid, aL2ers-High) possibly have the target knowledge. Thus learners’ knowledge can be obscured by researchers’ decisions about the cut-off points and data interpretation.

6.3 Theoretical Implications

6.3.1 Representational deficit.

The theories in favor of representational deficits in L2ers are often based on group data from L2 participants. For example, the most recent proposal along this line, the Shallow Structure Hypothesis (SSH—Clahsen & Felser, 2006), posited that “the syntactic representations adult L2 learners compute for comprehension are shallower and less detailed than those of native speakers,” but that “L2 learners are able to draw on their
lexical, pragmatic, and world knowledge to build up a semantic or conceptual representation of the sentence.” (p. 32). This hypothesis was formulated based on the group performance of L2ers. One example concerns attachment preferences for ambiguous relative clause modifiers (e.g., *the servant of the actress who was on the balcony*). Adult L2ers as groups displayed no clear attachment preference, so Clahsen and Felser argued that adult L2ers do not use structure-based ambiguity resolution strategies (e.g., predicate proximity strategy, recency strategy). In addition, the Fundamental Difference Hypothesis (FDH—Bley-Vroman, 1989, 1990), which claims fundamental differences between what drives child language development and what drives adult foreign language learning, was formulated based on general poor performance by adult L2ers.

However, in this dissertation, I showed that poor performance is not necessarily due to a deficit in relevant knowledge in two linguistic domains. First, misinterpretation of Korean OSV may stem (a) from a problem related to other areas of cognitive activity (e.g., perception failure), rather than from lack of knowledge (of case or scrambling), or (b) from a problem at the interface between morphology and syntax (i.e., a mapping problem between case morphology and Case features), in line with the Missing Surface Inflection Hypothesis, typically used to explain production errors but here from the comprehension side (Schwartz, 2010). Errors in case production (e.g., dative -hanthey for accusative -lul) suggest a mapping problem between Case feature and case morphology. Second, English TCs may be misinterpreted due to experimental artifacts that are linguistic in nature (intransitively-biased, optionally transitive verbs), rather than to absence of knowledge (of constraints on tough movement). The studies also highlighted how crucial contextualization can be in revealing or concealing linguistic knowledge; context can make the direct object a characteristic topic in the Korean OSV study or create a bias toward the (incorrect) agent/subject interpretation in the English TC study.

---

1 The SSH is a proposal primarily about L2 processing; however, the SSH ultimately maintains a knowledge problem. For discussion, see Dekydtspotter, Schwartz, and Sprouse (2006).
2 For counterevidence, see Witzel, Witzel, and Nicol (2012).
3 The FDH proposes that UG and the domain-specific learning mechanisms are responsible for child language development, while established L1 knowledge and the general problem-solving system are the basis for adult L2 learning. The general problem-solving system is not well-equipped to deal with language acquisition, and hence the process (and product) of adult L2 acquisition is fundamentally different from that of child language acquisition.
In addition, I showed that group data often mask the differences among the individuals in the group, as discussed in Section 6.2.2.1. The null results found in the study on L2ers’ attachment resolution, which is used to support Clahsen and Felser’s SSH, may come from individual differences. That is, when comprehending a phrase such as \textit{the servant of the actress who was on the balcony}, some L2ers might use the predicate proximity strategy (choosing \textit{the servant}) while others use the recency strategy (choosing \textit{the actress}), which could place the overall group performance within the realm of chance.

\subsection*{6.3.2 Fundamental difference.}

The English TC study compared L2 adult and L2 child development using identical testing methods (and holding the L1 constant). This comparison made it possible to test the extent to which these two populations make use of the same language-learning procedures (Schwartz, 1987, 1992, 2004). According to Schwartz, if they have the same developmental patterns, this suggests that they rely on the same mechanisms; if they have distinct patterns, this suggests that child L2ers rely on one set of mechanisms and adult L2ers rely on another.

The results showed that the low proficiency L2 adult and L2 child groups’ performance demonstrated a strong subject-interpretation bias both in the English TVJT and in the Korean TVJT. As for advanced learners, several participants in aL2ers-High group (7 out of 18) showed evidence of possessing the knowledge, having an object-interpretation preference in the English TVJT. Indeed, 1 adult L2er showed 100% accuracy (100% object interpretation) in the English TVJT, even in the conditions with intransitively-biased optionally transitive verbs and agent-biased contexts, which were designed to lead to the incorrect subject interpretation. Unlike the aL2ers-High group, all of the cL2ers-High group participants still held the subject-interpretation preference. No child showed significantly-above-chance performance. However, as mentioned above, in a pilot experiment that I conducted with child L2ers in Hawai‘i, I found that 1 child L2er showed 100% object interpretation (out of 16 test items).

If we concede that advanced child L2ers can come to know the English TC as some adult L2ers did in the current study, these results can be viewed as reflecting a developmental pattern from an L1 transfer stage to a TL acquisition stage in both children.
and adults, which is consistent with the findings reported in Unsworth (2005). That is, the developmental patterns of children and adults would not be different, which would suggest that the two groups are not fundamentally different in terms of TL development.\footnote{According to William O’Grady, these data can be seen as showing a developmental pattern from a stage with easy processing (subject interpretation) to a stage with harder processing (object interpretation). This is an insightful account, but to test this, a psycholinguistic experiment should have been conducted. This issue can be addressed in a further study.}

According to Schwartz (1992, 2004), this shows that the two L2er groups rely on the same language acquisition mechanisms.

According to the FDH (Bley-Vroman, 1989, 1990), when some individual adult L2ers exhibit convergence on the TL grammar, the FDH would attribute such native-like performance to (a) application of the corresponding L1 knowledge, or (b) an instruction effect, or (c) analogy from other TL phenomena. However, this account does not hold for the significantly-above-chance performance of the 7 adult L2ers in the English TC study. First, the Korean language does not have \textit{tough} movement, which is the core knowledge for the English TC. In addition, the closest Korean equivalents of the English TC strongly prefer the subject interpretation over the object interpretation. Second, the good performance on the part of these adult L2ers of English is highly unlikely to come from L2 instruction because the grammar of the English TC is rarely taught in English classes in Korea. Third, the TL input can indicate that the object interpretation is possible; crucially, however, there is no input which informs the L2er that the subject interpretation is not allowed in the English TC.

\subsection*{6.4 Learnability Problems}

There are two learnability questions in this dissertation. First, how do some adult L2ers (L1 = Korean) come to know the English TC, in spite of the absence of negative evidence dictating that the subject interpretation is not allowed? Second, conversely, how do some KHC show a deficit in their knowledge of the accusative case marker -\textit{lul}, in spite of hundreds or thousands of exposures? In fact, the first question is Plato’s problem, while the second question is Orwell’s problem.

For the first question, a plausible explanation can be obtained when we assume that (a) UG provides a hypothesis space for possible grammars and adults can access UG (e.g.,
Schwartz & Sprouse, 2013), and (b) language learning is driven by parsing failure (Berwick & Weinberg, 1984; Carroll, 2001; Gibson & Wexler, 1994; Schwartz & Sprouse, 1994, 1996; White, 1987). Let us explain the acquisition of the English TC by adult L2ers, based on these two assumptions. When a learner is exposed to English TC input, the parser builds a representation following a possible grammar that UG provides. When there is a parsing failure, the parser will build another representation, following another possible grammar that UG licenses. The grammar with which parsing is successful will be placed in the learner’s competence system, and the grammar with which parsing is unsuccessful will eventually disappear in the L2ers’ competence system.

Many L2ers who participated in this dissertation research relied on a subject interpretation for the English TC. If L2ers use grammars that are licensed by UG, the grammar for the subject interpretation should also be possible in UG. This seems true, according to Anderson (2005). The subject interpretation was possible in the history of English. Consider (1a-c).

(1) a. Hutchinson was neither easy to believe it, nor frightened at the example. (ca. 1665, OED)
   b. I have been very hard to sleep too, and last night I was all but sleepless. (ca. 1858, OED)
   c. In particular, I am difficult to believe, that [etc.] (ca. 1691, OED)

   (Anderson, 2005, p. 397, (4))

Based on historical evidence of the type shown in (1), Anderson claims that a 4-year-old native English-speaking child’s production of *This one (subject points to elephant) was hard to fight the tiger* (p. 398) suggests that the child used the possible grammar that allows the subject interpretation. That is, the grammar for the subject interpretation is not a wild grammar, but a licit grammar that is licensed by UG.

Still, a question remains: How do learners move from a grammar with subject interpretation to a grammar with object interpretation? In other words, if learners have

---

5 Anderson proposes the following syntactic analysis for the subject interpretation.

(i) The soldier, was easy [(for) e₁] [PRO₁ to fight the girl/PRO₂]. (p. 404, (10))
access to UG, then why do native English-speaking children and L2ers initially have a consistent preference for the grammar leading to the subject interpretation? For this question, I think that quantity of input may be the key.

TL input is of course important for the acquisition of the English TC. It serves as a source of data for the learners to test their hypothesis among the possible grammars. Learners with high proficiency probably have an advantage over learners with low proficiency in that those with high proficiency are likely to have had more opportunities for TL exposure than the low proficiency learners—that is, they have had more chances for hypothesis testing.

When learners encounter the English TC, one popular hypothesis of the parser will be to build a representation compatible with the common English control pattern (e.g., *Elmo is eager to see*), rather than the infrequent English TC (Mair, 1990, p. 58). Given that the parser is “tuned to the frequency of particular structures in the language the hearer hears” (Carroll, 2001, p. 10), as discussed in Chapter 2, it is likely that the parser builds a representation following the grammar that has been frequently used. If learners build a representation compatible with the English control pattern, they will end up with the subject interpretation.

This input account provides a possible explanation for the asymmetry between adult L2ers and child L2ers found in this dissertation research. Adult L2ers showed better performance than proficiency-matching child L2ers. The adult L2ers are likely to have had more opportunities for TL exposure than the proficiency-matching child L2ers—that is, again, they have had more chances for hypothesis testing.

---

6 although for the Korean learners of English, as discussed, L1 influence may be one reason.
7 I believe that the quality of input, in addition to the quantity of input, also matters for successful hypothesis testing. For example, input such as *Elmo is easy to talk to* is one of the best pieces of data for hypothesis testing in regard to the English TC, because the grammar for the subject interpretation is ill-formed in that the uninterpretable \[\alpha N\] feature in the preposition *to* cannot be checked off in the syntactic derivation.
8 Mair reported that there were only 85 instances of the English TC in the Survey of English Usage corpus, which contains 875,000 words overall.
9 Kamil Ud Deen (personal communication) proposes an insightful account for children’s strong subject interpretation: inability to reanalyze the initial parsing commitment. Children might use some heuristic, like ‘assume first NP = agent’, until they encounter disconfirming data. However, they are not good at revising their initial parsing commitment and thus end up with a strong subject-interpretation bias. This is an interesting proposal to test in a further study. This account may extend as well to adult L2ers, since they also demonstrate problems with online reanalysis.
Now, let us turn to the next question: How do some KHC show a deficit in their knowledge of the accusative case -*lul*, in spite of a great deal of exposure in the input? For this account, I consider the possibility that the amount of input to which the KHC have been exposed is not sufficient for learning -*lul*, which is a form that shows “non-optimal mapping between form and meaning” (O’Grady, Lee, & Lee, 2011, p. 28). O’Grady et al. claim that when there is a one-to-one mapping between form and meaning (i.e., optimal mapping), it is easy to learn the form-meaning mapping. However, they stress that KHC encounter “a complex statistical association between case and at least three interrelated factors (grammatical relation, animacy, and definiteness)” (p. 28) when learning the accusative -*lul*. O’Grady et al. explain that the use of the accusative -*lul* denotes not only a grammatical relation but also animacy and definiteness in that it is more commonly used when the direct object is animate and/or definite than when it is inanimate and/or indefinite (Lee, 2006). In addition, as discussed in Chapter 4, -*lul* can be used as a morphological resource to carry topic (Sim, 2005). That is, learners must identify whether -*lul* is a marker for denoting the direct object, a marker for denoting animacy, a marker for denoting definiteness, a marker for denoting topic, or if it has all of these properties. What makes mapping more difficult is, as O’Grady et al. themselves point out, the fact that the accusative case -*lul* is dropped very frequently in colloquial speech. See Figure 6.1, which adapts Figure 2 in O’Grady et al. (p. 28).

![Diagram](https://via.placeholder.com/150)

Figure 6.1. Optimal mapping and non-optimal mapping.

O’Grady et al. claim that a larger amount of input is necessary in a non-optimal mapping situation than in an optimal mapping situation. That is, what seems to be a large amount of input in general can actually be insufficient, in the situation of non-optimal mapping.
But this leads to yet another question: Why did the KHC generally show a mastery of the nominative -ka, despite -ka also instantiating a non-optimal mapping? Nominative -ka can deliver information about a grammatical relation, definiteness, animacy, and topic, just like -lul. I speculate that a larger amount of -ka input (vs. -lul input) was given to the KHC, and thus the children were able to overcome the non-optimal mapping situation. Recall that Cho (1981) reported that adults are more likely to drop accusative markers than nominative markers, based on longitudinal data from 3 mother-child dyads: The mothers used nominative case on subjects a little more than half the time (56.8%); but the accusative marker was used for direct objects about 10% of the time. Therefore, I hypothesize that the KHC might have been exposed to -ka far more frequently than -lul. To test this proposal, a corpus study is required to investigate how often KHC are exposed to (nominative) -ka and (accusative) -lul.

6.5 Concluding Remarks

Competence is linguistic knowledge, while performance is a psycholinguistic process. Poor performance is not necessarily due to a knowledge deficit. Crain, Thornton, and Murasugi (2009, p. 129) explain that sentence comprehension needs “lexical, semantic, pragmatic, planning, and inferential abilities” in addition to the application of syntactic structure; the failure of any one of these may result in poor language performance, and thus researchers must be cautious in concluding that there is a deficit in knowledge based on less-than-perfect performance. The 2 studies in this dissertation also provide evidence that the poor performance that learners show is not necessarily due to a knowledge deficit. I believe that learners at least sometimes know more than they demonstrate.
Appendix A: Norming studies for prosody manipulation

I conducted two norming procedures, described here, based on which I manipulated the case markers by increasing the pitch by a factor of 1.1, the intensity by a factor of 2, and the duration by a factor of 2.

1. Analysis of the Speech Production of Six Native Korean-speaking Adults

Six native Korean adults participated in a production task. First, they were asked to produce a canonical Korean SOV sentence, (1a), and its corresponding OSV sentence, (1b), without any context. I recorded their productions.

(1) a. Chelswu-ka Yengswu-lul ttayly-ess-e.
    Chelswu-NOM Yengswu-ACC beat-PAST-SES
    ‘Chelswu beat Yengswu.’

b. Yengswu-lul Chelswu-ka ttayly-ess-e.
    Yengswu-ACC Chelswu-NOM beat-PAST-SES
    ‘Yengswu, Chelswu beat.’

Next, the participants were given 2 contexts and then asked to read the scripts for one part of the dialogue, while I read the other part of the dialogue. The contexts were designed to make participants prosodically emphasize case markers because the communication breakdown in the contexts was due to misunderstanding the case markers. They were asked to read their part in the scripts and emphasize the case markers in the target sentences as naturally as they could. In Script 1, the target sentence in which the case markers should be emphasized is SOV; in Script 2, the target sentence in which the case markers should be emphasized is OSV.
(2) Script 1. **SOV sentence where the case markers are stressed**

Researcher: Nwukwu-lul nwu-ka ttayly-ess-e?
who-ACC who-NOM beat-PAST-SES
‘Whom, who beat?’

Participant: Chelswu-ka Yengswu-lul ttayly-ess-e.
Chelswu-NOM Yengswu-ACC beat-PAST-SES
‘Chelswu beat Yengswu.’

Researcher: Chelswu-lul Yengswu-ka ttayly-ess-ta-ko?
Chelswu-ACC Yengswu-NOM beat-PAST-DECL-COMP?
‘(Did you say that) Chelswu, Yengswu beat?’

Participant: Ani, Chelswu-KA Yengswu-LUL ttayly-ess-e.
No, Chelswu-NOM Yengswu-ACC beat-PAST-SES
‘No, Chelswu beat Yengswu.’

(3) Script 2. **OSV sentence where the case markers are stressed**

Researcher: Nwu-ka nwukwu-lul ttayly-ess-e?
who-NOM who-ACC beat-PAST-SES
‘Who beat whom?’

Participant: Yengswu-lul Chelswu-ka ttayly-ess-e.
Yengswu-ACC Chelswu-NOM beat-PAST-SES
‘Yengswu, Chelswu beat.’

Researcher: Yengswu-ka Chelswu-lul ttayly-ess-ta-ko?
Yengswu-NOM Chelswu-ACC beat-PAST-DECL-COMP
‘(Did you say that) Yengswu beat Chelswu?’

Participant: Ani, Yengswu-LUL Chelswu-KA ttayly-ess-e.
no Yengswu-ACC Chelswu-NOM beat-PAST-SES
‘No, Yengswu, Chelswu beat.’

Next, the differences (in ratio) in pitch, intensity, and duration between the unstressed case markers (as in (1a) and (1b)) and the stressed case markers (as in (2) and (3)) for
each participant were obtained: pitch (range = 0 – x1.1), intensity (range = x2 – x2.5), and duration (range = x2 – x2.5).

2. Acceptability Judgment Task

An acceptability judgment task (on a scale of 5: 1 = sounds bad; 5 = sounds natural) was designed based on the results from the above-described production task, and it was then administered to 94 native Korean adults. A total of 48 sentences (20 test items and 28 fillers) was obtained from the Text-To-Speech (TTS) program for this acceptability judgment task. These 20 test sentences had 5 conditions (i.e., 4 tokens each), which all used the same 4 transitive verbs (i.e., *cohahata* ‘like’, *masita* ‘drink’, *mekta* ‘eat’, *ttaylita* ‘beat’) with 2 arguments (but different noun phrases were used for all these arguments). Condition 1 was a baseline without any manipulation. For Conditions 2 to 5, the following prosody manipulations, based on the production-data analysis, were made on the 2 case markers: (a) x1.1pitch, x2 intensity, x2 duration; (b) x1.1 pitch, x2.5 intensity, x2 duration; (c) x1.1 pitch, x2 intensity, x2.5 duration; and (d) x1.1 pitch, x2.5 intensity, x2.5 duration. In addition, 28 fillers were used: 7 had such strongly odd prosody manipulation; the purpose of these was to make sure participants would use the full acceptability scale; 21 of those 28 fillers had no prosody manipulation and were in fact all used in the picture-selection comprehension tasks, 14 as actual experimental items in the baseline task ($k = 14$; 7 SOV items and 7 OSV items) and the other 7 as half of the fillers in all 3 picture-selection comprehension tasks; these latter 7 had been randomly chosen from the total list of 14 fillers used in the picture-selection comprehension tasks. Table A-1 shows the results of the acceptability judgment task.
Table A-1. 
**Acceptability judgment task results with 94 native Korean adults**

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Mean score (on a scale of 5: 1 = sounds bad; 5 = sounds natural)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test items</td>
<td>- Condition 1. No manipulation ($k = 4$): <strong>4.52</strong></td>
</tr>
<tr>
<td>($k = 20$)</td>
<td>- Condition 2. x1.1 pitch, x2 intensity, x2 duration ($k = 4$): <strong>4.00</strong></td>
</tr>
<tr>
<td></td>
<td>- Condition 3. x1.1 pitch, x2 intensity, x2.5 duration ($k = 4$): <strong>2.91</strong></td>
</tr>
<tr>
<td></td>
<td>- Condition 4. x1.1 pitch, x2.5 intensity, x2 duration ($k = 4$): <strong>3.51</strong></td>
</tr>
<tr>
<td></td>
<td>- Condition 5. x1.1pitch, x2.5 intensity, x2.5 duration ($k = 4$): <strong>3.15</strong></td>
</tr>
<tr>
<td>Fillers</td>
<td>- Experimental sentences for the baseline picture-selection</td>
</tr>
<tr>
<td>($k = 28$)</td>
<td>comprehension task ($k = 14$): <strong>4.12</strong></td>
</tr>
<tr>
<td></td>
<td>- Some fillers for the 3 picture-selection comprehension tasks</td>
</tr>
<tr>
<td></td>
<td>($k = 7$): <strong>4.02</strong></td>
</tr>
<tr>
<td></td>
<td>- Sentences with strong prosody manipulation ($k = 7$): <strong>1.34</strong></td>
</tr>
</tbody>
</table>

Based on the results of this acceptability judgment task, I decided that Condition 2 of the prosody manipulation [x1.1 pitch, x2 intensity, x2 duration] on both case markers was the best option to guarantee both prosodic salience and naturalness.
Appendix B: Proficiency data for the Korean OSV study

To measure Korean proficiency, a picture-narration task (Song & Schwartz, 2009) was used: The children were presented with 3 sets of 4 pictures depicting a series of events, presented below:

The procedure for calculating proficiency scores is as follows: The recordings from each participant were transcribed, independently, by 2 trained native speakers of Korean. That is, each recording was transcribed twice. For the cases where the 2 transcriptions did not agree, the 2 native speakers and the researcher resolved the disagreement through discussion. Subsequently, I converted the raw transcribed production data to a list of T-units. In this process, I removed any repeated words, meaningless words, or self-corrected words. For example, say a participant produced koyangi, uh, kay, kay-ka pap-ul mekeyo ‘The cat, uh, the dog, the dog is eating rice’ for a scene where a dog is eating some food; here I would have remove the repeated word (kay ‘dog’), the meaningless word (uh), and the self-corrected word (koyangi ‘cat’) and then inputted the utterance as kay-ka pap-ul mekeyo ‘The dog is eating rice’. Next, the native speaker assistants annotated errors (i.e., any morphological, syntactic, or lexical errors) from the converted T-units, independently. For the cases on which the annotators did not agree, the annotators and I resolved the disagreement through discussion.

I then calculated the raw proficiency scores of each participant, by using the mean word length of T-units and the rate of error-free T-units. To make the 2 scores from each measure equivalent, the accuracy measure was converted into a decimal figure by multiplying by 10 (e.g., 11.1% x 10 = 1.11). Subsequently, the range for the converted

1 A T-unit is “one main clause with all subordinate clauses attached to it” (Hunt, 1965, p. 20).
accuracy measure and the range for the complexity measure were each calculated across all participants (HLC and younger CL1ers). In order to correlate the ranges of each measure, the complexity measure score was multiplied by 1.7331 because the range of the accuracy measure (at 10) was 1.7331 times larger than that of the complexity measure (at 5.77). These two numbers from the (weighted) complexity measure and the accuracy measure were added together, resulting in a single proficiency score for each participant. The z-scores were then calculated, which divided the participants into 3 proficiency groups.
## Proficiency data: HLC

<table>
<thead>
<tr>
<th></th>
<th>Total number of T-units (A)</th>
<th>Total number of words (B)</th>
<th>Total number of T-units with errors (C)</th>
<th>Complexity scores (D)</th>
<th>Accuracy scores (E)</th>
<th>Weighted score (F)</th>
<th>Z-score (G)</th>
<th>Grouping by proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>5</td>
<td>14</td>
<td>5</td>
<td>2.80</td>
<td>0.00</td>
<td>4.85</td>
<td>-3.08</td>
<td>LOW</td>
</tr>
<tr>
<td>P2</td>
<td>10</td>
<td>26</td>
<td>7</td>
<td>2.60</td>
<td>3.00</td>
<td>7.51</td>
<td>-2.36</td>
<td>LOW</td>
</tr>
<tr>
<td>P3</td>
<td>22</td>
<td>76</td>
<td>13</td>
<td>3.45</td>
<td>4.09</td>
<td>10.08</td>
<td>-1.65</td>
<td>LOW</td>
</tr>
<tr>
<td>P4</td>
<td>9</td>
<td>25</td>
<td>4</td>
<td>2.78</td>
<td>5.56</td>
<td>10.37</td>
<td>-1.57</td>
<td>LOW</td>
</tr>
<tr>
<td>P5</td>
<td>13</td>
<td>26</td>
<td>4</td>
<td>2.00</td>
<td>6.92</td>
<td>10.39</td>
<td>-1.57</td>
<td>LOW</td>
</tr>
<tr>
<td>P6</td>
<td>14</td>
<td>63</td>
<td>10</td>
<td>4.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td>12</td>
<td>50</td>
<td>7</td>
<td>4.17</td>
<td>4.17</td>
<td>11.39</td>
<td>-1.30</td>
<td>LOW</td>
</tr>
<tr>
<td>P8</td>
<td>12</td>
<td>47</td>
<td>5</td>
<td>3.92</td>
<td>5.83</td>
<td>12.62</td>
<td>-0.96</td>
<td>LOW</td>
</tr>
<tr>
<td>P9</td>
<td>12</td>
<td>26</td>
<td>0</td>
<td>2.17</td>
<td>10.00</td>
<td>13.76</td>
<td>-0.65</td>
<td>LOW</td>
</tr>
<tr>
<td>P10</td>
<td>9</td>
<td>26</td>
<td>1</td>
<td>2.89</td>
<td>8.89</td>
<td>13.90</td>
<td>-0.61</td>
<td>LOW</td>
</tr>
<tr>
<td>P11</td>
<td>16</td>
<td>72</td>
<td>5</td>
<td>4.50</td>
<td>6.88</td>
<td>14.67</td>
<td>-0.40</td>
<td>MID</td>
</tr>
<tr>
<td>P12</td>
<td>13</td>
<td>71</td>
<td>6</td>
<td>5.46</td>
<td>5.38</td>
<td>14.85</td>
<td>-0.35</td>
<td>MID</td>
</tr>
<tr>
<td>P13</td>
<td>11</td>
<td>62</td>
<td>5</td>
<td>5.64</td>
<td>5.45</td>
<td>15.22</td>
<td>-0.25</td>
<td>MID</td>
</tr>
<tr>
<td>P14</td>
<td>13</td>
<td>59</td>
<td>3</td>
<td>4.54</td>
<td>7.69</td>
<td>15.56</td>
<td>-0.16</td>
<td>MID</td>
</tr>
<tr>
<td>P15</td>
<td>13</td>
<td>57</td>
<td>1</td>
<td>4.38</td>
<td>9.23</td>
<td>16.83</td>
<td>0.19</td>
<td>MID</td>
</tr>
<tr>
<td>P16</td>
<td>12</td>
<td>59</td>
<td>2</td>
<td>4.92</td>
<td>8.33</td>
<td>16.85</td>
<td>0.19</td>
<td>MID</td>
</tr>
<tr>
<td>P17</td>
<td>13</td>
<td>65</td>
<td>2</td>
<td>5.00</td>
<td>8.46</td>
<td>17.13</td>
<td>0.27</td>
<td>MID</td>
</tr>
<tr>
<td>P18</td>
<td>13</td>
<td>66</td>
<td>2</td>
<td>5.08</td>
<td>8.46</td>
<td>17.26</td>
<td>0.30</td>
<td>MID</td>
</tr>
<tr>
<td>P19</td>
<td>11</td>
<td>59</td>
<td>2</td>
<td>5.36</td>
<td>8.18</td>
<td>17.48</td>
<td>0.36</td>
<td>MID</td>
</tr>
<tr>
<td>P20</td>
<td>10</td>
<td>50</td>
<td>1</td>
<td>5.00</td>
<td>9.00</td>
<td>17.67</td>
<td>0.41</td>
<td>MID</td>
</tr>
<tr>
<td>P21</td>
<td>14</td>
<td>74</td>
<td>2</td>
<td>5.29</td>
<td>8.57</td>
<td>17.73</td>
<td>0.43</td>
<td>MID</td>
</tr>
<tr>
<td>P22</td>
<td>11</td>
<td>68</td>
<td>3</td>
<td>6.18</td>
<td>7.27</td>
<td>17.99</td>
<td>0.50</td>
<td>HIGH</td>
</tr>
<tr>
<td>P23</td>
<td>9</td>
<td>48</td>
<td>1</td>
<td>5.33</td>
<td>8.89</td>
<td>18.13</td>
<td>0.54</td>
<td>HIGH</td>
</tr>
<tr>
<td>P24</td>
<td>12</td>
<td>57</td>
<td>0</td>
<td>4.75</td>
<td>10.00</td>
<td>18.23</td>
<td>0.57</td>
<td>HIGH</td>
</tr>
<tr>
<td>P25</td>
<td>11</td>
<td>55</td>
<td>0</td>
<td>5.00</td>
<td>10.00</td>
<td>18.67</td>
<td>0.69</td>
<td>HIGH</td>
</tr>
<tr>
<td>P26</td>
<td>11</td>
<td>62</td>
<td>1</td>
<td>5.64</td>
<td>9.09</td>
<td>18.86</td>
<td>0.74</td>
<td>HIGH</td>
</tr>
<tr>
<td>P27</td>
<td>12</td>
<td>62</td>
<td>0</td>
<td>5.17</td>
<td>10.00</td>
<td>18.95</td>
<td>0.76</td>
<td>HIGH</td>
</tr>
<tr>
<td>P28</td>
<td>9</td>
<td>51</td>
<td>0</td>
<td>5.67</td>
<td>10.00</td>
<td>19.82</td>
<td>1.00</td>
<td>HIGH</td>
</tr>
<tr>
<td>P29</td>
<td>14</td>
<td>91</td>
<td>1</td>
<td>6.50</td>
<td>9.29</td>
<td>20.55</td>
<td>1.20</td>
<td>HIGH</td>
</tr>
<tr>
<td>P30</td>
<td>13</td>
<td>94</td>
<td>0</td>
<td>7.23</td>
<td>10.00</td>
<td>22.53</td>
<td>1.74</td>
<td>HIGH</td>
</tr>
<tr>
<td>P31</td>
<td>13</td>
<td>101</td>
<td>0</td>
<td>7.77</td>
<td>10.00</td>
<td>23.46</td>
<td>1.99</td>
<td>HIGH</td>
</tr>
</tbody>
</table>

Notes. $D = B/A$; $E = (A–C)/A*10$; $F = 1.7331*D+E$; $G = (F–\text{grand mean})/\text{standard deviation} = (F–16.15)/3.6704$
### Proficiency data: Young cL1ers

<table>
<thead>
<tr>
<th></th>
<th>Total number of T-units (A)</th>
<th>Total number of words (B)</th>
<th>Total number of T-units with errors (C)</th>
<th>Complexity scores (D)</th>
<th>Accuracy scores (E)</th>
<th>Weighted score (F)</th>
<th>Z-score (G)</th>
<th>Grouping by proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>13</td>
<td>43</td>
<td>3</td>
<td>3.31</td>
<td>7.69</td>
<td>13.42</td>
<td>-0.74</td>
<td>LOW</td>
</tr>
<tr>
<td>P2</td>
<td>12</td>
<td>48</td>
<td>2</td>
<td>4.00</td>
<td>6.67</td>
<td>13.60</td>
<td>-0.69</td>
<td>LOW</td>
</tr>
<tr>
<td>P3</td>
<td>12</td>
<td>48</td>
<td>2</td>
<td>4.00</td>
<td>6.67</td>
<td>13.60</td>
<td>-0.69</td>
<td>LOW</td>
</tr>
<tr>
<td>P4</td>
<td>9</td>
<td>30</td>
<td>1</td>
<td>3.33</td>
<td>8.89</td>
<td>14.67</td>
<td>-0.40</td>
<td>MID</td>
</tr>
<tr>
<td>P5</td>
<td>13</td>
<td>48</td>
<td>2</td>
<td>3.69</td>
<td>8.46</td>
<td>14.86</td>
<td>-0.35</td>
<td>MID</td>
</tr>
<tr>
<td>P6</td>
<td>14</td>
<td>63</td>
<td>4</td>
<td>4.50</td>
<td>7.14</td>
<td>14.94</td>
<td>-0.33</td>
<td>MID</td>
</tr>
<tr>
<td>P7</td>
<td>14</td>
<td>42</td>
<td>0</td>
<td>3.00</td>
<td>10.00</td>
<td>15.20</td>
<td>-0.26</td>
<td>MID</td>
</tr>
<tr>
<td>P8</td>
<td>12</td>
<td>37</td>
<td>0</td>
<td>3.08</td>
<td>10.00</td>
<td>15.34</td>
<td>-0.22</td>
<td>MID</td>
</tr>
<tr>
<td>P9</td>
<td>12</td>
<td>37</td>
<td>1</td>
<td>3.09</td>
<td>10.00</td>
<td>15.36</td>
<td>-0.22</td>
<td>MID</td>
</tr>
<tr>
<td>P10</td>
<td>12</td>
<td>51</td>
<td>2</td>
<td>4.25</td>
<td>8.33</td>
<td>15.70</td>
<td>-0.12</td>
<td>MID</td>
</tr>
<tr>
<td>P11</td>
<td>12</td>
<td>48</td>
<td>0</td>
<td>4.00</td>
<td>10.00</td>
<td>16.93</td>
<td>0.21</td>
<td>MID</td>
</tr>
<tr>
<td>P12</td>
<td>15</td>
<td>67</td>
<td>0</td>
<td>4.47</td>
<td>10.00</td>
<td>17.74</td>
<td>0.43</td>
<td>MID</td>
</tr>
<tr>
<td>P13</td>
<td>14</td>
<td>78</td>
<td>2</td>
<td>5.57</td>
<td>8.57</td>
<td>18.23</td>
<td>0.57</td>
<td>HIGH</td>
</tr>
<tr>
<td>P14</td>
<td>13</td>
<td>71</td>
<td>1</td>
<td>5.46</td>
<td>9.23</td>
<td>18.70</td>
<td>0.69</td>
<td>HIGH</td>
</tr>
<tr>
<td>P15</td>
<td>10</td>
<td>57</td>
<td>1</td>
<td>5.70</td>
<td>9.00</td>
<td>18.88</td>
<td>0.74</td>
<td>HIGH</td>
</tr>
<tr>
<td>P16</td>
<td>13</td>
<td>68</td>
<td>0</td>
<td>5.23</td>
<td>10.00</td>
<td>19.07</td>
<td>0.79</td>
<td>HIGH</td>
</tr>
<tr>
<td>P17</td>
<td>12</td>
<td>64</td>
<td>0</td>
<td>5.33</td>
<td>10.00</td>
<td>19.24</td>
<td>0.84</td>
<td>HIGH</td>
</tr>
<tr>
<td>P18</td>
<td>11</td>
<td>56</td>
<td>1</td>
<td>5.50</td>
<td>10.00</td>
<td>19.53</td>
<td>0.92</td>
<td>HIGH</td>
</tr>
<tr>
<td>P19</td>
<td>12</td>
<td>67</td>
<td>0</td>
<td>5.58</td>
<td>10.00</td>
<td>19.68</td>
<td>0.96</td>
<td>HIGH</td>
</tr>
<tr>
<td>P20</td>
<td>8</td>
<td>48</td>
<td>0</td>
<td>6.00</td>
<td>10.00</td>
<td>20.40</td>
<td>1.16</td>
<td>HIGH</td>
</tr>
<tr>
<td>P21</td>
<td>12</td>
<td>74</td>
<td>0</td>
<td>6.17</td>
<td>10.00</td>
<td>20.69</td>
<td>1.24</td>
<td>HIGH</td>
</tr>
</tbody>
</table>

Notes. \( D = B/A; \ E = (A–C)/A \times 10; \ F = 1.7331 \times D + E; \ G = (F–\text{grand mean})/\text{standard deviation} = (F–16.15)/3.6704 \)
Appendix C: Items used for the Korean OSV study

**Stimuli for the baseline and prosody tasks**
*(including test items, fillers, and practice items)*

<table>
<thead>
<tr>
<th>Audio stimuli</th>
<th>Visual stimuli</th>
</tr>
</thead>
</table>
| This is Cookie Monster.  
This is Elmo.  
This is Big Bird.  
This is Bert. | ![Who are they?](image) |
| Na-nun Big Bird-ipnita.  
I-top Big Bird-is  
‘I am Big Bird.’ | ![Practice](image) |
| Cookie Monster-ka Big Bird-lul  
ccochakayo.  
Cookie Monster-NOM Big Bird-ACC  
chase  
‘Cookie Monster is chasing Big Bird.’ | ![Practice](image) |
Bert-ka    Elmo-lul  ttalakayo.
Bert-NOM   Elmo-ACC  follow
‘Bert is following Elmo.’

Cookie Monster-ka   Big Bird-lul  pallochayo.
Cookie Monster-NOM  Big Bird-ACC  kick
‘Cookie Monster is kicking Big Bird.’

Na-nun   Elmo-ipnita.
I-TOP      Elmo-is
‘I am Elmo.’

Bert-lul   Big Bird-ka    mileyo.
Bert-ACC   Big Bird-NOM  push
‘Bert, Big Bird is pushing.’
Big Bird-ka Elmo-lul anayo.
Big Bird-NOM Elmo-ACC hug
‘Big Bird is hugging Elmo.’

Yeki Penguin isseyo.
here Penguin is
‘Here is Penguin.’

Cookie Monster-lul Big Bird-ka massagehayyo.
Cookie Monster-ACC Big Bird-NOM massage
‘Cookie Monster, Big Bird is massaging.’

Yeki Sponge Bob isseyo.
here Sponge Bob is
‘Here is Sponge Bob.’
Na-nun Penguin-ipnita.
I-TOP Penguin-is
‘I am Penguin.’

Bert-lul Elmo-ka mancyeyo.
Bert-ACC Elmo-NOM touch
‘Bert, Elmo is touching.’

Bert-ka Cookie Monster-lul cwusanohayo.
Bert-NOM Cookie Monster-ACC inject
‘Bert is injecting Cookie Monster.’

Bert-nun yeki isseyo.
Bert-TOP here is
‘Bert is here.’
<table>
<thead>
<tr>
<th>Image</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><strong>Big Bird-ka Elmo-lul chongssoayo.</strong>&lt;br&gt;Big Bird-NOM Elmo-ACC shoot&lt;br&gt;‘Big Bird shoots Elmo.’</td>
</tr>
<tr>
<td><img src="image2.png" alt="Image" /></td>
<td><strong>Yeki Sponge Bb isseyo.</strong>&lt;br&gt;here Sponge Bb is.&lt;br&gt;‘Here is Sponge Bob.’</td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
<td><strong>Elmo-lul Bert-ka pallochayo.</strong>&lt;br&gt;Elmo-ACC Bert-NOM kick&lt;br&gt;‘Elmo, Bert is kicking.’</td>
</tr>
<tr>
<td><img src="image4.png" alt="Image" /></td>
<td><strong>Yeki Abby isseyo.</strong>&lt;br&gt;here Abby is&lt;br&gt;‘Here is Abby.’</td>
</tr>
<tr>
<td>Page 167</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td></td>
</tr>
</tbody>
</table>
| **Cookie Monster-ka Elmo-lul mileyo.**  
*Cookie Monster-NOM Elmo-ACC push*  
‘Cookie Monster is pushing Elmo.’ |

| Na-nun Sponge Bob-ipnita.  
*I-TOP Sponge Bob-is*  
‘I am Sponge Bob.’ |

| Na-nun Bert-ipnita.  
*I-TOP Bert-is*  
‘I am Bert.’ |

| **Bert-lul Cookie Monster-ka anayo.**  
*Bert-ACC Cookie Monster-NOM hug*  
‘Bert, Cookie Monster is hugging.’ |
Cookie Monster-nun yeki isseyo.
Cookie Monster-top here is
‘Cookie Monster is here.’

<table>
<thead>
<tr>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
</tbody>
</table>

Elmo-ka Bert-lul massagehayyo.
Elmo-NOM Bert-ACC massage
‘Elmo is massaging Bert.’

<table>
<thead>
<tr>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
</tbody>
</table>

Yeki Big Bird isseyo.
here Big Bird is
‘Here is Big Bird.’

<table>
<thead>
<tr>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
</tbody>
</table>

Big Bird-ka Cookie Monster-lul maneyeyo.
Big Bird-NOM Cookie Monster-ACC touch
‘Big Bird is touching Cookie Monster.’

<table>
<thead>
<tr>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>Page</td>
</tr>
<tr>
<td>------</td>
</tr>
</tbody>
</table>
| 24   | ![Image](image1.png) | Yeki Elmo isseyo.  
here Elmo is  
‘Here is Elmo.’ |
| 25   | ![Image](image2.png) | Bert-lul Cookie Monster-ka  
chongssoayo.  
Bert-ACC Cookie Monster-NOM  
shoot  
‘Bert, Cookie Monster is shooting.’ |
| 26   | ![Image](image3.png) | Yeki Mickey Mouse isseyo.  
here Mickey Mouse is  
‘Here is Mickey Mouse.’ |
| 27   | ![Image](image4.png) | Na-nun Elmo-ipnita.  
I-TOP Elmo-is  
‘I am Elmo.’ |
Big Bird-lul Elmo-ka cwusanohayo.
Big Bird-ACC Elmo-NOM inject
‘Big Bird, Elmo is injecting.’
**Stimuli for the context task**

*(Fillers and practice items are the same as those of the other two tasks)*

<table>
<thead>
<tr>
<th>Audio stimuli</th>
<th>Visual stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I am in a classroom now. There, Cookie Monster does not get along well with his friends in the classroom. (Context given in Korean)</td>
<td>![Visual Stimuli 1]</td>
</tr>
</tbody>
</table>
| I Cookie Monster-ka Big Bird-lul pallochayo.  
this Cookie Monster-NOM Big Bird-ACC kick  
‘This Cookie Monster is kicking Big Bird.’ |  |
| 3. I am in an ice-skating rink now. There, Bert is skating with his friends. | ![Visual Stimuli 3] |
| I Bert-lul Big Bird-ka mileyo.  
this Bert-ACC Big Bird-NOM push  
‘This Bert, Big Bird is pushing.’ |  |
4. I am in an airport now. There, Big Bird just landed in his plane.

I Big Bird-ka Elmo-lul anayo. this Big Bird-NOM Elmo-ACC hugs
‘This Big Bird is hugging Elmo.’

6. I am in a health-care facility now. There, Cookie Monster is in the massage room.

I Cookie Monster-lul Big Bird-ka massage-hayyo.
this Cookie Monster-ACC Big Bird-NOM massage
‘This Cookie Monster, Big Bird is massaging.’
9. I am in a playground now. There, Bert is in a freeze-it game.

I Bert-lul Elmo-ka mancyeyo.
this Bert-ACC Elmo-NOM touch
‘This Bert, Elmo is touching.’

10. I am in a hospital now. There, Bert is in a ward.

I Bert-ka Cookie Monster-lul cwusanohayo.
this Bert-NOM Cookie Monster-ACC inject
‘This Bert is injecting Cookie Monster.’
12. I am in a park now. There, Big Bird is playing with his friends.

I Big Bird-ka Elmo-lul chongssoayo.
this Big Bird-NOM Elmo-ACC shoot
‘This Big Bird is shooting Elmo.’

14. I am in a classroom now. There, Elmo does not get along well with his friends in the classroom.

I Elmo-lul Bert-ka pallochayo.
this Elmo-ACC Bert-NOM kick
‘This Elmo, Bert is kicking.’
16. I am in an ice-skating rink now. There, Cookie Monster is skating with his friends.

<table>
<thead>
<tr>
<th>I</th>
<th>Cookie Monster-ka</th>
<th>Elmo-lul</th>
<th>mileyo.</th>
</tr>
</thead>
<tbody>
<tr>
<td>this</td>
<td>Cookie Monster-NOM</td>
<td>Elmo-ACC</td>
<td>push</td>
</tr>
</tbody>
</table>

‘This Cookie Monster is pushing Elmo.’

19. I am in an airport now. There, Bert just landed in his plane.

<table>
<thead>
<tr>
<th>I</th>
<th>Bert-lul</th>
<th>Cookie Monster-ka</th>
<th>anayo.</th>
</tr>
</thead>
<tbody>
<tr>
<td>this</td>
<td>Bert-ACC</td>
<td>Cookie Monster-NOM</td>
<td>hug</td>
</tr>
</tbody>
</table>

‘This Bert, Cookie Monster is hugging.’
21 I am in a health-care facility now. There, Elmo is in the massage room.

I Elmo-ka Bert-lul massage-hayyo.
this Elmo-NOM Bert-ACC massage
‘This Elmo is massaging Bert.’

23 I am in a playground now. There, Big Bird is in a freeze-it game.

I Big Bird-ka Cookie Monster-lul maneyeyo.
this Big Bird-NOM Cookie Monster-ACC touch
‘This Big Bird is touching Cookie Monster’
25. I am in a park now. There, Bert is playing with his friends.

I Bert-lul Cookie Monster-ka chongssoayo.
this Bert-ACC Cookie Monster-NOM shoot
‘This Bert, Cookie Monster is shooting.’

28. I am in a hospital now. There, Big Bird is in a ward.

I Big Bird-lul Elmo-ka ewusanohayo.
this Big Bird-ACC Elmo-NOM inject
‘This Big Bird, Elmo is injecting.’
### Appendix D: OSV accuracy on the two manipulation tasks

#### Table D-1

*OSV accuracy on the two manipulation tasks: KHC*

<table>
<thead>
<tr>
<th>Subject</th>
<th>Baseline OSV accuracy (A)</th>
<th>Prosody OSV accuracy (B)</th>
<th>Context OSV accuracy (C)</th>
<th>Manipulation tasks OSV accuracy (B+C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P5</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>P6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>P7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>P9</td>
<td>0</td>
<td>4</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>P10</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P11</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>P12</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P13</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>P14</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>P15</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>P16</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>P17</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>P18</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>P19</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>P20</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>P21</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>P22</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>P23</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>P24</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>P25</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>P26</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>P27</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>P28</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>P29</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>P30</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>P31</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>13</td>
</tr>
</tbody>
</table>
Table D-2

OSV accuracy on the two manipulation tasks: cL1ers

<table>
<thead>
<tr>
<th></th>
<th>Baseline OSV accuracy (A)</th>
<th>Prosody OSV accuracy (B)</th>
<th>Context OSV accuracy (C)</th>
<th>Manipulation tasks OSV accuracy (B+C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>P2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>P3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>P4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P5</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>P6</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>P7</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>P8</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>P9</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>P10</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>P11</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>P12</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>P13</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>P14</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>P15</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>P16</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>P17</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>P18</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>P19</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>P20</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>P21</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>14</td>
</tr>
</tbody>
</table>
### Appendix E: OSV accuracy vs. SOV inaccuracy

#### Table E-1

*OSV accuracy vs. OSV inaccuracy: HLC*

<table>
<thead>
<tr>
<th></th>
<th>Baseline task</th>
<th>Prosody task</th>
<th>Context task</th>
<th>Manipulation tasks (prosody + context)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OSV accuracy (A)</td>
<td>SOV inaccuracy (B)</td>
<td>OSV accuracy (C)</td>
<td>SOV inaccuracy (D)</td>
</tr>
<tr>
<td>P1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P5</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>P6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>P9</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>P10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P11</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>P12</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P13</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>P14</td>
<td>5</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>P15</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>P16</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>P17</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>P18</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>P19</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>P20</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>P21</td>
<td>2</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>P22</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>P23</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>P24</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>P25</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>P26</td>
<td>5</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>P27</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>P28</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>P29</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>P30</td>
<td>6</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>P31</td>
<td>7</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

180
<table>
<thead>
<tr>
<th></th>
<th>Baseline task</th>
<th>Prosody task</th>
<th>Context task</th>
<th>Manipulation tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OSV accuracy</td>
<td>SOV inaccuracy</td>
<td>OSV accuracy</td>
<td>OSV inaccuracy</td>
</tr>
<tr>
<td></td>
<td>(A)</td>
<td>(B)</td>
<td>(C)</td>
<td>(D)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>P2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>P3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>P6</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>P7</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>P8</td>
<td>7</td>
<td>2</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>P9</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P10</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>P11</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>P12</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>P13</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>P14</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>P15</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>P16</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>P17</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>P18</td>
<td>7</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>P19</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>P20</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>P21</td>
<td>6</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>
Appendix F: Stimuli for the Case-marker-elicitation task

**Stimuli for the case-marker-elicitation task**
*(including test items and practice items)*

<table>
<thead>
<tr>
<th>Audio stimuli</th>
<th>Visual stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>This is Cookie Monster. This is Elmo. This is Big Bird.</td>
<td><img src="image" alt="Characters" /></td>
</tr>
<tr>
<td>Describe this picture in Korean.</td>
<td><img src="image" alt="Image 1" /></td>
</tr>
<tr>
<td>Describe this picture in Korean.</td>
<td><img src="image" alt="Image 2" /></td>
</tr>
<tr>
<td>Describe this picture in Korean.</td>
<td><img src="image" alt="Image 3" /></td>
</tr>
</tbody>
</table>
Pakwuni an-ey koyangi twu mali iss-eyo?
basket in-at cat two CL exist-SES
‘Are there two cats in the basket?’

Cookie Monster-nun Elmo anayo.
Cookie Monster-TOP Elmo hug
‘Cookie Monster hugs Elmo.’

Big Bird-nun Elmo mileyo.
Big Bird-TOP Elmo push
‘Big Bird pushes Elmo.’

Elmo-nun Cookie Monster ttalylieyo.
Elmo-TOP Cookie Monster beat
‘Elmo beats Cookie Monster.’

Big Bird-nun Cookie Monster anayo.
Big Bird-TOP Cookie Monster hug
‘Big Bird hug Cookie Monster.’
| Cookie Monster-nun Big Bird mileyo.  
| Cookie Monster-TOP Big Bird push  
| ‘Cookie Monster pushes Big Bird.’ |
|---|---|---|
| Big Bird-nun Elmo ttalylieyo.  
| Big Bird-TOP Elmo beat  
| ‘Big Bird beats Elmo.’ |
Appendix G: Analysis of the closest Korean equivalent to the English TC

Consider (1) and (2).

(1) pro kyoswunim-i [manna-ki-ka elyep-ta.]
   professor-NOM meet-NML-NOM hard-DECL
   ‘The professor is difficult to meet.’

(2) Kyoswunim-i [pro manna-ki-ka] elyep-ta.
   professor-NOM meet-NML-NOM hard-DECL
   ‘It is difficult for the professor to meet someone.’

While (1) and (2) have the same surface realization, they have different underlying syntactic representations, which yield different interpretations. In (1), *kyoswunim* ‘professor’ is the “meetee,” while in (2), it is the “meeter.” Now, consider (3) and (4).

(3) Haksaying-hanthey kyoswunim-i [manna-ki-ka elyep-ta.]
   student-DAT professor-NOM meet-NML-NOM hard-DECL
   ‘The professor is difficult for the student to meet.’

   professor-NOM student-ACC meet-NML-NOM hard-DECL
   ‘It is difficult for the professor to meet the student.’

In (3), the subject of the entire sentence is *haksaying* ‘student’, while in (4), the subject of the entire sentence is *kyoswunim* ‘professor’. This can be tested with honorific agreement, which is a reliable subjecthood test in Korean (Hong, 1994). Consider (5) and (6).

(5) *Haksaying-hanthey kyoswunim-i [manna-ki-ka elyewu-si-ta.]
   student-DAT professor-NOM meet-NML-NOM hard-HON-DECL
   ‘The professor is difficult for the student to meet.’
Sentence (5) sounds unnatural while sentence (6) sounds natural. In a sentence with two arguments *haksaying* ‘student’ and *kyoswunim* ‘professor’, *haksaying* ‘student’ cannot be honorific; the honorific must go with *kyoswunim* ‘professor’. In (5), *haksaying* ‘student’ is the subject of the entire sentence and therefore incompatible with the honorific marker -si- on the matrix predicate. In contrast, the sentence in (6) sounds natural, because *kyoswunim* ‘professor’ is the subject of the entire sentence, and therefore compatible with -si- on the matrix predicate.
Appendix H: List of English constructions taught in secondary schools in Korea

The following is a list of the English constructions that are taught in secondary schools in Korea. Note that the English TC is not included in this list.

1. He takes a walk everyday. We went on a picnic yesterday. She is going (to go) abroad next year. The next test will probably be a little more difficult.
2. He is sleeping right now. I was studying when they came. She will be coming soon.
3. They have moved to a new house. He had already left when we arrived. I will have finished my homework by the time you get here.
4. I am not tired. It isn’t very cold. I don’t work on Sundays. You can’t turn right here. We didn’t enjoy the holiday very much. Mary hasn’t told him what to do. Tom will not be at the meeting tomorrow.
5. Are you ready? Is it raining? Do you like oranges? Don’t you like apples? Can you write a letter in English? Have you finished your homework yet? When will you come? Where can we get the bus? Why did he leave early? How did you come to school? Who can answer that question? Whose dolls are these? Which class are you in? Which ice cream would you like, chocolate or vanilla?
6. May I borrow your book? He may be sick. There might be some complaints. She must have been sick. I should study tonight. They must do well on the test. I have to go to the library. You don’t have to go to school tomorrow. Will you please help me with my English? He would take a walk every morning. She can play the violin. It can’t be true. Could you show me the way to the nearest post office? I will be able to help you get to the party tonight. Would you please pass me the salt? Shall we go on a picnic this weekend? You had better not say anything about this. They’ll fight to the end rather than give up. You needn’t worry about the test. You should do as he says. He suggested that they (should) play baseball.
7. To see is to believe. He wanted to go home. I hope to see you again soon. I have a book to read. He came to see me. You may go if you want to. He told me not to do it. The problem is where to get money (from). He seemed to have been ill (for some time). I heard the children sing. I expect John to pass the exam. I asked her to help
me. It was easy to find his house. To be frank (with you), I don’t like him. The book is simple enough to understand. It moves too quickly for us to see (it).

8. I made him carry the box. They had me repeat the message. You shouldn’t let him go there again. Bill had the car cleaned. I had my picture taken.

9. Playing baseball is fun. We enjoy swimming in the pool. I’m interested in watching horror movies. I saw him lying on the beach. They felt that Mary’s coming in late was bad. Did you go fishing last weekend? I don’t feel like sleeping now. I cannot help doing things like that. It’s no use trying to persuade me.

10. The novel was written by Mark Twain. The building was built in 1880. A prize was given to Jane. Susan will be invited to the party this evening. I was made to clean the room. Let it be done at once.

11. I have two pens. She bought some furniture. I don’t have any money. My mother bought many / lots of / a lot of apples. We didn’t buy much food. I’d like to ask you a few questions. I need a little more time to think (about it). Every/Each student is listening to music. The news is interesting. The Chinese drink tea. Most of the students in the class study hard. Mathematics is my favorite subject.

12. I don’t like the black coat, but I like the brown one. I have three books. One is mine. The others are yours. The biscuit was delicious. I’m going to have another one. These cups are dirty. Could I have some clean ones? We write to each other every week.

13. There is a book on the desk. There are so many children in the park.

14. It’s cold outside. It’s Wednesday. It’s half past four. It’s ten miles from here to my school.

15. You are a better singer than I am. They’ve got more money than they need. A car is much more expensive than a motor-bike. I prefer tea to coffee. The more she thought about it, the less she liked it.

16. Excuse me, where is the nearest bank? Money is important, but it is not the most important thing in life. What is the longest river in the world?

17. I don’t know as many people as you (do). I arrived at the same time as Tom (did). She’s old, but she’s not as old as he (is).
18. **What size** is this shirt? **What kind** of job do you want? **Which color** do you prefer? **What time** is it? **How old** is she? **How big** is the house? **How heavy** is your computer? **How much** does it cost? **How far** did you drive today?

19. Is there **enough hot water** for me to take a bath? Have you got **enough food** in the house to last the next few days? There is **hardly enough cake**.

20. Leave early so that you won’t miss the bus. We moved to London so that we could visit our friends more often. I hurried so that I wouldn’t be late. It not only rained yesterday, but it also snowed. You can have either tea or coffee. David both loves Joan and wants to marry her. The film was neither well made nor well acted. He came not to complain, but to help us. He went to bed because he was sleepy. When we arrived, she was talking on the phone. Since he left this morning, I haven’t seen him. We stayed home, since the weather was cold. The weather was so nice that we went to the zoo. We’re going to play baseball tomorrow unless it rains. Although / Though it was cold, I went swimming.

21. John plays the guitar, and his sister plays the piano. They are my neighbors, but I don’t know them well. I may stop by and see you tomorrow or (I) may just phone (you) late in the day.

22. Ted didn’t study at all. Therefore, he failed the test again. I’ll gave the book to Tom. However, he didn’t like it. I like spending my holidays in the mountains. On the other hand, my sister prefers the seaside.

23. The girl who is playing the piano is called Ann. Mrs. Lee, who teaches English, has two children. The pen which is on the desk is mine. This is the book that I bought yesterday. I met the girl whose father is a musician. She is the girl whom I told you about. The town in which I was born is very small. The house where she lives is very nice indeed.

24. **What she said** was interesting. I don’t know where he lives. Please tell me what happened. I wonder whose bicycle that is. I don’t know whether he will come (or not). Tell me how to make pizza. I think that he is a good singer.

25. She said, “I will help you.” She said that she would help me. He said to me, “I have been tired a lot lately.” He told me that he had been tired a lot lately. He said to me, “Please come at once.” He asked me to come at once. He said, “Do
you need a pen?” He asked if I needed a pen. Mary asked “Where do you live?” Mary asked where I lived.

26. Walking along the street, I met an old friend. Their father having left the room, the children began to watch TV. Being tired, he went to bed. Having seen that movie before, I didn’t want to see it again. Frankly speaking, I failed the test.

27. If I have enough time, I read a book every week. If the weather is fine tomorrow, we will go on a picnic. If I were a bird, I could fly home. If you had studied harder, you would have passed the exam. Had I had enough money, I would have bought a tape-recorder. It’s time you were in bed. I wish I spoke English well. He acts as if he knew you. I’d rather we had dinner now. Without / But for your advice, I would have failed.

28. Down came the rain. Here comes the bus. Not a word did he say. She was angry and so was I. I enjoyed the play and so did my friends. John didn’t see the accident and neither did Mary.

29. Won’t you try again? - Yes, I will try again. You must speak to the teacher. - I have spoken to him. I do hope you will succeed. It was John who/that told me the truth. It was by train that we traveled to London.

30. My brother was wearing a raincoat, and (he) didn’t get wet. One student has written a poem, and the other (has written) a short story. John understands the problem better than Mary (does). I believe (that) you are mistaken. Though (he had been) defeated, he remained a popular leader. You can borrow my pen, if you want to (do so).

31. Mr. Long, a neighbor of yours, will be visiting us this evening. The goal you wrote about in your letter, to become an astronaut, requires a great amount of time and energy. The fact that Mary was late didn’t surprise me.

32. At the station you will see a lady carrying a large umbrella. Any coins found on this site must be handed to the police. Something strange happened last night.

33. This / That is your book, isn’t it? Ann is here, isn’t she? You like oranges, don’t you? Bob isn’t here, is he? They haven’t left, have they?
34. As soon as it stops raining, we will leave. No sooner had I arrived at the airport than I realized I had forgotten my briefcase. As he was not interested in classical music, he decided not to go to the concert. She wept bitterly as she told her story.

35. It is cruel to tease animals. It is difficult for me to speak French. She found it difficult to learn Chinese.

Appendix I: Proficiency data for the English TC study

The procedure for calculating proficiency scores is as follows: The recording from each participant was transcribed twice, independently by 2 trained native speakers of English. For the cases where there was disagreement in the 2 transcriptions, the 2 native English speakers resolved the disagreement through discussion. I subsequently converted the raw transcribed production data to a list of T-units. In this process, I removed any repeated words, meaningless words, or self-corrected words. For example, say a participant produced The cat, uh, the dog, the dog is eating meat for a scene where a dog is eating some food; here I would have removed the repeated word (dog), the meaningless word (uh), and the self-corrected word (cat) and then inputted the utterance as The dog is eating meat. Next, the native speaker assistants annotated errors (i.e., any morphological, syntactic, or lexical errors) from the converted T-units, independently. For the cases on which the annotators did not agree, the annotators resolved the disagreement through discussion.

I then calculated the raw proficiency scores of each participant, by using the mean word length of T-units and the rate of error-free T-units. To make the 2 scores from each measure equivalent, the accuracy measure was converted into a decimal figure by multiplying by 10 (e.g., 11.1% x 10 = 1.11). Subsequently, the range for the converted accuracy measure and the range for the complexity measure were each calculated across all participants (adult L2ers and child L2ers). In order to correlate the ranges of each measure, the complexity measure score was multiplied by 1.2023 because the range of the accuracy measure (at 9.93) was 1.2023 times larger than that of the complexity measure (at 7.76). These 2 numbers from the (weighted) complexity measure and the accuracy measure were added together, resulting in a single proficiency score for each participant. The z-scores were then calculated, which divided the participants into 3 proficiency groups. For the materials used for the picture-narration task (Song & Schwartz, 2009), see Appendix B.

---

2 A T-unit is “one main clause with all subordinate clauses attached to it” (Hunt, 1965, p. 20).
## Proficiency data: Adult L2ers

<table>
<thead>
<tr>
<th></th>
<th>Total number of T-units</th>
<th>Total number of words</th>
<th>Total number of T-units with errors</th>
<th>Complexity scores (D)</th>
<th>Accuracy scores (E)</th>
<th>Weighted score (F)</th>
<th>Z-score (G)</th>
<th>Grouping by proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>12</td>
<td>70</td>
<td>12</td>
<td>5.83</td>
<td>0.00</td>
<td>7.01</td>
<td>-2.42</td>
<td>LOW</td>
</tr>
<tr>
<td>P2</td>
<td>18</td>
<td>99</td>
<td>13</td>
<td>5.50</td>
<td>2.78</td>
<td>9.39</td>
<td>-1.60</td>
<td>LOW</td>
</tr>
<tr>
<td>P3</td>
<td>11</td>
<td>89</td>
<td>11</td>
<td>8.09</td>
<td>0.00</td>
<td>9.73</td>
<td>-1.49</td>
<td>LOW</td>
</tr>
<tr>
<td>P4</td>
<td>13</td>
<td>91</td>
<td>11</td>
<td>7.00</td>
<td>1.54</td>
<td>9.95</td>
<td>-1.41</td>
<td>LOW</td>
</tr>
<tr>
<td>P5</td>
<td>23</td>
<td>176</td>
<td>21</td>
<td>7.65</td>
<td>0.87</td>
<td>10.07</td>
<td>-1.37</td>
<td>LOW</td>
</tr>
<tr>
<td>P6</td>
<td>14</td>
<td>112</td>
<td>13</td>
<td>8.00</td>
<td>0.71</td>
<td>10.33</td>
<td>-1.28</td>
<td>LOW</td>
</tr>
<tr>
<td>P7</td>
<td>27</td>
<td>183</td>
<td>21</td>
<td>6.78</td>
<td>2.22</td>
<td>10.37</td>
<td>-1.27</td>
<td>LOW</td>
</tr>
<tr>
<td>P8</td>
<td>20</td>
<td>119</td>
<td>13</td>
<td>5.95</td>
<td>3.50</td>
<td>10.65</td>
<td>-1.17</td>
<td>LOW</td>
</tr>
<tr>
<td>P9</td>
<td>13</td>
<td>74</td>
<td>8</td>
<td>5.69</td>
<td>3.85</td>
<td>10.69</td>
<td>-1.16</td>
<td>LOW</td>
</tr>
<tr>
<td>P10</td>
<td>14</td>
<td>133</td>
<td>14</td>
<td>9.50</td>
<td>0.00</td>
<td>11.42</td>
<td>-0.90</td>
<td>LOW</td>
</tr>
<tr>
<td>P11</td>
<td>27</td>
<td>212</td>
<td>19</td>
<td>7.85</td>
<td>2.96</td>
<td>12.40</td>
<td>-0.56</td>
<td>LOW</td>
</tr>
<tr>
<td>P12</td>
<td>10</td>
<td>45</td>
<td>3</td>
<td>4.50</td>
<td>7.00</td>
<td>12.41</td>
<td>-0.56</td>
<td>LOW</td>
</tr>
<tr>
<td>P13</td>
<td>18</td>
<td>128</td>
<td>11</td>
<td>7.11</td>
<td>3.89</td>
<td>12.44</td>
<td>-0.55</td>
<td>LOW</td>
</tr>
<tr>
<td>P14</td>
<td>12</td>
<td>102</td>
<td>9</td>
<td>8.50</td>
<td>2.50</td>
<td>12.72</td>
<td>-0.46</td>
<td>MID</td>
</tr>
<tr>
<td>P15</td>
<td>18</td>
<td>151</td>
<td>13</td>
<td>8.39</td>
<td>2.78</td>
<td>12.86</td>
<td>-0.41</td>
<td>MID</td>
</tr>
<tr>
<td>P16</td>
<td>19</td>
<td>122</td>
<td>9</td>
<td>6.42</td>
<td>5.26</td>
<td>12.98</td>
<td>-0.36</td>
<td>MID</td>
</tr>
<tr>
<td>P17</td>
<td>11</td>
<td>78</td>
<td>6</td>
<td>7.09</td>
<td>4.55</td>
<td>13.07</td>
<td>-0.33</td>
<td>MID</td>
</tr>
<tr>
<td>P18</td>
<td>29</td>
<td>267</td>
<td>23</td>
<td>9.21</td>
<td>2.07</td>
<td>13.14</td>
<td>-0.31</td>
<td>MID</td>
</tr>
<tr>
<td>P19</td>
<td>11</td>
<td>62</td>
<td>4</td>
<td>5.64</td>
<td>6.36</td>
<td>13.14</td>
<td>-0.31</td>
<td>MID</td>
</tr>
<tr>
<td>P20</td>
<td>11</td>
<td>54</td>
<td>3</td>
<td>4.91</td>
<td>7.27</td>
<td>13.17</td>
<td>-0.30</td>
<td>MID</td>
</tr>
<tr>
<td>P21</td>
<td>12</td>
<td>74</td>
<td>5</td>
<td>6.17</td>
<td>5.83</td>
<td>13.25</td>
<td>-0.27</td>
<td>MID</td>
</tr>
<tr>
<td>P22</td>
<td>13</td>
<td>80</td>
<td>5</td>
<td>6.15</td>
<td>6.15</td>
<td>13.55</td>
<td>-0.17</td>
<td>MID</td>
</tr>
<tr>
<td>P23</td>
<td>12</td>
<td>96</td>
<td>7</td>
<td>8.00</td>
<td>4.17</td>
<td>13.79</td>
<td>-0.09</td>
<td>MID</td>
</tr>
<tr>
<td>P24</td>
<td>26</td>
<td>261</td>
<td>21</td>
<td>10.04</td>
<td>1.92</td>
<td>13.99</td>
<td>-0.02</td>
<td>MID</td>
</tr>
<tr>
<td>P25</td>
<td>12</td>
<td>81</td>
<td>5</td>
<td>6.75</td>
<td>5.83</td>
<td>13.95</td>
<td>-0.03</td>
<td>MID</td>
</tr>
<tr>
<td>P26</td>
<td>12</td>
<td>82</td>
<td>5</td>
<td>6.83</td>
<td>5.83</td>
<td>14.05</td>
<td>0.00</td>
<td>MID</td>
</tr>
<tr>
<td>P27</td>
<td>12</td>
<td>72</td>
<td>3</td>
<td>6.00</td>
<td>7.50</td>
<td>14.71</td>
<td>0.23</td>
<td>MID</td>
</tr>
<tr>
<td>P28</td>
<td>12</td>
<td>83</td>
<td>4</td>
<td>6.92</td>
<td>6.67</td>
<td>14.98</td>
<td>0.33</td>
<td>MID</td>
</tr>
<tr>
<td>P29</td>
<td>12</td>
<td>110</td>
<td>7</td>
<td>9.17</td>
<td>4.17</td>
<td>15.19</td>
<td>0.40</td>
<td>MID</td>
</tr>
<tr>
<td>P30</td>
<td>21</td>
<td>143</td>
<td>6</td>
<td>6.81</td>
<td>7.14</td>
<td>15.33</td>
<td>0.44</td>
<td>MID</td>
</tr>
<tr>
<td>P31</td>
<td>13</td>
<td>101</td>
<td>5</td>
<td>7.77</td>
<td>6.15</td>
<td>15.49</td>
<td>0.49</td>
<td>MID</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>P32</td>
<td>16</td>
<td>140</td>
<td>8</td>
<td>8.75</td>
<td>5.00</td>
<td>15.52</td>
<td>0.51</td>
<td>HIGH</td>
</tr>
<tr>
<td>P33</td>
<td>42</td>
<td>379</td>
<td>22</td>
<td>9.02</td>
<td>4.76</td>
<td>15.61</td>
<td>0.54</td>
<td>HIGH</td>
</tr>
<tr>
<td>P34</td>
<td>10</td>
<td>73</td>
<td>3</td>
<td>7.30</td>
<td>7.00</td>
<td>15.78</td>
<td>0.60</td>
<td>HIGH</td>
</tr>
<tr>
<td>P35</td>
<td>14</td>
<td>104</td>
<td>4</td>
<td>7.43</td>
<td>7.14</td>
<td>16.07</td>
<td>0.70</td>
<td>HIGH</td>
</tr>
<tr>
<td>P36</td>
<td>16</td>
<td>100</td>
<td>2</td>
<td>6.25</td>
<td>8.75</td>
<td>16.26</td>
<td>0.77</td>
<td>HIGH</td>
</tr>
<tr>
<td>P37</td>
<td>15</td>
<td>162</td>
<td>10</td>
<td>10.80</td>
<td>3.33</td>
<td>16.32</td>
<td>0.79</td>
<td>HIGH</td>
</tr>
<tr>
<td>P38</td>
<td>12</td>
<td>105</td>
<td>5</td>
<td>8.75</td>
<td>5.83</td>
<td>16.35</td>
<td>0.80</td>
<td>HIGH</td>
</tr>
<tr>
<td>P39</td>
<td>15</td>
<td>156</td>
<td>9</td>
<td>10.40</td>
<td>4.00</td>
<td>16.50</td>
<td>0.85</td>
<td>HIGH</td>
</tr>
<tr>
<td>P40</td>
<td>10</td>
<td>89</td>
<td>4</td>
<td>8.90</td>
<td>6.00</td>
<td>16.70</td>
<td>0.92</td>
<td>HIGH</td>
</tr>
<tr>
<td>P41</td>
<td>16</td>
<td>156</td>
<td>8</td>
<td>9.75</td>
<td>5.00</td>
<td>16.72</td>
<td>0.93</td>
<td>HIGH</td>
</tr>
<tr>
<td>P42</td>
<td>12</td>
<td>94</td>
<td>3</td>
<td>7.83</td>
<td>7.50</td>
<td>16.92</td>
<td>0.99</td>
<td>HIGH</td>
</tr>
<tr>
<td>P43</td>
<td>21</td>
<td>206</td>
<td>10</td>
<td>9.81</td>
<td>5.24</td>
<td>17.03</td>
<td>1.03</td>
<td>HIGH</td>
</tr>
<tr>
<td>P44</td>
<td>20</td>
<td>185</td>
<td>6</td>
<td>9.25</td>
<td>7.00</td>
<td>18.12</td>
<td>1.41</td>
<td>HIGH</td>
</tr>
<tr>
<td>P45</td>
<td>19</td>
<td>233</td>
<td>11</td>
<td>12.26</td>
<td>4.21</td>
<td>18.95</td>
<td>1.70</td>
<td>HIGH</td>
</tr>
<tr>
<td>P46</td>
<td>24</td>
<td>225</td>
<td>4</td>
<td>9.38</td>
<td>8.33</td>
<td>19.60</td>
<td>1.92</td>
<td>HIGH</td>
</tr>
<tr>
<td>P47</td>
<td>24</td>
<td>252</td>
<td>7</td>
<td>10.50</td>
<td>7.08</td>
<td>19.71</td>
<td>1.96</td>
<td>HIGH</td>
</tr>
<tr>
<td>P48</td>
<td>38</td>
<td>427</td>
<td>13</td>
<td>11.24</td>
<td>6.58</td>
<td>20.09</td>
<td>2.09</td>
<td>HIGH</td>
</tr>
<tr>
<td>P49</td>
<td>34</td>
<td>377</td>
<td>9</td>
<td>11.09</td>
<td>7.35</td>
<td>20.68</td>
<td>2.29</td>
<td>HIGH</td>
</tr>
</tbody>
</table>

*Note.* D = B/A; E = (A–C)/A*10; F = 1.2023*D+E; G = (F–grand mean)/standard deviation = (F–14.04)/2.8988
### Proficiency data: Child L2ers

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of T-units (A)</td>
<td>14</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>10</td>
<td>14</td>
<td>12</td>
<td>15</td>
<td>13</td>
<td>11</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>14</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>14</td>
<td>12</td>
<td>15</td>
<td>14</td>
<td>11</td>
<td>12</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of words (B)</td>
<td>81</td>
<td>77</td>
<td>80</td>
<td>80</td>
<td>73</td>
<td>86</td>
<td>73</td>
<td>121</td>
<td>110</td>
<td>86</td>
<td>100</td>
<td>87</td>
<td>97</td>
<td>92</td>
<td>77</td>
<td>92</td>
<td>89</td>
<td>99</td>
<td>133</td>
<td>91</td>
<td>90</td>
<td>86</td>
<td>104</td>
<td>95</td>
<td>109</td>
<td>115</td>
<td>128</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of T-units with errors (C)</td>
<td>13</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>12</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy scores (E)</td>
<td>0.71</td>
<td>3.08</td>
<td>2.50</td>
<td>2.50</td>
<td>2.00</td>
<td>3.57</td>
<td>4.17</td>
<td>2.00</td>
<td>1.54</td>
<td>2.73</td>
<td>3.08</td>
<td>4.62</td>
<td>3.85</td>
<td>4.17</td>
<td>5.83</td>
<td>5.71</td>
<td>5.00</td>
<td>5.38</td>
<td>2.31</td>
<td>5.83</td>
<td>6.36</td>
<td>7.14</td>
<td>6.67</td>
<td>6.15</td>
<td>7.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z-score (G)</td>
<td>-2.20</td>
<td>-1.33</td>
<td>-1.22</td>
<td>-1.22</td>
<td>-1.13</td>
<td>-1.06</td>
<td>-0.88</td>
<td>-0.81</td>
<td>-0.80</td>
<td>-0.66</td>
<td>-0.59</td>
<td>-0.48</td>
<td>-0.42</td>
<td>-0.23</td>
<td>-0.17</td>
<td>-0.15</td>
<td>-0.04</td>
<td>0.17</td>
<td>0.20</td>
<td>0.31</td>
<td>0.42</td>
<td>0.59</td>
<td>0.70</td>
<td>0.74</td>
<td>0.76</td>
<td>0.87</td>
<td>0.92</td>
<td>1.17</td>
<td>1.26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** \( D = B/A; \) \( E = (A–C)/A*10; \) \( F = 1.2023*D+E; \) \( G = (F–\text{grand mean})/\text{standard deviation} = (F–14.04)/2.8988 \)
Appendix J: TVJT Items used for the English TC study

This section introduces each test item \((k = 16)\) for the English TC study, addressing why the context used in each item is neutral or agent-biased.

**Item 1 (Condition 1: Obligatorily transitive verb + Neutral context)**

| Elmo is the fastest in the class. He is better than Bert at playing tag because he can run fast. He can chase Bert and run away from him. | Today, Bert is wearing roller skates. He is very good at roller skating. When Bert and Elmo play tag, Bert is much faster than Elmo | **Bert is easy (hard) to catch today.** |

**Characteristic of the main property:**
- The roller skates: Likely to be advantageous no matter whether the wearer is a catcher (agent) or a catchee (patient).

**Predicted responses:**
- Correct object interpretation (i.e., *It is easy to catch Bert today*): **False**
- Incorrect subject interpretation (i.e., *It is easy for Bert to catch...today*): **True**
**Item 2 (Condition 2: Obligatorily transitive verb + Agent-biased context)**

| Cookie Monster and Big Bird like to play tag. Big Bird has Spider-man gloves. Big Bird can easily hold anything when he is wearing the magic gloves. He is better than Cookie Monster at playing tag because his gloves make him good at grabbing hold of Cookie Monster. | Today, Cookie Monster has borrowed the magic gloves from Big Bird. Cookie Monster and Big Bird are playing tag. Today, Cookie Monster is better than Big Bird at playing tag | **Cookie Monster is easy (hard) to catch today.** |

**Characteristic of the main property:**
- The magic gloves: Likely to be advantageous when the wearer is a catcher (agent).

**Predicted responses:**
- Correct object interpretation (i.e., *It is easy to catch Cookie Monster today*): **False**
- Incorrect subject interpretation (i.e., *It is easy for Cookie Monster to catch... today*): **True**
| Elmo and Bert used to play hide-and-seek in Bert’s village. Bert always played the game better than Elmo because Bert knew his village well. | Today, they are playing hide-and-seek in Elmo’s village. Elmo knows that the tree is the best place to look for Bert because, from the tree, he can see in all directions. He also knows that the best place to hide is among the many trees. Today, Elmo is better than Bert at playing hide-and-seek. | **Elmo is easy (hard) to find today** |

**Characteristic of the main property:**
- The tree: Likely to be advantageous no matter whether the user is a finder (agent) or a findee (patient).

**Predicted responses:**
- Correct object interpretation (i.e., *It is easy to catch Elmo today*): **False**
- Incorrect subject interpretation (i.e., *It is easy for Elmo to find...today*): **True**
**Item 4 (Condition 2: Obligatorily transitive verb + Agent-biased context)**

| Big Bird and Bert love to play hide-and-seek. Yesterday, Bert wore a blindfold as a penalty. Bert was not good at the game because he couldn’t see. | Today, Big Bird is wearing the blindfold. Big Bird is not good at the game today. | **Big Bird is easy (hard) to find today.** |

**Characteristic of the main property:**
- The blindfold: Likely to be disadvantageous when the wearer is a finder (agent)

**Predicted responses:**
- Correct object interpretation (i.e., *It is easy to catch Big Bird today*): **True**
- Incorrect subject interpretation (i.e., *It is easy for Big Bird to find...today*): **False**
**Item 5 (Condition 1: Obligatorily transitive verb + Neutral context)**

Bert and Cookie Monster used to play a toy hammer game. Bert was better than Cookie Monster in the game.

Today, Bert is stuck in a hole. He can’t do well at fighting. Cookie Monster is better than Bert at playing the toy hammer game today.

**Bert is easy (hard) to beat today.**

**Characteristic of the main property:**
- The hole: Likely to be disadvantageous no matter whether the stuck one is a beater (agent) or a beatee (patient).

**Predicted responses:**
- Correct object interpretation (i.e., *It is easy to beat Bert today*): **True**
- Incorrect subject interpretation (i.e., *It is easy for Bert to beat...today*): **False**
**Item 6 (Condition 2: Obligatorily transitive verb + Agent-biased context)**

<table>
<thead>
<tr>
<th>Big Bird and Elmo used to play a toy hammer game.</th>
<th>Today, Elmo has borrowed the hammer from Big Bird.</th>
<th>Elmo is easy (hard) to beat today.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yesterday, Big Bird was better than Elmo because Big Bird had a bigger, longer hammer.</td>
<td>Today, Elmo is better than Big Bird at playing the toy hammer game.</td>
<td></td>
</tr>
</tbody>
</table>

**Characteristic of the main property:**
- The hammer: Likely to be advantageous when the user is a beater (agent)

**Predicted responses:**
- Correct object interpretation (i.e., *It is easy to beat Elmo today*): False
- Incorrect subject interpretation (i.e., *It is easy for Elmo to beat...today*): True
Item 7 (Condition 1: Obligatorily transitive verb + Neutral context)

Elmo used to be better than Bert at playing tag.

Today, Bert is wearing a rocket. He is very fast and is using the rocket to fly through the sky. Bert is better than Elmo at playing tag today.

Bert is easy (hard) to touch today.

Characteristic of the main property:
- The rocket: Likely to be advantageous no matter whether the user is a toucher (agent) or a touch-ee (patient).

Predicted responses:
- Correct object interpretation (i.e., It is easy to touch Bert today): False
- Incorrect subject interpretation (i.e., It is easy for Bert to touch...today): True
**Item 8 (Condition 2: Obligatorily transitive verb + Agent-biased context)**

| Bert used to be better than Cookie Monster at playing tag because he had a magic stick that helped him stretch his arms as long as he wanted. | Today, Cookie Monster has borrowed the magic stick from Bert. Cookie Monster is better than Bert at playing tag today. | **Cookie Monster is easy (hard) to touch today.** |

**Characteristic of the main property:**

- The magic stick: Likely to be advantageous when the user is a toucher (agent).

**Predicted responses:**

- Correct object interpretation (i.e., *It is easy to touch Cookie Monster today*): **False**
- Incorrect subject interpretation (i.e., *It is easy for Cookie Monster to touch…today*): **True**
Item 9 (Condition 3: Intransitively-biased optionally transitive verb + Neutral context)

Cookie Monster is the fastest in class.

Today, Big Bird and Cookie Monster are skating. Big Bird is wearing skates. He is very good at skating. Big Bird is faster than Cookie Monster today.

Characteristic of the main property:
- The skates: Likely to be advantageous no matter whether the user is a racer (agent) or a racee (patient).

Predicted responses:
- Correct object interpretation (i.e., It is easy to race Big Bird today): False
- Incorrect subject interpretation (i.e., It is easy for Big Bird to race...today): True
**Item 10 (Condition 4: Intransitively-biased optionally transitive verb + Agent-biased context)**

| Bert and Elmo have bad eyesight. Bert used to be better than Elmo at skating because he wore good goggles. | Today, Elmo has borrowed the goggles from Bert. Elmo is better than Bert at skating today. | **Elmo is easy (hard) to race today.** |

Characteristic of the main property:
- The goggle: Likely to be advantageous when the user is a racer (agent).

**Predicted responses:**
- Correct object interpretation (i.e., *It is easy to race Elmo today*): *False*
- Incorrect subject interpretation (i.e., *It is easy for Elmo to race...today*): *True*
**Item 11 (Condition 3: Intransitively-biased optionally transitive verb + Neutral context)**

Bert and Big Bird used to practice taekwondo. Bert was always better than Big Bird.

Today, Bert is tied up. He cannot use any moves. Big Bird is better than Bert at playing Taekwondo today.

**Bert is easy (hard) to fight today.**

**Characteristic of the main property:**
- Tied up: Likely to be disadvantageous no matter whether the tied one is a fighter (agent) or a fightee (patient).

**Predicted responses:**
- Correct object interpretation (i.e., *It is easy to fight Bert today*): **True**
- Incorrect subject interpretation (i.e., *It is easy for Bert to fight...today*): **False**
Item 12 (Condition 4: Intransitively-biased optionally transitive verb + Agent-biased context)

Elmo and Cookie Monster used to practice taekwondo. Cookie Monster was better than Elmo because Cookie Monster had goggles that helped him see well.

Today, Elmo has borrowed the goggles. They are practicing taekwondo, and Elmo is better than Cookie Monster today

Elmo is easy (hard) to fight today.

Characteristic of the main property:
- Goggle: Likely to be advantageous when the user is a fighter (agent).

Predicted responses:
- Correct object interpretation (i.e., It is easy to fight Elmo today): False
  - Incorrect subject interpretation (i.e., It is easy for Elmo to fight...today): True

207
**Item 13 (Condition 3: Intransitively-biased optionally transitive verb + Neutral context)**

| A wolf tried to eat the baby animal. The wolf was stronger than the baby animal. | After some time, the baby animal became a lion. He became stronger. Today, the lion is trying to eat the wolf, and the wolf cannot run away. | The lion is easy (hard) to hunt today. |

**Characteristic of the main property:**
- Strong: Likely to be advantageous no matter whether the strong one is a hunter (agent) or a huntee (patient).

**Predicted responses:**
- Correct object interpretation (i.e., *It is easy to hunt the lion today*): **False**
  - Incorrect subject interpretation (i.e., *It is easy for the lion to hunt...today*): **True**
**Item 14 (Condition 4: Intransitively-biased optionally transitive verb + Agent-biased context)**

| A bear has a magic stick. He can freeze anything. Every animal is afraid of him. | Today, a tiger has stolen the magic stick from the bear. Today, the tiger is freezing the bear and trying to eat the bear. | **The tiger is easy** *(hard)* **to hunt today.** |

**Characteristic of the main property:**
- The magic stick: Likely to be advantageous when the user is a hunter (agent).

**Predicted responses:**
- Correct object interpretation (i.e., *It is easy to hunt the tiger today*): **False**
- Incorrect subject interpretation (i.e., *It is easy for the tiger to hunt...today*): **True**
### Item 15 (Condition 3: Intransitively-biased optionally transitive verb + Neutral context)

| Characteristic of the main property: | 
|-------------------------------------|---|
| - Strong and fast: Likely to be advantageous no matter whether the strong and fast one is a mover (agent) or a movee (patient). | 

<table>
<thead>
<tr>
<th>Predicted responses:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Correct object interpretation (i.e., <em>It is easy to move Elmo today</em>): <strong>False</strong></td>
</tr>
<tr>
<td>- Incorrect subject interpretation (i.e., <em>It is easy for Elmo to move...today</em>): <strong>True</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bert and Elmo love to practice Japanese sumo wrestling, a sport to push another wrestler out of the ring. Bert used to be better than Elmo at Japanese sumo wrestling. He was strong and fast.</th>
<th>Elmo practiced a lot. Today, Elmo is stronger and faster. Elmo is better than Bert at sumo wrestling today</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Elmo is easy (hard) to move today.</strong></td>
<td></td>
</tr>
</tbody>
</table>
Item 16 (Condition 4: Intransitively-biased optionally transitive verb + Agent-biased context)

Big Bird and Cookie Monster love to practice Japanese sumo wrestling, a sport to push another wrestler out of the ring. Big Bird was better than Cookie Monster because Big Bird wore sumo pants, which are very light and comfortable.

Today, Cookie Monster has borrowed the pants. He feels comfortable while wearing the pants. Cookie Monster is better than Big Bird at sumo wrestling today.

Cookie Monster is easy (hard) to move today.

Characteristic of the main property:
- Sumo pants: Likely to be advantageous when the user is a mover (agent).³

Predicted responses:
- Correct object interpretation (i.e., It is easy to move Cookie Monster today): False
- Incorrect subject interpretation (i.e., It is easy for Cookie Monster to move... today): True

³ The meanings of move are different between the transitive and intransitive meanings. In the transitive, move means pushing somebody else, whereas in the intransitive, move means moving one’s own body. The use of the light and comfortable sumo pants is aimed to bias to the subject interpretation in the intransitive reading, by helping the wearer move his own body easily.
**Filler item 1**

<table>
<thead>
<tr>
<th>Big Bird and Elmo are in an airport</th>
<th>Elmo is hugging Big Bird.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>True / False / I don’t know</em></td>
<td></td>
</tr>
</tbody>
</table>

**Filler item 2**

<table>
<thead>
<tr>
<th>Cookie Monster and Bert are in an airport</th>
<th>Bert is hugged by Cookie Monster.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>True / False / I don’t know</em></td>
<td></td>
</tr>
</tbody>
</table>

**Filler item 3**

<table>
<thead>
<tr>
<th>Big Bird and Cookie Monster are in a classroom</th>
<th>Big Bird is kicking Cookie Monster.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>True / False / I don’t know</em></td>
<td></td>
</tr>
</tbody>
</table>
| Filler item 4 | Elmo is kicked by Bert.  
(True / False / I don’t know) |
|---|---|
| Filler item 5 | Big Bird and Bert are on an ice-rink  
Bert is pushing Big Bird.  
(True / False / I don’t know) |
| Filler item 6 | Cookie Monster and Elmo are on an ice-rink  
Elmo is pushed by Cookie Monster.  
(True / False / I don’t know) |
Filler item 7

Bert and Elmo are in the classroom.  
Bert is scolding Elmo.  
(True / False / I don’t know)

Filler item 8

Cookie Monster and Big Bird are in the classroom.  
Cookie Monster is scolded by Big Bird.  
(True / False / I don’t know)
Introduction of the characters

Who are they?

Cookie Monster  Elmo  Big Bird  Bert

Practice item 1

Here are Cookie Monster and Big Bird.  **Big Bird is chasing Cookie Monster.**
(True / False / I don’t know)

Practice item 2

Here are Elmo and Bert.  **Bert is following Elmo.**
(True / False / I don’t know)
Appendix K: AJT items used for the English TC study

- Name: __________

<table>
<thead>
<tr>
<th>No.</th>
<th>Sentence</th>
<th>1—2—3—4—5—6 Very natural</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kim is touching Lee.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>2</td>
<td>Kim is finding.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>3</td>
<td>A tiger is hunting a rabbit.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>4</td>
<td>Kim is fighting Lee.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>5</td>
<td>Kim is racing.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>6</td>
<td>Kim is moving Lee.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>7</td>
<td>Kim is catching Lee.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>8</td>
<td>Kim is fighting.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>9</td>
<td>Kim is moving.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>10</td>
<td>Kim is finding Lee.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>11</td>
<td>Kim is catching.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>12</td>
<td>Kim is beating Lee.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>13</td>
<td>Kim is racing Lee.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>14</td>
<td>A tiger is hunting.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>15</td>
<td>Kim is touching.</td>
<td>Very unnatural</td>
</tr>
<tr>
<td>16</td>
<td>Kim is beating.</td>
<td>Very unnatural</td>
</tr>
</tbody>
</table>

- Write down any words that you do not know the meaning of.  ___________________________
Appendix L: Cloze test used for the English TC study

Directions:

1. Read the passage quickly to get the general meaning.
2. Write only one word in each blank in the column to the right. Contractions (example: don’t) and possessives (John’s bicycle) are one word.

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) but he didn’t hurt himself.

1. **down**
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 (2) 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 (34) in the Paleolithic period of man’s (35). Since no wood or bone tools (36) survived, we know of this man (37) his stone implements, with which he (38) kill animals, cut up the meat, (39) scrape the skins, as well as (40) pictures on the walls of the (41) where he lived during the winter.

(42) the warmer seasons, man wandered on (43) steppes of Europe without a fixed (44), always foraging for food. Perhaps the (45) carried nuts and berries in shells (46) skins or even in light, woven (47). Whenever they camped, the primitive people (48) fires by striking flint for sparks (49) using dried seeds, moss, and rotten (50) 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.
Appendix M: Number of object interpretations: English TVJT vs. Korean TVJT

Tables M-1 and M-2 present the number of object interpretations in the English TVJT and in the Korean TVJT by the adult L2 group and the child L2 group.

Table M-1. Number of object interpretations in the English TVJT vs. Korean TVJT: Adult L2ers

<table>
<thead>
<tr>
<th>Proficiency Group</th>
<th>Participant number</th>
<th>English TVJT (max = 16)</th>
<th>Korean TVJT (max = 16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>aL2ers-Low</td>
<td>P1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>P3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>P5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>P6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>P7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P9</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>P10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P11</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P12</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>P13</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>aL2ers-Mid</td>
<td>P14</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>P15</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>P16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P17</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>P18</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P19</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P20</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>P21</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>P22</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>P23</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>P24</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>P25</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>P26</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P27</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>P28</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>P29</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>P30</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>P31</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P32</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>aL2ers-High</td>
<td>P33</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>P34</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P35</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>P36</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>P37</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>P38</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>P39</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>P40</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>P41</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>P42</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P43</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>P44</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>P45</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>P46</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P47</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>P48</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P49</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>Proficiency Group</td>
<td>Participant number</td>
<td>English TVJT (max = 16)</td>
<td>Korean TVJT (max = 16)</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------</td>
<td>-------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>cL2ers-Low</td>
<td>P1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>P3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>P4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>P5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>P6</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P7</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>P8</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P9</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>P10</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>P11</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>cL2ers-Mid</td>
<td>P12</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>P13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P14</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>P15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P16</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>P17</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>P18</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>P19</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P20</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>P21</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>cL2ers-High</td>
<td>P22</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P23</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>P24</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P25</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P26</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>P27</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>P28</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>P29</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>P30</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>
REFERENCES


Bruening, B. (2012). *No such thing as “defective intervention.”* Unpublished manuscript, University of Delaware, Newark.


Hunt, K. (1965). Grammatical structures written at three grade levels. NCTE Research report No. 3. Champaign, IL: NCTE.


