

ABSTRACT

Title of Thesis: EXPLICIT AND IMPLICIT COGNITIVE
APTITUDES, L2 OUTCOME MEASURES,
AND LEARNING OF MORPHOSYNTAX
UNDER AN INCIDENTAL CONDITION

Ryo Maie, Master of Arts, 2018

Thesis Directed By: Professor Robert M. DeKeyser

This study is the first to combine recent methodological advances to the measurement of explicit and implicit knowledge in an investigation of learning under incidental exposure. Participants were exposed to a semi-artificial language, Japlish, and subsequently tested as to the extent to which they had developed explicit and/or implicit knowledge. Subjective measures of awareness, objective measures of linguistic knowledge, and explicit and implicit cognitive aptitudes were employed to triangulate learning outcomes at two testing sessions.

Overall results shed new light on the complexity of explicit and implicit learning under incidental conditions. Both learning types were confirmed in the experiment, but they occurred to a different degree and extent. Furthermore, the study identifies clear discrepancies among the four approaches to measuring explicit and implicit knowledge, with some being rigorous and others tending to underestimate or overestimate. The study calls for future research with more longitudinal and situated analyses of the phenomena.

EXPLICIT AND IMPLICIT COGNITIVE APTITUDES, L2 OUTCOME
MEASURES, AND LEARNING OF MORPHOSYNTAX UNDER AN
INCIDENTAL CONDITION

by

Ryo Maie

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Arts
2018

Advisory Committee
Professor Robert M. DeKeyser, Chair
Professor Michael H. Long
Professor Steven J. Ross

© Copyright by
Ryo Maie
2018

Acknowledgements

There are a quite few people I would like to send my sincere gratitude to. First and foremost, I would like to thank my advisor, Dr. Robert DeKeyser, for his continual guidance, kindness, and warmth. Without his mentorship and frequent support, this thesis would have never been possible, and I even might not have become a student at UMD in the first place. Robert's view on SLA certainly formed the basis of mine, and I hope this thesis embodies that.

In addition, I also would like to thank the other members of my thesis committee. Dr. Michael Long has inspired me ever since I started SLA and his professional and emotional encouragement helped me get through many setbacks in my master's study at UMD (and many soccer talks too). Dr. Steven Ross, my *sensei*, patiently helped me through statistical analyses in the study and through my frequent administrative issues (!!!). I believe NO ONE could ask for a better thesis committee than mine here.

I would also like to thank many contributors to the materials and the design of the present study. Bradford Salen, without any doubt, deserves the first mention, for his untiring collaboration and for our good friendship. My gratitude must also go to Dr. Yuichi Suzuki for his provision of the material for LLAMA-F test and consultation about the material designs, Dr. Alison Mackey and Dr. Paula Winke for their provision of the listening-span task, Dr. Beatriz Lado for her provision of the nonword repetition task, and Dr. Scott Barry Kaufman for his provision of the probabilistic serial reaction time task.

Finally, my *arigato* goes to my family, *Masaru*, *Yumiko*, and *Yusuke*, for their endless support throughout my life. I am forever indebted to the people who helped me throughout this study.

Table of Contents

Acknowledgements.....	ii
Table of Contents.....	iv
List of Tables.....	vi
List of Figures.....	vii
Chapter 1: Introduction.....	1
Chapter 2: Review of the Literature.....	4
2.1 Explicit and Implicit Learning in SLA.....	4
2.1.1 Learning, Condition, and Knowledge: Definitions.....	4
2.1.2 Experimental Studies of L2 Learning under Incidental Conditions.....	8
2.1.3 Limitations in Previous Studies.....	21
2.1.4 Use of Artificial and Semi-Artificial Languages in SLA Research.....	26
2.2 Cognitive Aptitudes.....	29
2.2.1 Theoretical Underpinnings.....	29
2.2.2 Cognitive Aptitudes in SLA.....	32
2.2.3 Explicit and Implicit Cognitive Aptitudes.....	34
2.3 Gaps in the Literature.....	43
Chapter 3: Research Questions and Hypotheses.....	46
Chapter 4: Methods.....	54
4.1 Participants.....	54
4.2 Materials.....	55
4.2.1 Exposure Task Materials.....	55
4.2.2 Untimed Auditory Grammaticality Judgment Task.....	56
4.2.3 Word-Monitoring Task.....	58
4.2.4 LLAMA-F.....	61
4.2.5 Listening-Span Task.....	62
4.2.6 Phonological Short-Term Memory Task.....	63
4.2.7 Serial Reaction Time Task.....	64
4.3 Procedure.....	65
4.4 Analysis.....	69
4.4.1 Untimed Auditory Grammaticality Judgment Tasks.....	69
4.4.2 Word-Monitoring Task.....	72
4.4.3 Retrospective Verbal Reports.....	73
Chapter 5: Results.....	75
5.1 Immediate Posttests.....	75
5.1.1 Untimed Auditory Grammaticality Judgment Task.....	75
5.1.2 The Role of Cognitive Aptitudes.....	79
5.1.3 The Role of Confidence Ratings and Source Attributions.....	82
5.1.4 Word-Monitoring Task.....	88
5.1.5 The Role of Cognitive Aptitudes.....	90
5.1.6 Summary of Immediate Posttests.....	92
5.2 Delayed Posttests.....	94
5.2.1 Untimed Auditory Grammaticality Judgment Task.....	94
5.2.2 The Role of Cognitive Aptitudes.....	99
5.2.3 The Role of Confidence Ratings and Source Attributions.....	100
5.2.4 Word-Monitoring Task.....	104

5.2.5 The Role of Cognitive Aptitudes	108
5.2.6 Summary for the Delayed Posttest.....	109
5.3 Post-Experimental Verbal Reports	111
Chapter 6: Discussion	114
6.1 Research Questions and Hypotheses	114
6.2. General Discussion	124
Chapter 7: Conclusions and Limitations.....	131
Appendices.....	135
Appendix A: Exposure Task Stimuli	135
Appendix B: AGJT Stimuli	142
Appendix C: WMT Stimuli	147
Appendix D: Biographical Questionnaire.....	157
Appendix E: Post-Experimental Questionnaire	158
Bibliography	162

List of Tables

Table 1. <i>Untimed AGJT: Item Types</i>	57
Table 2. <i>Procedure of the Study</i>	66
Table 3. <i>Mean Percentile Accuracy for the Untimed AGJT at the Immediate Posttest</i>	75
Table 4. <i>d-prime Scores on the Untimed AGJT at the Immediate Posttest</i>	76
Table 5. <i>Post-hoc Difference between the Construction Types</i>	79
Table 6. <i>Descriptive Statistics of the Cognitive Aptitude Measures</i>	80
Table 7. <i>Correlational Matrix for the Cognitive Aptitude Measures</i>	80
Table 8. <i>Correlational Matrix for NWRT and Six Construction Types</i>	82
Table 9. <i>Results of Logistic Regression on Confidence Ratings at the Immediate Posttest</i>	84
Table 10. <i>Mean Proportion of Accuracy for Each Source Attribution Category at the Immediate Posttest</i>	87
Table 11. <i>Mean Word-Monitoring Latencies at the Immediate Posttest</i>	88
Table 12. <i>Mean Percentile Accuracy for the Untimed AGJT at the Delayed Posttest</i>	94
Table 13. <i>d-prime Scores on the Untimed AGJT at the Delayed Posttest</i>	95
Table 14. <i>Post-hoc Differences between the Construction Types at Delayed Posttest</i>	99
Table 15. <i>Results of Logistic Regression on Confidence Ratings at Delayed Posttest</i>	101
Table 16. <i>Mean Proportion of Accuracy for Each Source Attribution Category at the Delayed Posttest</i>	104
Table 17. <i>Mean Word-Monitoring Latencies at the Delayed Posttest</i>	105
Table 18. <i>Post-hoc RT Difference between the Construction Types: Experimental</i>	107
Table 19. <i>Post-hoc RT Difference between the Construction Types: Control</i>	107
Table 20. <i>Post-Experimental Retrospective Verbal Reports</i>	111
Table 21. <i>Results of One-Sample t-test for the Control Group</i>	113

List of Figures

<i>Figure 1.</i> Procedure for the untimed AGJT.....	59
<i>Figure 2.</i> Procedure for the WMT.....	61
<i>Figure 3.</i> Histogram of d-prime Scores for the Experimental Group at the Immediate Posttest.....	77
<i>Figure 4.</i> Boxplot of the AGJT at the Immediate Posttest.....	78
<i>Figure 5.</i> Scatterplot of the Residuals and the Predicted Values on AGJT at the Immediate Posttest.....	81
<i>Figure 6.</i> Scatterplot Matrix: NWRT, OSV, OSIV, and OSSVV at the Immediate Posttest.....	83
<i>Figure 7.</i> Scatterplot Matrix: NWRT, OSSIVV, CaseMis, and CaseMix at the Immediate Posttest.....	83
<i>Figure 8.</i> Confidence Level and Overall Accuracy at the Immediate Posttest.....	85
<i>Figure 9.</i> Confidence Level and OSV Accuracy at the Immediate Posttest.....	86
<i>Figure 10.</i> Confidence Level and CaseMix Accuracy at the Immediate Posttest.....	86
<i>Figure 11.</i> Histogram of TransRTs for the Control Group at the Immediate Posttest.....	89
<i>Figure 12.</i> Scatterplot of the Residuals and the Predicted Values on WMT at the Immediate Posttest.....	91
<i>Figure 13.</i> Histogram of d-prime Scores for the Experimental Group at the Delayed Posttest.....	96
<i>Figure 14.</i> Histogram of d-prime Scores for the Control Group at the Delayed Posttest.....	96
<i>Figure 15.</i> Boxplot of the AGJT at the Delayed Posttest.....	98
<i>Figure 16.</i> Boxplot for the Immediate and Delayed Posttest.....	99
<i>Figure 17.</i> Scatterplot of the Residuals and the Predicted Values on AGJT at the Delayed Posttest.....	100
<i>Figure 18.</i> Confidence Level and Overall Accuracy at the Delayed Posttest.....	102
<i>Figure 19.</i> Confidence Level and OSV Accuracy at the Delayed Posttest.....	102
<i>Figure 20.</i> Confidence Level and CaseMix Accuracy at the Delayed Posttest.....	103
<i>Figure 21.</i> Scatterplot of the Residuals and the Predicted Values on WMT at the Delayed Posttest.....	109

Chapter 1: Introduction

Constructs of explicit and implicit learning of second language (L2) are now central to the understanding of learning processes underlying second language acquisition (SLA). According to Rebuschat (2008) and Rebuschat and Williams (2009, 2012), the interest in this debate was primarily sparked by the theoretical work of Stephen Krashen's Monitor Model (1981, 1982, 1994), and since then, empirical research in SLA has witnessed a fruitful number of investigations exploring whether and how explicit and/or implicit learning take place in L2 learning, and also what kinds of L2 knowledge such learning processes lead to (for a recent review, see Rebuschat, 2015). Among those investigations, in particular, recent studies seek to examine whether adult L2 learners still retain the ability to implicitly induce regularities and patterns of the target language, and if any, what constraints there are on such learning process. The study presented here serves to further our understanding of (a) how explicit and implicit learning processes interact with complexity of a language, (b) what types of knowledge these forms of learning lead to, and (c) how explicit and implicit cognitive aptitudes moderate the above (a)-(b).

According to R. Ellis (2015), there are two ways of investigating explicit and implicit learning of language: (i) by investigating whether and in what conditions the learning process takes place with or without awareness and (ii) by examining whether learners are aware of the knowledge that resulted from the learning. In the very same vein, Leow (2015) and Leow and Hama (2013) also describe that two methodological approaches exist in the current research on explicit and implicit learning, one with its

focus on learning as *processes* and the other with its focus on learning as *products*. While the former often operationalizes the construct of awareness assessed at the stage of knowledge encoding with such methodologies as think-aloud protocols, the latter examines the awareness at the stage of knowledge retrieval, directly assessing whether learners are aware of the knowledge at all.

In addition to the two methodological approaches to explicit and implicit learning of language, recent SLA research produces another type of approach which seeks to infer the learning processes by inspecting how explicit and implicit cognitive aptitudes relate to resulting learning outcomes (e.g., Granena, 2013a; Robinson, 2005a; Suzuki, 2015; Suzuki & DeKeyser, 2017; Yi, in press). This is based on a recent claim by DeKeyser (2012) that the interaction of learning outcomes and individual differences in them allows us to infer the nature of learning processes because “a treatment variable interacts with an ID variable because the treatment variable requires a mental process that is facilitated/hampered by the value of the ID variable” (p. 190). This thesis study draws on such recent wealth of methodological approaches in SLA, and attempts to combine the latter two approaches by investigating how explicit and implicit cognitive aptitudes moderate learning of a semi-artificial language from brief incidental exposure, assessed by two L2 outcome measures hypothesized to gauge explicit and implicit knowledge, respectively.

In what follows, experimental studies of explicit and implicit learning under incidental conditions will be reviewed with the focus on studies conducted in controlled laboratory settings. An incidental condition here is defined as a learning condition in which subjects are exposed to exemplar sentences without being told that

they would be learning a language nor that there would be posttests that would assess their learning outcomes. Furthermore, a recent line of inquiry on explicit and implicit cognitive aptitudes will also be discussed, and then four cognitive aptitude measures employed in the present investigation will be introduced. It is hypothesized that those four cognitive aptitude measures tap into domain-general cognitive abilities that lie on a continuum of explicit to implicit processing. Lastly, gaps in the current literature on explicit and implicit learning are identified in relation to the cognitive aptitudes, leading up to the description of the current investigation.

Chapter 2: Review of the Literature

2.1 Explicit and Implicit Learning in SLA

When Krashen first formulated the acquisition-learning hypothesis, it did not explicitly distinguish between incidental and implicit learning and intentional and explicit learning. Even though it is true that their occurrences would naturally correlate in that incidental learning can lead to implicit learning and intentional learning entails explicit, conscious awareness of learning experiences (see DeKeyser, 2003; Hulstijn, 2003 for a discussion of the implicit-incidental and explicit-intentional issue). The two pairs of learning processes do differ by definition and also in nature (Hulstijn, 2003, 2013), and thus any SLA studies on this matter must first clearly define and distinguish them. In the following, definitions and terminologies regarding explicit and implicit learning, condition, and knowledge are introduced briefly and distinguished from those of incidental and intentional learning. Then what follows is a body of SLA literature on explicit and implicit learning with its focus on learning of target linguistic features under incidental conditions.

2.1.1 Learning, Condition, and Knowledge: Definitions

In brief, *implicit learning* can be defined as “the acquisition of knowledge about the underlying structure of a complex stimulus environment by a process that takes place naturally, simply, and without conscious operations” (N. Ellis, 1994, p.1). This learning process, according to Rebuschat and Williams (2012, 2013), must satisfy two criteria, *intentionality* and *awareness*, in that during implicit learning, learners are not intending to learn (i.e., - *intentionality*); nor are they aware of the

learning that is taking place, resulting in unconscious, implicit knowledge (i.e., -*awareness*). The term was first adopted by Arthur Reber in his seminal study of implicit learning of an artificial grammar (Reber, 1967, 1969) and subsequently characterized as a process that is “fast, effortless, unconscious, procedural, domain-independent, bottom up, intuitive, and associative” (Reber, 2011, p. 30). Explicit learning, on the other hand, is of more conscious operations, during which learners formulate and test hypotheses about the underlying. In contrast to implicit learning, explicit learning is slow, effortful, conscious, declarative, domain-dependent, top down, rational, and linear (Reber, 2011, as cited in N. Ellis, 2015). As such, it often results in L2 knowledge that learners are aware of possessing, explicit knowledge. Although these two learning processes have been pitted against one another in SLA, they are not mutually exclusive in reality, and both learning types can occur during incidental exposure to a target language.

While explicit and implicit learning are often distinguished by the criterion of awareness, incidental learning and intentional learning are contrasted with reference to the existence of intentionality. According to Schmidt (1994), the field of SLA has defined the term incidental learning in three overlapping senses: (a) learning without intention to learn, (b) learning of an aspect of a stimulus while the primary attention is on some other aspect, and (c) learning of formal features of language while the primary attention is on meaning. In contrast, intentional learning is a more conscious and explicit process during which learners intentionally engage in learning or search for the underlying structure (Hulstijn, 2003, 2013). As already discussed, incidental learning is a process that can lead to implicit learning, provided that learners are not

aware of the knowledge they have acquired, but it can also lead to explicit learning as well, as learners become aware of the existence of underlying patterns (DeKeyser, 2003). Thus, implicit learning implicates more than what incidental learning conveys, and the presence of the former always entails the latter but the reverse is not necessarily true (Hulstijn, 2003).

In addition to such learning processes, research in SLA is also prolific in producing empirical studies on explicit and implicit *knowledge* and *conditions*. The distinction between explicit and implicit knowledge is a subject of heated debates in the field, most of which are in relation to the issue of their interface (e.g., DeKeyser, 2015; N. Ellis, 2005, 2015; Hulstijn, 2002; Paradis, 2009). According to the literature, the degree of awareness can be the primary criterion for distinguishing between explicit and implicit knowledge (DeKeyser, 2003; Suzuki, 2017; Vafaei, Suzuki, & Kachinske, 2017; Williams, 2009). Implicit knowledge is tacit knowledge that language users are not aware of possessing, and it can be deployed without awareness of doing so. Explicit knowledge, on the other hand, is conscious knowledge that language users are aware of possessing, and its use often requires conscious awareness of doing so¹.

Thus, at least in the short run, implicit learning is a process that primarily results in implicit knowledge and explicit learning is a process that results in explicit

¹ However, use of explicit knowledge can be functionally equivalent to that of implicit knowledge through automatization of its access and processing (e.g., DeKeyser, 2015). That is, explicit knowledge can be accessed automatically, such that their use does not involve conscious awareness anymore. As Logan (1988) argues, this is primarily phenomenal in that it is extremely difficult to introspect on things that are executed automatically.

knowledge. However, it is not always true that instructional conditions that are deemed conducive to cognitive processes underlying such learning processes actually guarantee that they would happen (Long, 2017). Instructional conditions can be implicit, explicit, incidental, or intentional, depending on the cognitive processes *intended* by the instructor, but nevertheless, the learning could progress independently of the instructor's intent. In meta-analytic studies on the effectiveness of various instructional conditions, explicit types of instruction (or conditions) are defined as treatments that involve rule explanation, provision of metalinguistic feedback, and/or some other means through which teachers directly draw student's attention to linguistic forms; conversely, implicit types of instruction are ones that do not involve them (Goo, Granena, Yilmaz, & Novella, 2015; Norris & Ortega, 2000; Spada & Tomita, 2010). In the present study, subjects were exposed to a semi-artificial language, *Japlish*, under an incidental condition. They were neither notified that the language consisted of some underlying word order and case marking patterns, nor that there would be immediate and delayed posttests assessing their learning of the structures. Thus, the study made no prior assumptions about the types of learning processes and L2 knowledge resulting from the learning. However, it sought to investigate them through four triangulated methodological approaches to measuring the construct of awareness and explicit and implicit knowledge: (a) two outcome measures (i.e., an untimed auditory grammaticality judgment task: AGJT, and a word-monitoring task: WMT), (b) subjective confidence ratings and source attributions, (c) subsequent retrospective verbal reports, and (d) explicit and implicit cognitive aptitudes.

2.1.2 Experimental Studies of L2 Learning under Incidental Conditions

To date, a number of empirical studies have been conducted with the aim of uncovering how explicit and implicit learning progress under incidental conditions, and what types of knowledge those learning processes result in. This section reviews laboratory studies of L2 learning under incidental conditions, focusing on five empirical questions concerning explicit and implicit learning by adult L2 learners: (a) Is incidental/implicit learning of L2 possible?, (b) What types of structures are amenable to implicit learning process?, (c) To what extent does incidental exposure result in implicit knowledge?, (d) Is knowledge learned under incidental conditions durable?, and (e) Do individual differences play a role under incidental conditions?

Is Incidental/Implicit Learning of L2 Possible?

The question of whether adult L2 learners can learn aspects of the target language without intention and/or without being aware has been the subject of numerous discussions (e.g., Hama & Leow, 2010; Leow, 1997, 2000; Leow & Hama, 2013; Williams, 2009) as well as experimental studies in SLA (DeKeyser, 1994, 1995; Graham & Williams, 2016; Grey, Williams, & Rebuschat, 2014; Hamrick & Rebuschat, 2013; Kachinske, Osthus, Solovyeva, & Long, 2015; Kerz & Wiechmann, 2017; Leung & Williams, 2012, 2014, 2015; Paciorek & Williams, 2015; Rebuschat, 2008; Rebuschat, Hamrick, Riestenberg Sachs, & Ziegler, 2015; Rebuschat & Williams, 2012; Rogers, Révész, & Williams, 2016; Williams, 2005; Williams & Kuribara, 2008). The first well-known study of implicit learning, though in experimental psychology rather than SLA, was conducted by Reber (1967, 1969). In that study, subjects were presented with letter sequences generated by a finite-state

grammar and instructed to memorize those sequences. Here, the subjects were not notified that the letter arrangements followed rules of the grammar; nor were they told that they would be tested on their knowledge of the grammar later in the session. In a subsequent testing of recognizing old and new letter arrangements, Reber found that the subjects classified 78.9% of the testing sentences correctly, suggesting that they acquired an underlying abstract representation of the grammar rules. Also, this was typically true despite the fact that the subjects were unable to verbalize the rules (Reber, 1989). A large number of studies in cognitive psychology followed this line of research and there seems to be a consensus that implicit learning from incidental exposure to a stimulus domain is possible (see Berry & Dienes, 1993; Cleermans, Destrebecqz, & Boyer, 1998; Perruchet, 2008; Pothos, 2007; Reber, 1993; Shanks, 2005, for a review). However, it is yet to be discovered in regards to the nature of the resulting knowledge implicit learning leads to. In contrast to the claim by Reber that abstract rules of grammar can be implicitly acquirable, an alternative account has also been proposed that fragment-based knowledge of bigrams or trigrams can explain performance by participants on the classification task.

In the field of SLA, Williams (2004, 2005) is widely considered to have initiated the line of recent implicit L2 learning research. In Williams (2005), subjects were exposed to a semi-artificial language, consisting of English lexical items and four artificial article determiners (*gi*, *ro*, *ul*, and *ne*), which were described to encode distance of nouns they modify (*gi* and *ro* for near objects; *ul* and *ne* for far objects). Unbeknownst to the participants, however, these articles also encoded the animacy of the nouns (*gi* and *ul* for animate objects; *ro* and *ne* for inanimate objects). In doing so,

each subject was told to indicate whether the determiners referred to a near or far object and to make a mental imagery of the situation conveyed by the sentences during an exposure phase. After the training, a surprise testing session was announced, during which the participants were presented with two noun phrases which contained the same noun modified by two different types of determiners only differing in animacy (e.g., *gi cushion* vs. *ro cushion*). The results showed that the subjects scored 61% accuracy (i.e., above chance) on the test, and a subsequent interview also revealed that 71% of them remained unconscious about the animacy rule. Based on these results, the author concluded that “the present experiments show that, at least for some individuals, it is possible to learn form-meaning connections without awareness of what those connections are” (p. 293).

In sum, there is a sufficient amount of empirical studies from which we can conclude that implicit learning by adult L2 learners is indeed feasible under incidental conditions. However, some researchers pointed out that existing studies have not provided any strict evidence showing implicit learning indeed occurred during those experiments (e.g., Hama & Leow, 2010; Jackson, 2013). Furthermore, retrospective verbal reports have been criticized for being an insensitive measure of awareness; it might be the case that subjects explicitly and intentionally learned the target features, but they were just unable to verbalize the rules due to several issues inherent to the methodology (cf. Hama & Leow, 2010). Thus, it is evident that new SLA studies on explicit and implicit learning must adopt more than one measure of awareness and carefully triangulate whether subjects are aware of the knowledge they have acquired (Rebuschat, Hamrick, Riestenberg, Sachs, & Ziegler, 2015).

What Types of Structures Are Amenable to Implicit Learning Processes?

Most studies of implicit learning have primarily focused on acquisition of morphosyntax. For instance, Rebuschat and Williams (2009, 2012) conducted two experiments to investigate whether adult L2 learners were able to learn non-native word order rules of a semi-artificial language (i.e., English vocabulary plus German word orders). In Experiment 1, the learners in the experimental group, who were exposed to 128 training sentences, correctly classified test sentences at a significantly above-chance level. However, the learning was only evidenced for those who were consciously aware of having acquired some kind of knowledge. In Experiment 2, the authors reduced the number of rules so the subjects could process the word order more directly and trained them with elicited imitation tasks in addition to the plausibility judgments. The results showed that the subjects performed significantly above chance, and the training resulted in both conscious and unconscious knowledge. Similarly, Williams and Kuribara (2008) adopted a different semi-artificial language, Japlish, which combined English lexis with Japanese syntax and case marking. They examined whether subjects can learn rules of Japanese syntax (i.e., head-direction and scrambling) under an incidental condition. The results showed that the experimental group outperformed the control group on GJTs at the posttest, despite the fact that they were unable to explicitly verbalize the rules. Since then, several SLA studies to date have replicated the findings on the implicit learning of non-native syntax (e.g., Grey, Williams, & Rebuschat, 2014; Kachinske, Osthus, Solovyeva, & Long, 2015; Rebuschat, 2008; Rebuschat & Williams, 2009, 2012; Williams & Rebuschat, 2012).

Besides simple syntactic constructions, Leung and Williams (2011) examined implicit learning of form-meaning mappings. Using the same four article determiners (*gi*, *ro*, *ul* and *ne*) adopted in Williams (2005), they instructed the subjects that these articles encode distance, but unbeknownst to the subjects they also referred to the thematic role of the noun they modified (i.e., *gi* and *ul* for agents vs. *ro* and *ne* for patients). After being trained with 88 exemplar sentences, the subjects' reaction time significantly slowed down when the articles in test sentences violated the agreement rule of the thematic role (e.g., *gi* and *ul* for patients). Furthermore, 20 out of 25 participants remained unaware of the relationship between the determiners and thematic role in a subsequent debriefing. Leung and Williams interpreted this as evidence of implicit learning of the form-meaning mappings. However, it seems that not all form-meaning connections are learnable from incidental exposure. For instance, another study by Leung and Williams (2012) with the same methodology found that a form-meaning mapping of the relative size of objects cannot be learned incidentally. The authors concluded that implicit learning of form-meaning mappings might be constrained by the nature of the meaning involved. Another type of form-meaning mapping that has been found to be amenable to implicit learning is a case-marking system (i.e., two types of Czech noun declension) that is incorporated in a semi-artificial language (Rogers, Révész, & Rebuschat, 2016).

Finally, recent studies have dealt with implicit learning of new areas of language. Paciorek and Williams (2015), for example, found that their subjects were able to learn semantic preference (i.e., the tendency of a word to co-occur with words sharing similar semantic features) in verb-noun collocations. In a similar vein,

Graham and Williams (2016) focused on a suprasegmental component of artificial words and reported that Latin stress regularities can be learned implicitly. In sum, many aspects of language (e.g., word orders, form-meaning mappings, semantic preference, and stress regularities) have been shown to be learnable incidentally, and in some cases, implicitly (according to the authors of the studies in question). However, as Leung and Williams (2012) suggested, there is a limit to what implicit learning is capable of producing.

To What Extent Does Incidental Exposure Result in Implicit Knowledge?

Whether and to what extent adult L2 learners are able to acquire implicit knowledge from incidental exposure is the most controversial area in this inquiry, on which researchers diverge to a significant degree. In part, this controversy is due to the fact that studies differ in their methodological approaches to the measurement of awareness (for a review, see Rebuschat, 2013). For instance, many studies using post-experimental verbal reports have found a majority of their participants remained unaware of rules of the target structures, suggesting that knowledge acquired from the exposure was implicit knowledge (Graham & Williams, 2016; Leung & Williams, 2011, 2012; Paciorek & Williams, 2015; Williams, 2005). However, offline verbal reports have been criticized for being an insensitive measure of awareness and suffer from several methodological issues, including memory decay and fabrication (e.g., Shanks & St. John, 1994). Indeed, this was the underlying motivation behind Hama and Leow (2010), who adopted a concurrent online measure of awareness (i.e., think-aloud protocols), in addition to the offline verbal reports, finding “at the encoding stage, unaware learners do not appear capable of selecting or producing the correct

determiner-noun combination when both animacy and distance information was included” (p. 487).

In addition to the online and offline measures of awareness, a recent line of SLA studies has begun to adopt new methods, *confidence ratings* and *source attributions* (Grey, Williams, & Rebuschat, 2014; Hamrick, 2013; Hamrick & Rebuschat, 2011; Jackson, 2014; Rebuschat, 2008; Rebuschat & Williams, 2009, 2012). These are normally coupled with a judgment or classification task, during which subjects are instructed to report how confident they are on each response and also what the basis of their knowledge is. The knowledge is deemed to be unconscious if the subjects claim to be ‘guessing’ when their actual performance is nevertheless above the baseline (*the guessing criterion*: Dienes et al., 1995) and/or the level of confidence they report is unrelated to the accuracy of their performance (*the zero-correlation criterion*: Chan, 1992; Dienes, et al., 1995). With these new methods, recent SLA studies have certainly added new insight into the understanding of how learners develop different types of L2 knowledge under incidental conditions, as they typically find that subjects developed both implicit and explicit knowledge from incidental exposure (e.g., Grey, Williams, & Rebuschat, 2014; Hamrick, 2013; Hamrick & Rebuschat, 2013; Rebuschat, 2008, experiment 6; Tagarelli, Borges-Mota, & Rebuschat, 2011).

Is Knowledge Learned under Incidental Conditions Durable?

In L2 interaction research, there is an abundance of experimental and quasi-experimental studies documenting that the memory traces learners develop from implicit types of corrective feedback are long-lasting. Indeed, two of the meta-

analyses on the effectiveness of corrective feedback report that implicit feedback (e.g., recast) is relatively more effective and longer-lasting than explicit feedback (e.g., metalinguistic explanation), even though their effects emerge more slowly (Li, 2010; Mackey & Goo, 2007). However, Grey, Williams, and Rebuschat (2014) lamented the fact that very few studies exist that have tested the durability of knowledge acquired from incidental exposure to the target language. One exception is Robinson (2002a), which included two delayed posttests at 1 week and 6 months after the training phase. In that study, native speakers of Japanese were incidentally exposed to a miniature version of Samoan after they were trained to criterion with vocabulary of that language. Three types of case marking (i.e., locative marker, ergative marker, and noun incorporation) served as the target of the study. Each sentence was presented on a computer screen, and the participants were instructed to try to understand the meaning of the sentences as much as they could. The results on GJTs at immediate testing showed that they were able to judge old items accurately regardless of the structure types, but the knowledge was generalized to novel items only for the locative marker. This result remained the same at the 1-week-delayed and 6-month-delayed posttests.

Following Robinson (2002a), Grey et al. (2014) also conducted a study examining to what extent incidental exposure to Japlish results in durable knowledge. Thirty-six undergraduate university students listened to 128 Japlish sentences containing eight word order types and three case markers. They then took GJTs and picture-word matching tasks (PWMs) for testing. The results at the immediate testing indicated that learning of those two types of target structure was only evidenced on

the GJTs. Interestingly however, learning of those structures was observed for the PWMs at the delayed testing while the accuracy on the GJTs was maintained. Thus, not only was the knowledge acquired during the exposure session durable across the time span, but also it improved after a delay with no exposure and no training (cf. performance on PWMs). This result is consistent with several studies of implicit learning of L2 that similarly have found that learners' performance rather improves (or increases) after a certain delay (Ellis, Loewen, & Erlam, 2006; Li, 2010; Mackey, 1999; Mackey & Goo, 2007; Morgan-Short & Bowden, 2006; Morgan-Short, Finger, Grey, & Ullman, 2012). In sum, the studies above, though few in number, seem to show that a brief period of incidental exposure to target structures may result in memory traces durable enough to manifest themselves at delayed posttests. However, it must also be recognized that the studies did not clearly identify whether the retained knowledge was explicit or implicit, for which the present study can give new insight, using the carefully triangulated approaches to measurement of explicit and implicit knowledge at an immediate and a delayed posttest.

Do Individual Differences Play a Role under Incidental Conditions?

Reber (1989) first made the claim of “primacy of the implicit” (Reber, Walkenfeld, & Hernstadt, 1991, p. 888) on account of explaining whether and how the ability of implicit learning varies across humans, depending on individual difference (ID) factors. According to the claim, implicit processes are “the functional instantiations of a phylogenetically primitive system that developed before the emergence of conscious functioning” (p. 888). This means that biological substrates underlying implicit learning are evolutionarily older than those that underlie explicit

learning. Hence, Reber continues to state that implicit processes are more robust and uniformly operative to individuals than are explicit processes.

In SLA, Krashen (1982) made a similar claim to that of Reber and his colleagues. Unconscious incidental learning processes, denoted as *acquisition* in his model, were conceptualized to be insensitive to IDs in explicit cognitive processes assessed by language aptitude tests such as the Modern Language Aptitude Test (Carroll & Sapon, 1959). Since then, several SLA researchers have tested Reber (1989), Reber et al. (1991), and Krashen's claim. Robinson (2005a) replicated Reber et al.'s explicit and implicit artificial grammar (AG) learning and compared it to learning of a natural language, Samoan, under an incidental condition. In the study, 54 Japanese learners of English (very experienced) were assigned to one of three conditions: (a) learning of an AG through the explicit series solution task, (ii) learning of an AG through the implicit task, and (iii) learning of Samoan (a language completely unknown to the participants) under an incidental condition. Of particular interest to the present study is that Robinson hypothesized that, while IDs in cognitive abilities do not influence the implicit AG learning, they do influence learning of Samoan under an incidental condition. The results showed a slightly different pattern from that of Reber et al.'s. Posttest performance of the explicit group was significantly positively correlated with a language aptitude measure ($r = .38$; Language Aptitude Battery for Japanese developed by Sasaki, 1996), but performance by the implicit group was negatively related to their IQ score ($r = -.34$). Furthermore, Robinson also found that performance on one of the posttest measures (i.e., a listening GJT) by the incidental group was significantly correlated with a WM task (r

= .42; Reading-span task: Rspan). Robinson took this as the evidence that the process of natural L2 learning under incidental conditions is different from that of implicit AG learning, stating that “it could be argued that, on the whole, incidental learning entails a variety of conscious, explicit learning, rather than simply unconscious and implicit, although associative learning plays a role there too” (p. 261). Moreover, he characterized this learning as involving “the ability to process for meaning while simultaneously switching attention to form during problems in semantic processing” (Robinson 2005b, p. 55). Several SLA studies to date have followed up on the study by Robinson (2005a), and explored the extent to which learning under incidental conditions is influenced by learners’ cognitive IDs. However, those studies yielded rather mixed findings, as some obtained very similar results (e.g., Denhovska, Serratrice, & Payne, 2016; Jackson, 2014, 2016a), while others failed to do so (e.g., Brooks & Kempe, 2013; Grey, Williams, & Rebuschat, 2015; Tagarelli, Borges-Mota, & Rebuschat, 2011).

The apparent discrepancy of research findings cited above is of great interest to the present study. One hypothesis which could potentially explain the mixed findings is that the role of IDs under incidental conditions might be modulated by the actual learning processes that would have taken place during the exposure phase. Going back to the claim by DeKeyser (2012), it seems to make sense to posit that those learners for whom IDs are found to be significantly correlated, learned the target structures more explicitly, while the others for whom individual IDs did not play a role, learned the targets more implicitly. This hypothesis fits with the learning conditions employed by the studies cited above. For instance, Jackson (2014, 2016a)

investigated the learning of artificial morphological constructions under incidental conditions in which exemplars were presented in variation sets (i.e., contiguous utterances consisting of partial repetition and expressing a roughly uniform intention, Jackson, 2014, p. 10). In other words, the input was modified in order to facilitate the discovery of the structure underlying the constructions (Onnis, Waterfall, & Edelman, 2008). It has been argued that input modification (or modified input) induces noticing of target structures (Long, 1996, 2015), and in turn, noticing in Schmidt's sense (1990, 1995, 2001) is known to be related to learners' working memory capacity (WMC) (e.g., Bergsleithner, 2011; Mackey, Adams, Stafford, & Winke, 2010; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Skehan, 2002; 2016). Hence, it follows that the learning of the artificial constructions in Jackson's study should be significantly correlated with the subjects' WMC because noticing necessarily involves some level of conscious awareness on the learner's part (Schmidt, 2001).

Although previous studies seem to have provided evidence of implicit learning being impervious to ID factors, this conclusion might be confounded with the nature of the ID construct each study attempts to measure. The previous studies have exclusively focused upon ID factors, especially cognitive aspects, that are biased towards explicit psychological processes (e.g., Granena, 2013a, 2016a; Kaufman et al., 2010; Woltz, 1999, 2003). Hamrick (2015) and Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, and Wong (2014) are among the very rare innovative exceptions among the literature in the sense that the studies took account of cognitive IDs for both explicit and implicit aspects: declarative memory and procedural

memory². Specifically, in Hamrick (2015), declarative memory was operationalized as a score on LLAMA-B, a vocabulary learning task (Meara, 2005), and procedural memory was measured by the serial reaction time task (SRT: Nissen & Bullemer, 1987). Of particular interest to the present study is that recent research in cognitive psychology and SLA has provided evidence that the SRT taps into *implicit* (sequence) learning ability (e.g., Granena, 2012, 2013a, 2016a; Kaufman, Young, Gray, Jimenez, Brown, and Mackintosh, 2010; Siegelman & Frost, 2015; Suzuki, 2015; Suzuki & DeKeyser, 2017), and that results in Hamrick (2015) showed that while IDs in declarative memory predicted performance on the immediate recognition memory test, those in procedural memory (measured by the SRT) predicted performance on the delayed, but not immediate, posttest. Morgan-short et al. (2014) report exactly the same results.

In sum, it can be concluded that the literature has confirmed that learners' IDs significantly moderate learning of the target language under incidental conditions. However, the extent to which and what types of the ID factors influence the learning outcomes seems to depend on the learning processes (i.e., explicit vs. implicit) that actually took place. In particular, previous studies have employed cognitive ID measures that are exclusively predictive of explicit learning processes, while very few have included measures predictive of implicit learning. The present study contributes new insights to this literature, as it adopts four cognitive aptitudes that are hypothesized to lie on the continuum of explicit to implicit processing.

² Although the dichotomy of declarative and procedural memory (or learning) is not the same as that of explicit and implicit learning, the task used in their study (i.e., SRT) is also known as a measure of implicit sequence learning ability.

Summary

The literature confirms that incidental learning, in some cases implicit learning, is still an option for adult L2 learners in learning various aspects of language. This implies that incidental exposure to the target language can result in unconscious, implicit knowledge, of which memory traces seem to last for a long period of time. Furthermore, it was noted that both explicit and implicit learning processes can take place under incidental conditions, and that this leads to the attainment of both explicit and implicit knowledge. However, there is still a gap in our understanding of explicit and implicit learning of L2. For instance, the literature reviewed above still needs empirical studies that adopt multiple methods of assessing conscious awareness in order to triangulate and scrutinize the nature of the resulting knowledge. Moreover, it is still not clear as to whether and to what extent ID factors moderate learning of target structures under incidental conditions.

2.1.3 Limitations in Previous Studies

The traditional procedure for explicit and implicit learning research adopts a treatment-posttest-debriefing design. Subjects in a study are incidentally exposed to some stimulus domains and then tested afterward on the extent to which they have acquired knowledge of and about aspects of the stimuli. A post-experimental debriefing session then ensues, during which the subjects are interviewed and/or given a questionnaire that probes the level of awareness they possess with regard to the acquired knowledge. With this design, existing studies have concluded that subjects develop both explicit and implicit knowledge from incidental exposure (e.g., Grey, Williams, & Rebuschat, 2014; Hamrick, 2013; Hamrick & Rebuschat, 2013;

Rebuschat, 2008, experiment 6; Tagarelli, Borges-Mota, & Rebuschat, 2011), and that the acquisition of the implicit knowledge is, at least to some extent, the result of implicit learning. However, several methodological limitations in these studies must be pointed out, for they could potentially jeopardize the internal and external validity of the studies.

Firstly, almost all of the laboratory studies of implicit learning only exposed their subjects to the target language very briefly (but see DeKeyser, 1995, for an exception). The exposure phase normally takes place for 20-30 minutes in most of the studies, which, given the significant amount of time and input that implicit learning requires, is far from ideal (e.g., Ellis, 1993, 1994; DeKeyser, 2003). Furthermore, the exposure in a short period of time also results in an unnatural, extreme intensity of exposure to target structures, which, in normal L2 learning settings, is randomly distributed over an extended period of time. Language learning, according to the usage-based perspective, is an incremental piecemeal process (e.g., Ellis, 1996, 2006; Ellis, Römer, & O'Donnell, 2016).

Secondly, many recent studies of implicit learning in SLA employ rather conservative measures of awareness. In particular, one approach taken by such studies is to rely on retrospective verbal reports, subjective confidence ratings, and source attributions (Dienes, 2008, 2012; Dienes & Scott, 2005; Rebuschat, 2013). Motivation behind the use of these subjective measures has resulted from a criticism leveled at the use of more objective tests of awareness. Reingold and Merikle (1988, 1999) first argued that objective tests are inadequate and insensitive measures of conscious awareness in that they lack exclusivity. For instance, R. Ellis (2005), in his

seminal validation study of explicit and implicit knowledge measures, acknowledges that the measures at issue such as GJTs and elicited imitations are not process-pure, and thus “it would be impossible to construct tasks that would provide pure measures of the two types of knowledge” (p. 153). Given this reasoning, more sensitive, subjective measures are sometimes claimed to be more preferable measures of awareness. In the most typical case, subjects are tested on a single outcome measure, mostly GJTs, and confidence rating and source attribution indices are collected for each item response (for a review of the subjective measures of awareness, see Rebuschat, 2008, 2013).

However, a problem can be identified with regard to this methodology: the sole reliance on GJTs complemented with the subjective confidence ratings and source attributions is rather problematic and inconsistent with results of recent validation studies of explicit and implicit knowledge measures. Studies by Suzuki (2015, 2017), Suzuki and DeKeyser (2015), and Vafaei, Suzuki, and Kachinske (2016) show that GJTs, even with a timed condition, do not really tap into implicit L2 knowledge. Therefore, it seems rather inconsistent to examine the extent to which subjects draw on implicit knowledge for an explicit knowledge measure (e.g., GJTs) by using the subjective measures of awareness. In order to overcome this issue, the present study adopted two objective outcome measures that involved either controlled-explicit or automatic and possibly, implicit processing of L2 knowledge, complemented with the confidence rating and source attribution techniques and retrospective verbal reports. This is in fact in line with the recent call for including multiple measures of conscious awareness and L2 knowledge (e.g., Hama and Leow,

2010; Rebuschat, Hamrick, Riestenberg, Sachs, & Ziegler, 2015). More to the point, this methodology should allow us to capture more reliably the extent to which brief incidental exposure results in explicit and/or implicit knowledge, which is beyond the mere understanding of whether or not subjects are aware of the knowledge they have acquired. Révész (2012), for instance, employed three different types of L2 outcome measures to further understand the effect of learning from recasts. The results of many-facet Rasch measurement and correlation analyses indicated that recasts most contributed to gains in performance on oral production tests, followed by gains in written production tests to a lesser extent, and finally an untimed GJT least of all. Indeed, this differential effect of recasts on different tasks exemplifies the importance of the adoption of multiple outcome measures to understand the multidimensional and highly complex nature of participants' developing L2 systems and the effects of instructional treatments on this development (Norris & Ortega, 2003).

Two outcome measures utilized in the present study are an untimed auditory grammaticality judgment task (AGJT) and a word-monitoring task (WMT). Untimed GJTs, whether auditory or visual, have been suggested to tap into subjects' explicit knowledge of a target language (Bowles, 2011; Ellis, 2005; Ellis & Loewen, 2007; Gutiérrez, 2013; Suzuki, 2017; Vafaei, Suzuki, & Kachinske, 2016). In particular, its untimed nature makes it more likely that subjects monitor the processing of testing stimuli, allowing controlled use of L2 knowledge. Furthermore, as Suzuki and DeKeyser (2017) point out, instructions of the task necessarily orient the subjects to focus on accuracy and form of the language in question. Thus, even though it does

not exclude the use of implicit knowledge, there is no doubt that it primarily measures subjects' explicit knowledge of L2.

In contrast to the AGJT, which is accuracy- and form-focused, the WMT involves more online processing of grammatical structures, which orients the subjects to the processing of meaning. Recent validation studies of implicit L2 knowledge measures suggest that it is a promising measure of implicit knowledge (Granena, 2012, 2013a; Suzuki, 2015, 2017; Suzuki & DeKeyser, 2015, 2017; Vafaei, Suzuki, & Kachinske, 2016). The newest study by Suzuki (2017) conducted a confirmatory factor analysis and multitrait-multimethod analysis of six measures of L2 knowledge (three hypothesized implicit and three hypothesized explicit measures): an eye-tracking visual-world paradigm task (Eye), word monitor task (WMT), self-paced reading task (SPR), timed auditory and visual grammaticality judgment task (T-AGJT; T-VGJT), and the timed SPOT task (T-SPOT)³. Results of the two analyses showed that for those with a longer length of residence (who can be expected to have developed both implicit and explicit knowledge of the target structures), the two-factor model including “implicit knowledge” and “automatized explicit knowledge” better explained the participants' performance than the one-factor model including only one construct, “language knowledge”. Suzuki also found that Eye Tracking, WMT, and SPR loaded onto the same factor labeled, “implicit knowledge”, whereas T-AGJT, T-VGJT, and T-SPOT loaded onto the other factor, “automatized explicit

³ Simple Performance-Oriented Test (SPOT) is a written fill-in blank test which explicitly asks learners to fill in blanks with target grammatical structures in written sentences.

knowledge”. Thus, in addition to the previous validation studies of explicit and implicit knowledge measures, which approached their analysis exploratorily (e.g., Ellis, 2005), there is now also evidence from research that tested the validity of outcome measures with the confirmatory approach. Reflecting such research results, the present study adopts a WMT as a measure of implicit knowledge.

2.1.4 Use of Artificial and Semi-Artificial Languages in SLA Research

In research explicit and implicit learning in SLA, the use of artificial linguistic systems has been prevalent, for they provide researchers control of linguistic factors known to be crucial for learning of a L2. These factors include prior knowledge about the target language subjects bring to the experiment and also the number of exemplars they had been exposed to prior to the experiment. Reflecting this advantage of greater control over linguistic materials, several researchers have strongly advocated the use of artificial languages (e.g., Ellis & Schmidt, 1998; Hulstijn, 1989). Semi-artificial languages, in particular, are of increasing popularity these days. Since vocabulary items in semi-artificial languages typically utilize subjects’ L1 lexis, it obviates the need for learning phases in which the researchers pre-train their subjects on the lexical items. This methodological advantage greatly reduces the burden of workload on the researchers’ side, and allows for the possibility that an entire experiment can be carried out in a short period of time – in some cases, just a single session (Rogers, 2015; Rogers, Révész, & Rebuschat, 2015).

However, in spite of the aforementioned usefulness of semi-artificial languages, limitations on and caveats about their use (and the use of artificial

languages in general) must also be explicitly acknowledged. First, many researchers have already voiced concerns about the extent to which results from artificial language studies can be generalized to learning of natural languages. These include the aforementioned small and insufficient amount of exposure during a training phase and the resulting unnatural intensity of exposure to target structures. Furthermore, some have pointed out that complexity of artificial languages and natural languages differ, thus jeopardizing the external validity of the studies (e.g., DeKeyser, 1994; Ellis, 1999; McLaughlin, 1980). Secondly, some SLA researchers contend that the use of semi-artificial languages entails an additional validity issue. It is that the conditions of learning semi-artificial languages might differ from those of natural L2 learning due to the increased saliency (DeKeyser, 2016; Godfroid, 2016). Since in an artificial language morphological targets are attached to the ending of L1 words (e.g., Grey, Williams, & Rebuschat, 2014, 2015; Williams & Kuribara, 2008), or unknown words with target morphology are mixed in L1 sentences (e.g., Hulstijn, 1989; Rogers, 2016; Rogers, Révész, Rebuschat, 2015), those unfamiliar targets become potentially more salient than they would otherwise be in a completely new language. Hence, this could trigger selective attention to such forms, which eventually leads to noticing, jeopardizing implicit processes assumed to be taking place. For instance, Rogers, Révész, & Rebuschat (2016) report that *all* of their participants at least noticed morphological inflections at the end of foreign words during a training exposure. Furthermore, subjects in experiments (even of implicit learning) normally have some intention to learn and some conscious awareness of learning, even though they might end up learning something they did not intend to learn in the first place

(DeKeyser, 2003). Such intention and awareness should be all the more apparent for conditions in which they are exposed to artificial languages.

To return to the present study, a semi-artificial language, *Japlish*, was adopted from Grey, Rebuschat, and Williams (2014, 2015) and Williams and Kuribara (2008) as the target of the experiment (for specific structures, see *The Language*). While admittedly all of the aforementioned limitations and caveats also apply to the current study, two reasons justify the use of Japlish here. First, a recent study by Godfroid (2016) successfully replicated results of the implicit L2 learning studies, using vowel alterations in strong German verbs. In other words, there is now evidence, though only a single study, that results of implicit learning studies with artificial linguistic systems can be generalized to settings of natural L2 learning (but see DeKeyser, 2016, for a criticism of participants' pre-existing experience with the language and structures similar to the target). Second, the use of a semi-artificial language is better suited for the aim of the current experiment. The present study examines how explicit and implicit cognitive aptitudes moderate learning of simple and complex word orders in Japlish. The concepts of language complexity and difficulty are known to be particularly difficult to define and operationalize, depending on miscellaneous factors, such as linguistic, context, and learner characteristics (DeKeyser, 2005, 2016; Housen & Simone, 2016). By using a semi-artificial language, several of the linguistic factors can be readily controlled and thus manipulated, including frequency, familiarity, and saliency of target features, which is difficult to do in natural languages. Therefore, a decision was made that the study would take into account the

criticisms toward the use of artificial languages, but still employ them for their many methodological advantages.

Finally, research questions and hypotheses of the present study were motivated by insight from the previous studies of implicit L2 learning that the salience (whether or not it is increased by the artificiality of a semi-artificial language) of language structures in question might interact with their complexity. For instance, in Grey, Williams, and Rebuschat (2014), 23 out of 34 were able to state the correct word order rule for simple sentence types (i.e., S-O-V, O-S-V, S-I-O-V, and O-S-I-V) but very few even tried for complex sentence types (i.e., S-[S-O-V]-V, O-S-[S-V]-V, S-[S-I-O-V]-V, and O-S-[S-I-V]-V). It might be that if target structures are complex enough, subjects may remain unaware that they are learning them. This coincides with what Reber (1989, 1993) explained in his study, “a rich and complex stimulus domain is a prerequisite for the occurrence of implicit learning” (1989, p.220). Thus, it seems that for targets to be learned implicitly, they need to be complex so that subjects cannot encode them explicitly.

2.2 Cognitive Aptitudes

2.2.1 Theoretical Underpinnings

Language aptitude has been understood as a special talent for language learning, with John Carroll (1973, 1981, 1993) first defining it as a composite of cognitive and perceptual abilities that facilitate learning of a language. Through a series of factor-analytic studies, Carroll identified four learning abilities underlying the construct: (a) phonetic coding ability, (b) grammatical sensitivity, (c) inductive

language learning ability, and (d) associative memory, all of which together are conceptualized as constituting a latent variable, namely, language aptitude. Based on this model, he developed the Modern Language Aptitude Test (MLAT) with Stanley Sapon (Carroll & Sapon, 1959), and this traditional model and battery of tests have been popular in use. At least five characteristics of aptitude measured by the MLAT have been originally postulated, some of which have not been confirmed by current SLA research. First, the set of abilities in Carroll's framework are thought to be distinct from other individual difference factors (e.g., motivation, anxiety, and personality). Second, it is fairly stable across time and impervious to external influences (e.g., training). Third, it predicts the rate and ease of learning, rather than the ultimate attainment of L2 ability. Fourth, it is relevant in instructed rather than in naturalistic settings. Fifth, it is predictive at initial stages of learning, but the extent to which it affects advanced L2 learners is unknown.

The construct of aptitude has been subject to many empirical studies to date, and most of its parts have undergone some extent of refutations and changes. For instance, Skehan (2002) pointed out that two of the underlying abilities of the aptitude construct, namely grammatical sensitivity and inductive language learning ability, share in common the analytic aspects of learning, uniting and relabeling them as *language analytic ability*. Furthermore, another example is that a body of research has shown that aptitude is not only predictive of the rate of learning in instructed settings but also of the ultimate attainment in naturalistic settings (e.g., Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000; Granena, 2016b; Granena & Long, 2013). Some

alleged methodological issues notwithstanding (e.g., Long, 2013), language aptitude is known to play a compensatory role for a late start of L2 learning.

One of the long-lasting issues in the conceptualization of language aptitude is what individual difference factors are to be included. Originally, language aptitude was conceptualized as a unitary domain-specific construct comprising a set of abilities (Carroll, 1973). However, as Li (2016) points out, the construct has been characterized rather differently in different disciplines. For instance, Snow (1992) defines aptitude as something that is a composite of more dynamically interacting individual characteristics, as opposed to Carroll's more cognitively oriented approach, thereby necessitating inclusion of affective and contextual factors (e.g., motivation). Furthermore, reflecting the current understanding of its contribution to L2 learning (Linck et al., 2014), the inclusion of working memory (WM) capacity is also advocated by many L2 researchers (e.g., DeKeyser & Koeth, 2011; Kormos, 2013; Miyake & Friedman, 1998; Sawyer & Ranta, 2001; Wen & Skehan, 2011). A recent review of language aptitude by Wen, Biedroń, & Skehan (2017) even suggests that WM capacity might be the central component of language aptitude constructs. This clearly reflects the recent call for reconceptualizing aptitude from a more multifaceted perspective (e.g., Skehan, 2002, 2012), referring to 'aptitudes' rather than 'aptitude', which comprise a set of domain-general learning mechanisms and abilities. The present study focuses on cognitive aptitudes, which are defined as individual differences in cognitive and perceptual abilities that predispose individuals to learn a language quickly and easily.

2.2.2 Cognitive Aptitudes in SLA

At the time of Carroll's work, cognitive mechanisms and processes relevant in L2 learning were largely yet to be discovered. However, as theoretical and empirical understandings of psychological constructs underlying L2 learning have substantially progressed, the construct of language aptitude should also evolve by incorporating those understandings (Robinson, 2012). Furthermore, the MLAT was developed at a time when audiolingualism was the predominant language teaching methodology in society. Thus, its components are hypothesized to predict success in such a learning environment. In order to reflect today's communicative, task-based, or immersion classrooms, a new construct of language aptitude which captures cognitive abilities required for today's classroom learning is necessary. As a result, two new models of language aptitude that are informed by attested findings of SLA research and L2 pedagogy have been proposed. First, Skehan (1998, 2002, 2016) proposed a model of language aptitude that is based on developments of accumulating SLA research. The model connects nine SLA processing stages to putative components of aptitude. For instance, 'noticing', one of the L2 cognitive processes in the model, has been recognized as crucial in L2 learning (e.g., Schmidt, 2001). However, it has also been known to occur variably from individual to individual. As Dörnyei and Skehan (2003) explain, this would be in part due to individual differences in noticing abilities, which are, in the model, identified as working memory capacity and phonetic coding ability (as argued in Sawyer & Ranta, 2001). In this way, Skehan argues that aptitude can be used for an approach that serves to explain the underlying causes of individual differences in learning outcomes through relating major SLA

processing to aptitude components, rather than using them merely for their predictive value (see, Dörnyei & Skehan, 2003; Skehan, 2002, 2012, 2015, 2016; Wen, Biedroń, & Skehan, 2017 for a review of Skehan's processing model of language aptitude).

Second, Robinson (2002a, 2005b, 2007, 2012) proposed a more pedagogically oriented model of language aptitude. This model is based on two hypotheses based on the interactionist approach to aptitude: the Aptitude Complex Hypothesis and Ability Differentiation Hypothesis. The Aptitude Complex Hypothesis is based on an idea from Richard Snow (1987, 1994) and claims that a set of primary cognitive abilities (e.g., phonological working memory, perceptual speed) can be combined to form higher-order aptitude complexes, where each two sets of which are drawn on, in a particular instructional or input exposure condition. The Ability Differentiation Hypothesis claims that individual L2 learners vary in their strength on particular aptitude complexes, creating a unique aptitude profile specific to the individuals. For instance, learning from recasts is hypothesized to draw on two aptitude complexes, noticing the gap (NTG), and memory for contingent speech (MCS). NTG, in turn, consists of two primary (but measurable) cognitive abilities, namely, perceptual speed and sound-symbol correspondence. Similarly, MCS can also be further broken down into two different abilities, phonological working memory (PWM) capacity and speed of PWM. In Robinson's account, a learner can either be high in both aptitude complexes, high in one but low in the other, or low in both. Thus, the model offers a new perspective of language aptitude, sub-sets of which interact with a range of basic instructional conditions, to which learners are optimally matched according to their strengths and weaknesses shown by the aptitude profile.

In sum, these two contemporary models of language aptitude share a common feature that seeks to provide an explanation of differential final learning outcomes by individuals (Wen, Biedroń, & Skehan, 2017) instead of trying to predict them. The current study adopts this approach and tries to explain learning processes under an incidental exposure condition using currently available cognitive aptitude measures. Fortunately, there is an increasing number of studies exploring new measures of cognitive aptitudes (e.g., Granena, 2012, 2013a, b, c, 2016a; Kaufman et al., 2010; Linck et al., 2013; Wen & Skehan, 2011; Woltz, 1999, 2003). The development of new aptitude measures and constructs is partly a response to the criticism made by some that the traditional language aptitude constructs are extremely biased towards explicit cognitive processes rather than implicit processes (e.g., Granena, 2013a, b, c, 2016a). Measures of our types of cognitive aptitudes that are hypothesized to involve a continuum of explicit processing and implicit processing are employed: (a) language analytic ability, (b) working memory, (c) phonological short-term memory, and (d) implicit sequence-learning ability. They are reviewed one by one below, with some evidence of their association with explicit and implicit learning processes.

2.2.3 Explicit and Implicit Cognitive Aptitudes

Language Analytic Ability

Language analytic ability (LAA) is defined as “the capacity to infer rules of language and make linguistic generalizations or extrapolations” (Skehan, 1998, p. 207), which is represented by grammatical sensitivity and inductive learning ability in Carroll’s model. Measures of LAA include the Words-in-Sentences subtest of the

MLAT (Carroll & Sapon, 1959), the Language Analysis subtest of the Pimsleur Language Aptitude Battery (Pimsleur, 1966), and the LLAMA- F subtest of LLAMA language aptitude test (Meara, 2005). According to Skehan's model, LAA is utilized for pattern identification, complexification, and handling of feedback. Similarly, in Robinson's model, grammatical sensitivity (a type of LAA) and rote memory combine to form a higher-order aptitude complex, called 'metalinguistic rule rehearsal', which is hypothesized to facilitate explicit rule learning. Furthermore, in a recent literature review, Skehan (2016) further categorized this ability as an explicit cognitive aptitude. This categorization seems reasonable, based on results of SLA studies that investigated links between a set of cognitive aptitudes and learning under particular conditions. For instance, DeKeyser (2000) examined how verbal analytical ability of Hungarian child and adult immigrants predict their ultimate attainment in English morphosyntax. The aptitude score measured by a Hungarian version of the MLAT subtest, Words-in-Sentences, was found to be a significant predictor of performance scores on GJTs by the adult acquirers, but no such relationship was found for child acquirers. DeKeyser took this as the evidence that child and adult L2 acquirers employ different learning mechanisms, namely implicit domain-specific mechanisms and explicit domain-general problem-solving capacities, respectively (as claimed by Bley-Vroman, 1989). This finding is in line with other studies of the role of LAA in immersion settings, such as Harley and Hart (1997) who found that LAA is related to late-immersion rather than early-immersion learners.

Another line of experimental studies that explore the role of LAA focus on the relationship between learner aptitude profiles and acquisition of L2 forms from

corrective feedback. Those studies uniformly show that LAA moderates the effectiveness of explicit and/or metalinguistic corrections (e.g., Ranta, 2002; Sheen, 2007; Trofimovich, Ammar, & Gatbonton, 2007; Yilmaz, 2013; Yilmaz & Granena, 2016a). For instance, Sheen (2007) investigated how LAA moderated the extent to which learners benefit from two corrective feedback types, recasts and metalinguistic corrections. The results found a strong correlation between LAA and performance by learners in metalinguistic correction group both at the immediate- and delayed-posttesting, while no such finding was obtained for the recast group. Yilmaz (2013) and Yilmaz and Granena (2016a) found similar results.

Reflecting the models and the results from the experimental studies, it seems valid to treat LAA as an explicit cognitive aptitude. The current study follows Skehan (2016)'s categorization and assumes that if participants' performance correlates with their LAA, it must be the case that they engaged in explicit cognitive processes (and possibly with metalinguistic focus) while they were incidentally exposed to training materials. Ranta (2005)'s characterization of LAA fits with this assumption as she describes LAA as "the ability of an individual to focus on the structural properties of linguistic utterances rather than on their meaning" (p. 101).

Working Memory

Working memory (WM), since its introduction by Baddeley and Hitch (1974; also Baddeley, 2007, 2012), has been conceptualized in different ways in different models (e.g., Cowan, 1999; Engle, Kane, & Tuholski, 1999; for a review, see Miyake & Shah, 1999). In the current study, WM is defined as a system of dual functions, temporary storage and attentional control (Baddeley, 2007). An important distinction

to be made here based on the definition is between WM in a broad sense and phonological short-term memory (PSTM). In the present investigation, WM is measured with a complex span task (e.g., a listening span task), which involves both storage and processing phonological information through sounds, whereas PSTM is measured with a simple repetition task (e.g., a nonword repetition), which only involves the storage of information.

In L2 learning, it has been attested that WM plays a significant role (e.g., Linck et al., 2014; also, Juffs & Harrington, 2011; Williams, 2012 for a review) at almost all stages of learning. For instance, reflecting this contribution of WM, Skehan (2016) proposes the centrality of WM in the language aptitude construct and argues that it is involved in all stages of SLA processing except the last stage, lexicalization. In particular, a number of SLA research studies to date have found that WM is important in noticing (e.g., Bergsleithner, 2011; Mackey, Adams, Stafford, & Winke, 2010; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Skehan, 2002; 2016). However, as reviewed above, it is not clear whether WM enhances learning under incidental conditions (e.g., Jackson, 2016a; Robinson 2005a, vs. Brooks & Kempe, 2013; Tagarelli, Borges-Mota, & Rebuschat, 2011). Robinson (2002a, 2005, 2007, 2012), for instance, proposes in his model that incidental learning from oral input containing a flood of particular forms requires two types of aptitude complexes, memory for contingent speech (MCS) and deep semantic processing (DSP). MCS can be further subdivided into two primary cognitive abilities, phonological WM (PWM) and its speed, and DSP can similarly be divided into two, analogies and inference of word meaning. The present study focuses on subjects' WM capacity and employs a

listening span task to measure it. This test has been proposed as a suitable measure of one of the primary abilities contributing to MCS (Robinson, 2012).

WM is also considered as an explicit cognitive aptitude, as is LAA, and related to conscious processing (Çeçen & Erçetin, 2016; Paradis, 2009; Skehan, 2016). Support for this also comes from experimental studies of SLA, for instance, Robinson (2005a). Although WM capacity can be categorized as a type of explicit cognitive aptitude, it is associated with a different level of awareness than that of LAA. As described, LAA is related to awareness at the level of understanding due to its relationship with metalinguistic awareness. Several SLA studies of WM however provide evidence that WM is not closely related to metalinguistic awareness (as cited in Jackson, 2016b). For instance, Robinson (2002b) found no association between subjects' WM capacity (measured by a reading span task) and written untimed GJTs. Bell's (2009) study fits with Robinson's finding, as she found that inductive learning ability (i.e., a subset type of LAA) predicted metalinguistic awareness of target features, whereas WM capacity did not. Also, Roehr & Gánem-Gutiérrez (2009) did not find a relationship between subjects' WM capacity and their metalinguistic knowledge, concluding that LAA entails "a higher-level mental faculty, such as analytic reasoning about language", whereas WM involves "a lower-level and thus more generic mental faculty like online storage and processing of linguistic information" (p. 175-176). Thus, it seems that having a high WM capacity does not necessarily enhance acquisition and maintenance of awareness at the level of understanding.

However, this is by no means to argue that WM is not involved with implicit cognitive processes. Although the results of the past SLA literature are mixed as to whether learners' WMC is related to learning under incidental/implicit conditions (see, 2.1.2), researchers in experimental and cognitive psychology have come to a consensus that there is an implicit aspect to WM processes and components (Hassin, Bargh, Engell, & McCulloh, 2009; Janacsek & Nemeth, 2013, 2015; Martini, Sachse, Furtner, & Gaschler, 2015; Soto & Silvanto, 2014). For instance, Hassin, Bargh, Engell, and McCulloh (2009) conducted five experiments in which they showed that three processes underlying WM (i.e., maintenance of ordered information for short periods time, context-relevant updating of information and goal-relevant computations involving representations, and rapid biasing of task-relevant cognitions and behaviors) can operate outside of conscious awareness. Therefore, it seems valid to categorize WM as involving both explicit and implicit processing but the role of the former dominantly outweighs the latter. This fits with the results of the SLA studies that the relationship between working memory and L2 learning consistently emerges when learning is explicit and intentional (cf. Janacsek & Nemeth, 2013, for the same observation in the field of experimental psychology).

Phonological Short-Term Memory

In the present study, PSTM refers to the phonological storage component of the entire WM system. In Baddeley's multi-component model (Baddeley, 2007), PSTM is represented by the phonological loop, a domain-specific slave system which comprises a phonological store and an articulatory control (rehearsal) process. To date, SLA studies have demonstrated that a better PSTM capacity facilitates L2

vocabulary learning (Baddeley, Gathercole, & Papagno, 1998; Baddeley, Papagno, & Valler, 1988; Martin & Ellis, 2012; Papagno, Valentine, & Baddeley, 1991), grammar learning (Ellis & Sinclair, 1996; Kormos & Sáfár, 2008; Martin & Ellis, 2012; Service, 1992; Service & Kohonen, 1995), and overall L2 fluency development (French & O'Brien, 2008). However, while WM capacity has been implicated frequently in L2 learning processes, there are few theoretical accounts that address the role of PSTM per se. For instance, neither Skehan nor Robinson include PSTM in their SLA model of aptitude.

From a usage-based perspective, N. Ellis (1996) argues that language learning is fundamentally a form of sequence learning that results in a stock of lexical sequences from which learners can abstract grammatical and distributional regularities. In his account, this abstraction process takes place largely in the form of an implicit positional analysis of distributional information, and PSTM is the system that determines its success, because “[r]epetition of sequences in PSTM allows their consolidation in phonological LTM” (p. 108), upon which representation in the L2 system will be tuned by frequency of exposure. Thus, IDs in PSTM affect the quality and quantity of sequences that are held in the phonological loop, and this eventually affects the quality and quantity of language learning too. As an example, Williams and Lovatt (2003) conducted an experiment within this framework. They examined the relationship between subjects’ PSTM capacity and the ability to generalize determiner-noun gender agreement rules, to which they were incidentally exposed. Their results indicated a strong correlation between the PSTM assessed by an immediate serial recall test and subjects’ performance on generalization tests.

Furthermore, the result was still consistent even “when generalization test performance depended on an abstract categorization of nouns into word groups, as opposed to direct associations between morphemes that occurred together in training items” (p. 106). They thus concluded that PSTM is indeed involved in the acquisition of very abstract aspects of grammar.

However, even though the usage-based approach proposes the role of PSTM in implicit cognitive processes described above, it is still not clear from an empirical perspective whether PSTM alone constitutes an explicit or implicit cognitive aptitude. Skehan (2016) categorizes it as “[n]ot clearly explicit or implicit” (p. 31), suggesting the role of PSTM in both explicit and implicit processes. Also, in his exploration of implicit and explicit aptitudes and knowledge measures, Suzuki (2015) claimed that PSTM contributes to both explicit and implicit inductive learning. However, the result failed to find any substantial relationship between PSTM and acquisition of automatized explicit knowledge and implicit knowledge, from which he concluded that, “this may mean that more high-order cognitive aptitudes, language inductive analytic and probabilistic sequence learning ability, play more important roles in grammar learning” (p. 156). As do Skehan (2016) and Suzuki (2015), the present study argues that PSTM is involved in both explicit and implicit cognitive processes. However, compared to WMC, the present study also claims that PSTM is relatively more relevant to automatic and implicit processes of language learning and less to explicit aspects of it. Support for this categorization can be seen in an aptitude study by Harley and Hart (1997), who found that memory ability significantly predicted

performance of early immersion learners (i.e., starting from Grade 1) whereas analytic verbal ability was predictive for late learners (i.e., starting from Grade 7).

Implicit Sequence-Learning Ability

Implicit sequence-learning ability is oftentimes measured by the serial reaction time (SRT) task originally developed by Nissen and Bullemer (1987). More recent studies in cognitive psychology have repeatedly demonstrated that SRT entails implicit learning and acquisition of implicit knowledge (e.g., Destrebecqz & Cleeremans, 2001; Jiménez, Vaquero, & Lupiáñez, 2006; Kaufman et al., 2010). In the field of SLA, too, calls have been made for incorporating implicit cognitive aptitudes, including the implicit sequence-learning ability measured by SRT (e.g., Doughty et al., 2010; Granena, 2013a, b, c; Linck et al., 2013). Answering this call, several SLA researchers have tested experimentally to what extent implicit cognitive aptitudes relate to L2 learning. For instance, Granena (2012) investigated the extent to which explicit and implicit cognitive aptitudes (i.e., LLAMA-D: a sound recognition task, and SRT) moderate ultimate L2 attainment of early and late naturalistic L2 acquirers. Of particular interest to the present study is that Granena used a series of language attainment measures reflecting a continuum of controlled and automatic use of L2 knowledge. The results showed that the implicit aptitudes significantly predicted the L2 attainment of both early and late learners on measures that require automatic processing of L2 knowledge. On the other hand, explicit aptitudes were found to predict outcomes by both early and late learners indexed by measures of controlled L2 knowledge use. These results yield two implications that are particularly relevant to the current study. First, they suggest that adult L2 learners as

well as child learners can engage in implicit learning processes, which is consistent with results of implicit learning research reviewed above. Second, different types of cognitive aptitudes interact with the modality and type of L2 outcome measures employed. Specifically, Granena (2012; also, 2013c) showed that the implicit type of aptitudes was related to performance on measures that require automatic processing of L2 knowledge, whereas the explicit type of aptitudes was related to measures of controlled processing of L2 knowledge. Suzuki (2015) also reports such association of aptitude between explicit and implicit cognitive aptitudes and types of L2 outcome measures (i.e, explicit and implicit knowledge measures).

2.3 Gaps in the Literature

The review of the literature above has unearthed several gaps in our understanding of explicit and implicit learning of L2 under incidental conditions. First, there are few studies that compared various methodological approaches to acquisition of explicit and implicit knowledge. One rare exception is Rebuschat, Hamrick, Riestenberg, Sachs, and Ziegler (2015) who triangulated measures of conscious awareness through concurrent verbal reports (i.e., think-aloud protocols), retrospective verbal reports, and subjective measures. However, there is currently no study that adopted multiple objective measures of L2 knowledge including a WMT. At the same time, the results in the previous studies are inconsistent as to the extent and degree to which incidental exposure to a language resulted in explicit and/or implicit knowledge (e.g., Hama & Leow, 2010; Rebuschat & Williams, 2012; Williams, 2005). Thus, it should be instructive to the field of SLA (and also Instructed SLA) to examine the extent to which those outcome measures would

homogeneously and/or heterogeneously index L2 learners' acquisition of explicit and implicit knowledge and conscious awareness associated with them. One of the issues raised against the use of the subjective measures of conscious awareness was that they were rather conservative. Thus, it is wise for us to ask a question such as what happens if those measures of awareness are pitted against the objective outcome measures of explicit and implicit knowledge, as it can inform us about the complex and multifaceted nature of learner's developing L2 system.

Second, despite the fact that the literature on explicit and implicit learning is productive in investigating how ID variables moderate learning of target structures under incidental conditions, there is currently no study that draws upon the measures of both explicit and implicit cognitive aptitudes. Thus, the present study is novel in this domain of inquiry in that it adopts four explicit and implicit cognitive aptitude measures that tap into domain-general cognitive abilities hypothesized to lie on a continuum of explicit and implicit cognitive processing. Although there are a few exceptions that already examined how explicit and implicit cognitive aptitudes moderate learning outcomes (Brooks, Kwoka, & Kempe, 2017; Granena, 2012; Suzuki, 2015), none of them looked at explicit and implicit learning of L2 under incidental conditions.

Lastly, the present study is novel in that it investigates explicit and implicit learning of different types of linguistic constructions under an incidental condition and how learning of these constructions interacts with explicit and implicit L2 outcome measures as well as cognitive aptitudes. Thus, the purpose of the study is to understand and decipher the complexity of the developing L2 knowledge system

resulting from incidental exposure and to contribute to understanding the roles of explicit and implicit learning of L2 under incidental conditions.

Chapter 3: Research Questions and Hypotheses

Taking into account the gaps in the literature identified in the previous chapter, a laboratory experiment was conducted with a randomized design during which participants assigned to either the experimental group or the control group were briefly exposed to exemplar sentences of a semi-artificial language, *Japlish*. The participants were subsequently tested on the extent to which they learned aspects of the target constructions explicitly and/or implicitly. The experimental and control groups differed in that while the former group was exposed to Japlish sentences, participants in the latter were trained with sentences whose word order and position of case markers were pseudo-randomized (see 4.3 for pseudo-randomization of word orders and case markers). The entire experiment took place with a treatment-posttest-debriefing design. Moreover, two outcome measures, an untimed auditory GJT (AGJT) and a word-monitoring task (WMT), were employed at both immediate and delayed testing. Explicit as well as implicit cognitive aptitude measures were administered. Based on the current understanding of explicit and implicit learning of language, methodological issues of the previous studies, and relationships between cognitive aptitudes and outcome measures, eight research questions were formulated. The hypotheses and their rationales follow:

- Research Question 1: Are adult L2 learners able to learn the *simple* word order types of constructions in Japlish, as measured by two different types of L2 outcome measures at an immediate posttest: AGJT, a measure which allows for controlled use of acquired knowledge, and WMT, a measure which requires automatic use of acquired knowledge?

Hypothesis 1a: For the simple word order types, the experimental group will significantly outperform the control group on the AGJT (reasoning below).

Hypothesis 1b: For the simple word order types, subjective measures of awareness (i.e., confidence ratings and source attributions) on the AGJT will show that those who are aware of the knowledge they have acquired perform better on the task. In other words, the confidence level and knowledge base that participants report on each grammaticality judgment will significantly correlate with the accuracy of their performance, thus *not* satisfying the guessing criterion and zero-correlation criterion of unconscious implicit knowledge (Chan, 1992; Dienes, et al., 1995).

Hypothesis 1c: For the simple word order types, the experimental group will *not* significantly differ from the control group on the WMT.

The reasoning behind the hypotheses for Research Question 1 is the suggestion by Reber (1989, 1993) and the results by Grey, Williams, and Rebuschat (2014), which suggest that there might be a correlation between the salience of language structures in question and their complexity. Thus, it was expected that learners would notice the simple word orders, and this noticing would result in conscious awareness on the participants' part, leading to intentional and explicit learning, the result of which is explicit knowledge, only indexed by the AGJT. Furthermore, as the participants were expected to notice the simple constructions, they were also expected to draw on their explicit and conscious knowledge on the task, which was operationalized as the level of

confidence participants report on the confidence ratings and also the knowledge basis of their response, indicated by the source attributions.

- Research Question 2: Are adult L2 learners able to learn the *complex* word order types of Japlish, measured by the two different types of L2 outcome measures at an immediate posttest?

Hypothesis 2a: For the complex word order types, the experimental group will significantly outperform the control group on the AGJT.

Hypothesis 2b: For the complex word order types, subjective measures of awareness on the AGJT will show that there is *no* systematic correlation between the level of confidence and knowledge basis of their response that participants report on each grammaticality judgement and their accuracy of performance on the task. In other words, the guessing criterion and the zero-correlation criterion will be satisfied, confirming that the participants acquired unconscious, implicit knowledge.

Hypothesis 2c: For the complex word order types, the experimental group will significantly differ from the control group on the WMT.

The rationale behind these hypotheses is also based on the finding from Grey et al. (2014). Thus, it was expected that the complex types of word order constructions were complex enough for the participants to not notice those features. Thus, the participants were predicted to learn the constructions incidentally and implicitly, resulting in implicit knowledge (though not exclusive to it). As pointed out by Reingold and Merikle (1988, 1990), objective direct tests such as GJTs are not process-pure. As a result, the untimed AGJT in the present study was expected to

allow for the use of implicit knowledge along with explicit knowledge. Thus, the experimental group was hypothesized to outperform the control group not only on the WMT, but also on the AGJT.

- Research Question 3: Are adult L2 learners able to learn *morphological case markings* of Japlish, measured by the two different types of L2 outcome measures at an immediate posttest?

Hypothesis 3a: For the case markings, the experimental group will significantly outperform the control group on the AGJT.

Hypothesis 3b: For the case markings, subjective measures of awareness on the AGJT will show that those who are aware of the knowledge they have acquired perform better on the task. Thus, it is hypothesized that the confidence level and the knowledge basis of responses would correlate with the accuracy of their performance.

Hypothesis 3c: For the case markings, the experimental group will *not* significantly differ from the control group on the WMT.

The artificial nature of Japlish, especially case markers attached to participants' L1 vocabulary, will elicit their selective attention to and noticing of such features (Godfroid, 2016). Hence, explicit cognitive processes are expected to be employed during the exposure to exemplar sentences, and this is thus predicted to lead to the acquisition of explicit knowledge assessed by the untimed AGJT. It is deemed possible that participants could have developed a small extent of implicit knowledge due to the fact that they still have to process each exemplar sentence for meaning. However, the amount of exposure in the current experiment was judged to

be insufficient for them to acquire implicit knowledge reliable enough to be detected by the WMT.

- Research Question 4: Which cognitive aptitudes moderate learning of the simple word order types, measured by the two L2 outcome measures?

Hypothesis 4: The learning of the simple word order constructions indexed by the untimed AGJT will be significantly moderated by participants' WM capacity (i.e., the listening span) and their language analytic ability (i.e., LLAMA-F).

This hypothesis hinges on Hypothesis 1a-c participants will notice the simple word order types and thus primarily acquire explicit knowledge. WM is known to be related to noticing of language forms (e.g., Bergsleithner, 2011; Mackey, Adams, Stafford, & Winke, 2010; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Skehan, 2002; 2016) and LLAMA-F to metalinguistic awareness and learning of explicit rules (e.g., Ranta, 2002; Robinson, 2002a, 2005b, 2007; Sheen, 2007; Trofimovich, Ammar, & Gatbonton, 2007; Yilmaz, 2013; Yilmaz & Granena, 2016). Furthermore, Granena (2012)'s finding of an interaction between L2 outcome measure types and explicit-implicit cognitive aptitudes supports this hypothesis.⁴

- Research Question 5: Which cognitive aptitudes moderate learning of the complex word order types, measured by the two L2 outcome measures?

⁴ It was also expected that implicit learning ability and PSTM would significantly moderate participants' learning of the simple word orders measured by the WMT. However, the correlation between the aptitudes and participants' performance on this task should be small, given that Hypothesis 1 stipulates that this learning will be primarily explicit, mostly resulting in explicit knowledge, and therefore minimal variance in implicit knowledge.

Hypothesis 5: The learning of the complex word order types indexed by the AGJT as well as WMT will be significantly moderated by subjects' implicit sequence-learning ability (i.e., SRT) and PSTM capacity (i.e., the nonword repetition).

This hypothesis hinges on Hypotheses 2a-c that for learning of the complex word order constructions, implicit learning processes will operate. These two cognitive aptitudes are argued to underlie implicit sequence-learning (e.g., SRT: Destrebecqz & Cleeremans, 2001; Jiménez, Vaquero, & Lupiáñez, 2006; Kaufman et al., 2010; PSTM: Ellis, 1996; Williams, 2003, 2005). Furthermore, given the assumption that the associative sequence learning mechanism underlies word order learning (e.g., Williams, 2010), these two cognitive aptitudes should be significantly related to the learning of the complex word order types.

- Research Question 6: Which cognitive aptitudes moderate learning of the case marking types, measured by the two L2 outcome measures?

Hypothesis 6: The learning of case markings assessed by the untimed AGJT will be significantly moderated by WM capacity and LAA.

This hypothesis hinges on Hypothesis 3a-c that the experimental group will outperform the control group on the untimed AGJT, but not on the WMT. As with Hypothesis 4, the increased saliency engages the participants in explicit learning processes, and thus the learning outcome will correlate with the explicit cognitive aptitudes.

- Research Question 7: To what extent do the results in Research Question 1-3, learning of simple and complex word orders and case marking constructions, change at the delayed posttests?

As the literature on explicit and implicit learning under incidental conditions is limited in the number of experimental studies that included delayed posttests, this research question was addressed with an exploratory approach. However, one speculation could be made by drawing on those existing studies: it was expected that the experimental group would still significantly outperform the control group on the untimed AGJT for all of the target constructions. This was to reflect the results by Robinson (2002) and Grey et al. (2014) that the knowledge that adult L2 learners acquire can be durable enough to be measured at delayed posttesting. However, it is yet unclear as to whether the retained knowledge was explicit or implicit, and RQ7 exploratorily examines whether explicit and/or implicit knowledge can be maintained over two weeks of no exposure to the language, and investigates patterns in which the two types of knowledge emerge. The meta-analytic reports of the effectiveness of explicit and implicit instruction suggest that while explicit knowledge often declines, implicit knowledge is retained or rather improves over a delay period (e.g., Li, 2010; Mackey & Goo, 2007).

- Research Question 8: To what extent do the results in Research Question 4-6, the roles of explicit and implicit cognitive aptitudes in learning of the target constructions under an incidental condition, change at the delayed posttests?

As with Research Question 7, this was also addressed with an exploratory approach. However, in order to reflect the findings of Hamrick (2015) and

Morgan-Short et al. (2014), it was expected that the retention of the target constructions would be significantly moderated by subjects' implicit sequence-learning ability and PSTM capacity regardless of the complexity of the constructions.

Overall, research questions and hypotheses were formed based on the theoretical discussions and experimental studies in the literature on explicit and implicit learning and their relationship with cognitive aptitudes. It was expected that the simple nature of the constructions and the artificiality of the language would jointly contribute to facilitating noticing of the simple word order and case marking constructions. As noticing entails conscious awareness of the learning experience, it was predicted to engage the participants in explicit learning, the result of which was to be explicit knowledge indexed by the untimed AGJT and the subjective measures of awareness. On the other hand, the complexity of the complex word order constructions was hypothesized to make the forms extremely difficult to encode (based on the claim by Reber, 1989), and thus if learning was to occur, it would be predicted to be implicit learning, the result of which is implicit knowledge assessed by the WMT and the subjective measures of awareness (i.e., the guessing criterion and the zero-correlation criterion).

Chapter 4: Methods

4.1 Participants

A total of 63 native speakers of English who had no experience with Japanese nor any case-marking languages (i.e., German, Greek, Japanese, Korean, Latin, and Russian) participated in the study. They were recruited either via the University of Maryland Psychology Research Sign-Up System to receive course credits ($n = 56$) or seeing via a flyer ($n = 7$) for a participatory compensation of \$40. They were all undergraduate or graduate students at University of Maryland, majoring in subjects other than linguistics. Out of this sample, 14 participants were excluded, because they did not return for the delayed posttest ($n = 8$), did not follow the instructions ($n = 4$), or they produced a mean WMT RT larger than 2500 ms ($n = 2$). As a result, 49 participants (11 males and 38 females) constituted the final sample. Although their native language was English, they also knew one or more second languages ($M = 1.29$, $SD = \text{Min} = 0$, and $\text{Max} = 3$) at various levels of proficiency (beginner to advanced): Albanian, Arabic, ASL, Chinese, French, Hebrew, Hindi, Igbo, Italian, Spanish, Telugu, Twi, and Yiddish. At the time of the experiment, no one had been to any countries where a case-marking language is spoken for a period longer than two weeks. The mean age at the time of the study was 19.47 ($SD = 1.78$, $\text{Min} = 18$, and $\text{Max} = 27$). All of the participants were randomly assigned to either the experimental group ($N = 28$) or the control group ($N = 21$) at the beginning of the experiment.

4.2 Materials

4.2.1 Exposure Task Materials

As the target to be learned, the present study utilized a semi-artificial language, Japlish. The language, originally used by Williams and Kuribara (2008), consists of English lexis and Japanese syntax and case markers. In the present study, four word order patterns, OSV, OSIV, OSSVV, and OSSIVV, and three case markers, -ga (subject marker), -o (direct object marker), -ni (indirect object marker), were specifically chosen as target constructions to be learned, as in the examples provided below⁵:

1. O-S-V (simple)
That wall-o Mary-ga painted.
“Mary painted that wall.”
2. O-S-I-V (simple)
The picture-o John-ga his friends-ni sent
“John sent the picture to his friends.”
3. O-S-[S-V]-V (complex)
The tuition-o Mary-ga her school-ga raised said
“Mary said that her school raised the tuition.”
4. O-S-[S-I-V]-V (complex)
Those documents-o John-ga his workmate-ga their boss-ni faxed told.
“John told that his workmate faxed those documents to their boss.”

A total of 100 exemplar sentences of Japlish were constructed and subsequently checked with a male native speaker of American English as to whether

⁵ Though the previous research also focused upon other word order types, SOV, SIOV, SSOVV, and SSIOVV, a decision was made not to include them in the current investigation, so as to avoid their potential conflict with participants’ L1 English word orders (i.e., SVO, SVIO, SVSVO, and SVSVIO).

each sentence made sense or not. The same native speaker was also asked to read aloud the sentences with a normal speech rate and the reading was audio-recorded so as to create a sound file for each sentence. Out of the entire set, 25 sentences corresponded to each word order type, each of which contained thirteen semantically plausible and twelve implausible sentences. The entire set of sentences was presented twice to the participants (amounting to 200 trials in total), and the orders of the presentation were randomized regardless of sentence complexity, following Grey, Williams, and Rebuschat (2014)'s suggestion. For each word order type, the number of words in the sentences was controlled (OSV: four words, OSIV: six words, OSSVV: seven words, and OSSIVV: nine words). Appendix A lists the exposure stimuli used in the study.

4.2.2 Untimed Auditory Grammaticality Judgment Task

The untimed AGJT served as a measure of explicit knowledge, which allowed participants to deploy controlled processing. Eighty Japlish sentences were constructed; of the 80 test items, 32 were grammatical and 48 were ungrammatical sentences (8 items for each grammatical and ungrammatical item type).

Ungrammatical items were devised such that they either had an illicit word order, contained a noun whose case marker was missing, or positions of the case markers mixed. Table 1 illustrates each grammatical and ungrammatical item type introduced in the study. For all items with a case marking violation, sentences were presented in the OSIV word order type for two reasons. First, this word order type contained all case markers in question (Grey et al., 2014), and second, the control of the type of word order used for these items allowed for a reliable measurement of the extent to

Table 1. *The Untimed AGJT: Item Types*

Type	Example
<u>Grammatical</u>	
OSV	<i>This ink-o Stacey-ga spilled.</i>
OSIV	<i>This language-o Mike-ga his student-ni taught.</i>
OSSVV	<i>The diamond-o John-ga that man-ga stole said</i>
OSSIVV	<i>A bribe-o Karen-ga her colleague-ga their boss-ni offered mentioned.</i>
<u>Ungrammatical</u>	
<u>Word Order</u>	
OVS	<i>The fire-o lighted Angela-ga.</i>
OSVI	<i>The letter-o Mary-ga faxed her boss-ni.</i>
OSVSV	<i>That vase-o Stacey-ga broke her spouse-ga thought.</i>
OSSVIV	<i>The document-o Mary-ga her workmate-ga faxed her boss-ni that told.</i>
<u>Case Missing</u>	
-ga	<i>A tip-o Tim the driver-ni gave.</i>
-o	<i>The clothes Pamela-ga her daughter-ni chose.</i>
-ni	<i>A letter-o Steve-ga the mayor wrote.</i>
<u>Case Mixing</u>	
-ga, -o	<i>A tip-ga Tim-o the driver-ni gave.</i>
-ga, -ni	<i>The clothes-o Pamela-ni her daughter-ga chose.</i>
-o, -ni	<i>A letter-ni Steve-ga the mayor-o wrote.</i>

Note. Bold facing indicates ungrammatical elements.

which the participants developed explicit knowledge of the case markings alone, without confounding it with the knowledge of the word orders. Four lists of presentation were developed by first creating two lists of items with entirely different sentences, and subsequently counterbalancing the sentences by grammaticality. The participants were presented with two of the presentation lists at the immediate and delayed posttest and care was taken so that they would not see the same sentences twice in the two testing sessions. Appendix B lists the materials for the untimed AGJT. During the task, the participants first saw an asterisk for 500 milliseconds (ms) and subsequently listened to a Japlish sentence. They were instructed to judge whether each sentence conformed to the patterns of the language, by pressing either

the YES-key if they thought the sentence was grammatical or the NO-key if ungrammatical. For each response, participants were further asked to indicate the level of confidence that they had and also the basis for that response. For the confidence ratings, the participants used a scale of 1-5 key on a keyboard, which corresponded to each confidence level, 1 = “guess: 50%”, 2 = “somewhat confident: 60-70%”, 3 = “confident: 70-80%”, 4 = “very confident: 80-90%”, and 5 = “absolutely certain: 100%”, respectively. For the source attributions, keys 1-4 were assigned, each of which corresponded to 1 = “guess”, 2 = “intuition”, 3 = “memory of items from the exposure phase”, and 4 = “rule”. The granularity of the scales was adopted from a study by Kachinske, Osthus, Solovyeva, and Long (2015) with slightly different labels (see Rebuschat, 2013 for guidelines). The meaning of each confidence level and source attribution category was carefully explained to the participants; for instance, participants were instructed to choose “guess: 50%” (confidence) and “guess” (source) category when their response was based on a complete guess (i.e., 50/50), and “intuition” when they felt the sentence was grammatical (or ungrammatical) but they did not know as to why. Figure 1 graphically summarizes the entire procedure of the task. The reliability of the task based on Cronbach’s alpha was $\alpha = 0.92$ and 0.94 , respectively, at the immediate posttest and the delayed posttest.

4.2.3 Word-Monitoring Task

The WMT served as a measure that required automatic processing of language. In the task, participants monitored incoming auditory input sentences for a

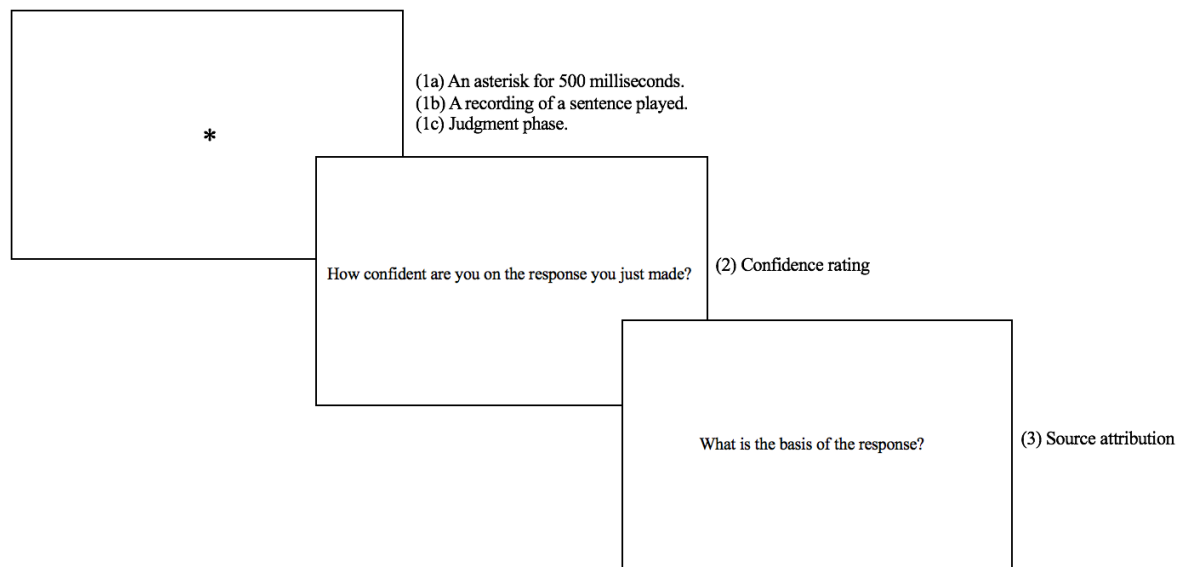


Figure 1. Procedure for the untimed AGJT

target word for detection. One hundred thirty Japlish sentences were constructed for the task; 96 sentences were target sentences, half of which were grammatical and the other half were ungrammatical (8 items for each grammatical and ungrammatical item type). The rest of the sentences were used as grammatical distractors. It must be noted that all grammatical and ungrammatical item types were the same as those for the untimed AGJT, and four lists of presentation were also created in the same manner (see Appendix C for the entire items). However, all item sentences in the task were different from those in the untimed AGJT, so as to ensure that the participants would never hear the same sentences twice. Furthermore, the items also contained a multi-word adverb phrase at the beginning and the end. This was to ensure that ungrammatical elements would not come at the place of the first or the second word, and the last or the second to last word of the sentences. As pointed out by Jiang

(2012), participants might not be as focused at the beginning or the end of a sentence as they are in the middle. Thus, this manipulation was deemed justifiable, as it was intended to prevent participants from missing the target word due to the positional effect. Each target word to be monitored was chosen such that it immediately followed the ungrammatical element in the sentence.

During the task, the participants first saw an asterisk for 500 ms; it subsequently turned into a target word to monitor. The target word was presented visually and remained on the screen while a recording of a carrier sentence was being played (i.e., cross-modal presentation). The participants were instructed to focus on the meaning of the carrier sentence as well as the appearance of the target word, and to press a corresponding key as soon as they hear the target. For one sentence out of two (i.e., 50%), a comprehension question followed the carrier sentence in order to ensure that the participants would focus on the meaning of each sentence. In this dual-task condition, the word-monitoring and comprehension of meaning minimized the possibility that the participants would use any explicit knowledge of grammar or conscious strategies (e.g., Kilborn & Moss, 1996). The time window for the measurement of monitoring latencies started at the onset of the presentation of the target word. By using split recordings of the target stimuli, it became possible to use the same recording of the first half of the stimuli for both grammatical and ungrammatical items, which served to neutralize the expected variability in recorded speech (see Jiang, 2012 for a methodological guideline of a word-monitoring task). Figure 2 graphically presents the entire procedure of the WMT. The reliability of the

task based on Spearman-Brown prophecy formula with a split-halves method was $r = .95$ and $.81$, respectively, at the immediate posttest and the delayed posttest.

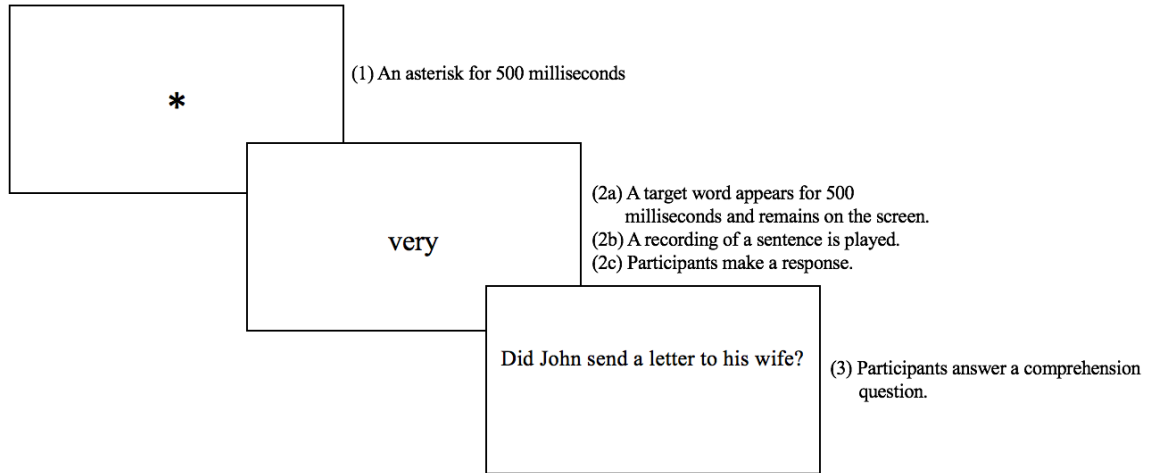


Figure 2. Procedure for the WMT

4.2.4 LLAMA-F

The construct of language-analytic ability was operationalized with a score on the adaptive version of LLAMA F, *Grammatical Inferencing* (Meara, 2005). Task materials were adopted from the IRIS Repository (Marsden, Mackey, & Plonsky, 2016); these were the same materials from Suzuki and DeKeyser (2017)⁶, submitted to IRIS by the authors. The entire task consisted of two phases, a learning phase and a testing phase. In the former, the participants were given five minutes to study an unknown language grammar whose sentence exemplars matched a corresponding picture conveying the meaning of the sentence. The testing phase immediately followed the learning phase during which the participants saw a picture and two

⁶ For the task materials, please visit the IRIS website (<https://www.iris database.org/iris/app/home/detail?id=york%3a932045&ref=search>, for Suzuki & DeKeyser, 2017).

sentences, one grammatical and one ungrammatical. They were instructed to choose the one they judged to be grammatical. The task contained 30 items, and the participants were not allowed to change their answer once they marked either one of the options. The reliability of the task as measured by Cronbach's alpha was $\alpha = 0.75$, and the scores of the experimental and the control groups were not significantly different from each other, $t(47) = 0.475, p = .673$.

4.2.5 Listening-Span Task

Due to the auditory nature of the materials in the study, a listening-span task was used to measure the participants' L1 working memory capacity. Again, the task materials were adopted from the IRIS Repository, which were originally developed by Mackey, Philp, Egi, Fujii, and Tasumi (2002) and based on previous work on other span tests in cognitive psychology (Daneman & Carpenter, 1980; Turner & Engle, 1989; Waters & Caplan, 1996). The task was constructed so as to involve both storage and processing components of the participants' WM capacity, which were deemed to be critical in language processing. The task consisted of 48 sentences, half grammatical and half semantically plausible, which were grouped into 12 presentation sets: four sets of three, four, and five sentences. The participants were instructed to judge the grammaticality *and* plausibility of the sentences. At the end of each presentation set, the participants were prompted to recall the last word of each sentence in order. These words were common, non-compound, concrete nouns of one to three syllables (Winke, 2013)⁷. In the present investigation, the participants

⁷ Please visit the IRIS website for the task (<https://www.iris-database.org/iris/app/home/detail?id=york%3a806454&ref=search>).

responded in a paper-and-pencil format but listened to the sentences through headphones (i.e., computer-delivered). A recall of each word was scored as one point and each grammaticality and semantic plausibility judgment was given a half point. This made 96 points, the maximum score possible. The reliability of the task as measured by Cronbach's alpha was $\alpha = 0.79$. The scores of the two groups were not significantly different from each other, $t(43) = 1.813, p = .077$.

4.2.6 Phonological Short-Term Memory Task

A nonword repetition task developed by Lado (2008, 2017), also used by Grey, Cox, Serafini, and Sanz (2015) and Grey, Williams, and Rebuschat (2015)⁸, was adopted from the IRIS Repository. The task contained sixteen pairs of nonwords, each of which was presented individually. The participants were instructed to repeat the words aloud in the order presented after a beep sound that followed the presentation of each pair. An audio recording of the task was played through headphones and the responses were made into a microphone attached to them. Those sixteen sets of nonwords gradually increased in length, from 3 to 8 syllables (i.e., 16 syllables at the largest). Following Grey, Cox, Serafini, and Sanz (2015), the dependent variable was scored by assigning one point for each correctly recalled word without more than one error on the two-item set (i.e., a pair). The reliability of the task as measured by Cronbach's alpha was $\alpha = 0.79$. Again, the scores of the two groups were not significantly different from each other, $t(44) = 0.177, p = .860$.

⁸ Please visit the IRIS website for the task (https://www.iris-database.org/iris_app/home/detail?id=york%3a933009&ref=search).

4.2.7 Serial Reaction Time Task

A probabilistic version of the serial reaction time task originally developed by Kaufman et al. (2010) measured the participants' implicit sequence-learning ability. In the task, the participants saw a black dot successively appear in one of four prescribed locations on the computer screen and tried to respond as quickly and accurately as possible by pressing corresponding keys to the locations (i.e., 'V', 'B', 'N', 'M' keys). Unbeknownst to the participants, however, the serial presentation of stimuli followed a probabilistic rule wherein 85 % of the stimuli followed one sequence, whereas the remaining 15 % followed the other sequence, both of which were governed by the second-order conditionals (i.e., a location of the stimulus is determined by locations of the two previous stimuli). Specifically, two sequences from Kaufman et al. (2010) were employed in the present study: Sequence A (1–2–1–4–3–2–4–1–3–4–2–3) and Sequence B (3–2–3–4–1–2–4–3–1–4–2–1). Thus, in one block, the stimulus sequence followed Sequence A with a probability of .85, and Sequence B with of .15. The probabilistic nature and this complex second-order conditional rule made the task complex so that it was difficult for the participants to encode the sequences explicitly.

Although previous studies included a surprise posttask recognition test so as to ascertain whether participants did not acquire any explicit knowledge about the sequences (e.g., Granena, 2012, 2013a; Suzuki, 2015, 2017; Suzuki & DeKeyser, 2017), the present study did not have any recognition tests after the task. This decision was justified considering that none of the SLA studies reviewed here found a participant that came to be consciously aware of the second conditional rules that

underlay the sequences. Lastly, the difference between the mean RTs in the trials of the correct sequence and those of another sequence was used as the participants' implicit sequence-learning ability. The reliability of the task as measured by the Spearman-Brown prophecy formula with the split-halves method was $r = .72$, and the scores from the two groups were not significantly different from each other, $t(41) = -0.419, p = .677$.

4.3 Procedure

Participants first signed up for the study on the University of Maryland Psychology Research Sign-Up System or contacted the experimenter directly to take part in the experiment. The participants were prescreened at the time of the contact such that only those who had had no experience with Japanese nor other case-marking languages and who were between 18 and 40 years of age were invited. In addition, they were allowed to schedule an appointment only if they stated they would be able to return for the delayed posttest, exactly two weeks after the first session. The entire procedure of the study followed a fixed order of presentation shown in Table 2.

At the beginning of the first session, the participants were asked to sign a consent form in which an overview of the entire study was presented. The study was introduced to them as a project in which the experimenter investigated whether or not native speakers of English were able to comprehend an artificial language that had been developed recently. Although the participants were notified that there would be a few tests that follow the training phase, they were also told that the tests would

Table 2. *Procedure of the Study*

	Task	Min.
1	Consent Form and Biographical Questionnaire	10
2	Exposure Task	30
3	Intermediate Debriefing Session	5
4	Word-Monitoring Task	30
5	Break	10
6	Untimed Auditory Grammaticality Judgment Task	20
7	Nonword Repetition Task	4
	↓	
1	Word-Monitoring Task	30
2	Untimed Auditory Grammaticality Judgment Task	20
3	Break	10
4	Serial Reaction Time Task	10
5	Listening Span Task	18
6	LLAMA-F	10
7	Post-experimental Questionnaire	10
		Total = 217

be on comprehension of sentences. A biographical questionnaire accompanied the consent form, which inquired about their gender, age, major, university year, native language, known second language(s), estimated proficiency in the second language(s), experience with the second language(s), and foreign countries they had visited (see, Appendix D). The participants were guided to the exposure task only if they agreed to participate in the study and if they filled out the questionnaire. At this point, each participant was randomly assigned to either the experimental group or the control group.

During the exposure phase, the participants performed a semantic plausibility judgment on each sentence. Using two contrasting sentences in English (i.e., *John ate an apple* vs. *John ate a chair*), it was described that their task was to judge the plausibility of each sentence in meaning. They were informed that the language they would be listening to was not English, but they would be able to understand it. As the task began, those in the experimental group listened to 100 Japlish sentences with

patterns (i.e., four word order rules and three case markers), whereas those in the control group listened to Japlish sentences where word order and positions of case markers were pseudo-randomized such that they would listen to every possible word order and case marking configuration with an (almost) equal number of exemplars. For instance, the control group on OSV order was exposed to OSV, OVS, SOV, SVO, VOS, and VSO order, with an equal frequency (i.e., 5 or 4; see Appendix B for details). The order of presentation was randomized and the task was repeated twice, amounting to 200 trials in total.

Immediately following the exposure task, the participants were told that the study was actually on learning of the artificial language, not comprehension, and that they would be subsequently tested upon the knowledge they had acquired from the exposure phase. Another consent form, which spelled out the true aims of the experiment, was provided to the participants; they were instructed to sign if they still agreed to participate⁹. This phase served as an intermediate debriefing session, during which the true nature of the study was explained and the participants were encouraged to ask questions regarding the entire experiment.

The WMT followed the debriefing session as the first testing task. The task was described as a test of language processing and comprehension, so as not to disclose the existence of ungrammatical items in the task. The participants were instructed that they must concentrate at times and that they must attend to both the meaning of sentences and the appearance of the target word. The task was structured in a way that provided a short 1-2 minute break after every 30-item set of sentences.

⁹ None declined to participate here.

Despite the presence of the ungrammatical items, no one reported their existence in the task.

For the untimed AGJT, the participants were instructed to make a judgment as to whether each sentence was grammatical or not, based on the experience they had in the exposure phase. Additionally, they were also asked to perform the confidence rating and source attributions on each judgment. The items were randomly presented, and a short break was taken after every 30 items. As suggested by Hamrick and Sachs (2017), the instructions throughout the study were kept the same for both the experimental and the control groups. The DMDX software developed by Forster and Forster (2003) was used to carry out the exposure task, WMT, and AGJT, and the latter two had eight practice items before the beginning of the test trials.

The NWRT came as the last task in the first session. Before the commencement of the task, the experimenter described what was meant by *nonwords* and that the participants were to try their best to hold the pair of nonwords in memory and to repeat them back in the order they were presented. After completion of the NWRT, the participants were reminded of the second session and received half of the course credit or \$20 as the compensation for the first session.

Two weeks later, the participants came back to the laboratory and started the second session with the WMT and untimed AGJT. Subsequently, they were asked to take the SRT. They were reminded that their task was to respond as quickly *and* accurately as possible and that they must keep their second and third finger from both hands on the keyboard. The LSPAN task immediately followed the SRT, and the experimenter provided a pen and an answer sheet for the task, and described the task

to the participants. The task was carried out by playing the recording of the material, while the participants listened to sentences and wrote down their answers on the answer sheet (i.e., a plausibility and grammaticality judgment for each sentence and also a recall of the last word of each sentence at every end of an item set). Lastly, the participants took the LLAMA-F test, in which they were instructed to figure out the grammar of an unknown language. The study phase was set to take place in 5 mins and testing immediately followed.

At the end of the second session, the participants answered a post-experimental questionnaire, in which they were asked to perform retrospective verbal reports on their noticing and understanding about the structures of the language. The questionnaire was constructed such that questions became increasingly more explicit and directing in probing the level of the awareness they possessed (i.e., awareness at the level of noticing vs. understanding) and also when they came to be aware. There were nine questions in the survey which inquired about the three aspects of the language in question: simple word order types, complex word order types, and case marking types. Upon completing the questionnaire, they were thanked for their participation and received the other half of the course credit or \$20 as final compensation.

4.4 Analysis

4.4.1 Untimed Auditory Grammaticality Judgment Tasks

Mean percentile accuracy rate for overall, grammatical, and ungrammatical items were calculated across different construction types. In the main analysis, d' (d-

prime) index was computed for each participant on each grammatical and ungrammatical construction type. This allowed the analysis to take into account any response bias that the participants might have had during the decision-making process (Kunimoto, Miller, & Pashler, 2001; Rebuschat, 2013). First, the group differences on the d-prime scores were examined through a multivariate analysis of variance (MANOVA) with Group as a between-subjects factor (two levels: Experimental and Control) and d-prime scores on six construction types as dependent variables (i.e., OSV, OSIV, OSSVV, OSSIVV, CaseMis, and CaseMix). If the multivariate analysis detected a significant main effect of Group, follow-up univariate analyses of variance were conducted (ANOVAs) to examine on which dependent variables the experimental group and the control group differed significantly from each other. Subsequently, the extent to which the four cognitive aptitudes moderated the learning of Japlish was explored, using a multivariate analysis of covariance (MANCOVA), this time with the experimental group only, with d-prime scores on six construction types as dependent variables, and with the four aptitude measures as covariates. Provided that the MANCOVA detected the multivariate effect of any of the covariates, follow-up univariate analyses of covariance (ANCOVAs) were conducted to see for which construction types the aptitude(s) significantly moderated the participants' performance on the AGJT.

For the subjective measures of awareness, two approaches were taken to examine the relationship between the accuracy scores and the confidence ratings and the source attributions. First, a multiple logistic regression model was built to examine the correlation between the confidence level and the accuracy score. Here,

the dependent variable was on the nominal scale, correct or incorrect (i.e., 1 or 0), and the predictor variables included the level of confidence that the participants reported and the six construction types that were dummy-coded through the effect coding method.

Second, a mean proportion of correct to incorrect answers was calculated for each source attribution category (i.e., Guess, Intuition, Memory, and Rule) on each construction type (24 data points). Consequently, one sample t-tests were run to examine if each proportion of correct answers was significantly better than the chance level. Here, the above-chance level was *not* defined to be 50% accuracy, but rather, it was based on the proportion of correct to incorrect answers calculated in the following way. First, binary correct and incorrect scores were randomly generated by *R* software (version 3.3.3, R Core Team, 2017) with *sample* function, which were as many as the number of responses observed in each category. Second, the proportion of correct answers was calculated over the randomly generated scores, and it was used as a population estimate against which the observed proportion score was compared in the t-test. This procedure was necessary in order to avoid the assumption that the chance-level performance in the population was 50% accuracy (which is not necessarily true in reality). Bonferroni correction was made to the alpha-level in order to avoid committing a Type-I error.

These two processes for the confidence ratings and source attributions were implemented in order to reliably inspect whether the zero correlation criterion and the guessing criterion were met or not met in this study (Dienes, Altmann, Kwan, & Goode, 1995; Dienes & Scott, 2005). Note that the analysis concerning the subjective

measures of awareness was carried out only for the experimental group. The same procedure was followed in the analysis of the delayed posttest as well.

Mean and d-prime scores for one participant at the immediate posttests were lost due to technological difficulties. Since all other data for the person were available including language tests and cognitive aptitude measures, six d-prime data points were imputed using *mice* package (version 2.46.0, van Buuren & Groothuis-Oudshoorn, 2011) in *R* software. The data were imputed for five times and the means of the five iterations were taken as the estimates.

4.4.2 Word-Monitoring Task

For a descriptive purpose, mean raw RTs on overall, grammatical, and ungrammatical items were calculated across different construction types. Before summarizing, data points that exceeded $\pm 3SD$ of the person's mean and that were larger than 2500 ms and smaller than 100 ms were excluded from the analysis. This resulted in an exclusion of 4.08 % (immediate) and 4.18% (delayed) of the entire data set for the WMT task. The mean monitoring latencies were transformed into their reciprocals and multiplied by -1000 (TransRTs) so as to reduce the positively skewing nature of RT data. First, the group differences on the monitoring latencies were examined through a MANOVA with Group as a between-subjects factor, with Grammaticality as a within-subjects factor, and with TransRTs on each construction type as dependent variables. Subsequently, univariate ANOVAs were conducted to see on which construction types the experimental group and the control group were significantly different from each other in their word-monitoring latencies to grammatical and ungrammatical items.

For the analysis of the moderation effect of the cognitive aptitudes on the WMT, mean raw RTs on grammatical items were subtracted from those of ungrammatical items to derive the Grammaticality Sensitivity Index (GSI). GSIs were used in previous studies to operationalize an implicit sensitivity to violation of grammatical patterns (Granena, 2012; Suzuki, 2015). Another MANCOVA was conducted with GSIs on each construction type as dependent variables and the four cognitive aptitude measures as covariates (only for the experimental group). Subsequent ANCOVAs were also conducted in order to examine on which construction types the four aptitudes significantly moderated the participants' performance on the WMT. Note that the same procedure was followed in the analysis of the delayed posttest as well.

For all of the null-hypothesis significance testing, an alpha-level was set at .05. For the ANOVA and ANCOVA analyses, a partial eta-squared (η_p^2) was used for index of the effect size, and R^2 for the correlational and regression analyses. Following Cohen (1988), η_p^2 (or R^2) between 0.01 and 0.06 was considered as a small effect size, between 0.06 and 0.14 as a medium effect size, and of more than 0.14 as a large effect size. In the entire analysis, care was taken to examine every possible violation of statistical assumptions associated with the statistical methods.

4.4.3 Retrospective Verbal Reports

The first and second questions of the questionnaire asked in general whether or not the participants noticed or understood any features of the experiment and the language and, when they did, they were asked to state what and when they noticed. The third, fourth, and fifth questions specifically provided the label “word order” and

“case-markers” (i.e., word endings) and asked to state any forms or rules that the participants noticed or understood about, and when they did, when they noticed. The sixth, seventh, and eighth question explicitly mentioned that the language contained simple and complex word order types, and which ones they thought they saw in the experiment. Lastly, the ninth question directly provided -ga, -o, and -ni and asked if the participants were able to tell what their roles were in the language.

All of the nine questions were structured in order of increasing explicitness and directness and examined the extent to which the participants became aware of the patterns in the language. The present study employed three levels of conscious awareness: *No Report*, *Noticing*, and *Understanding* (Schmidt, 1990, 1995, 2001). Consciousness at the level of understanding here was operationalized as a correct provision of rule of specific word orders or case markers. On the other hand, the status of conscious awareness that did not yield any correct rules but showed awareness of their existence was considered sufficient evidence of noticing. Hence, any mention of Japlish word orders and case markers sufficed to as evidence of noticing. These two constructs of awareness have been widely used in SLA research (see Robinson, Mackey & Gass, 2012 for a recent review) and it seemed justified to employ this categorization scheme. The present study only discusses the retrospective verbal reports in descriptive terms.

Chapter 5: Results

5.1 Immediate Posttests

5.1.1 Untimed Auditory Grammaticality Judgment Task

Mean percentile accuracy and d-prime scores for the experimental and the control groups are summarized in Table 3 and 4, respectively. The experimental group surpassed the control group in all respects, except the mean accuracy for the grammatical items of OSV sentences (the experimental group, 84.35 and the control group 88.33). Before running the inferential statistics, the assumption of normality, homogeneity of variance, and sphericity were checked. The Shapiro-Wilk Test for the

Table 3. *Mean Percentile Accuracy for the AGJT at the Immediate Posttest*

	Overall	OSV	OSIV	OSSVV	OSSIVV	Case Mis	Case Mix
Experimental							
Overall							
<i>Mean</i>	73.65	84.35	75.29	72.99	68.34	72.56	68.36
<i>SD</i>	27.76	21.88	28.70	23.73	30.67	29.16	29.63
Grammatical							
<i>Mean</i>	82.53	83.89	88.07	71.43	75.62	88.07	88.07
<i>SD</i>	20.64	20.65	13.88	24.81	27.23	13.88	13.88
Ungrammatical							
<i>Mean</i>	64.77	84.82	62.50	74.55	61.05	57.05	48.66
<i>SD</i>	31.01	23.41	33.85	22.94	32.05	32.27	28.12
Control							
Overall							
<i>Mean</i>	56.48	66.04	59.02	54.06	46.34	56.12	57.32
<i>SD</i>	32.82	34.20	32.51	27.21	30.36	34.76	34.01
Grammatical							
<i>Mean</i>	73.47	88.33	82.41	59.38	45.89	82.41	82.41
<i>SD</i>	26.47	19.38	18.30	22.17	31.95	18.30	18.30
Ungrammatical							
<i>Mean</i>	39.49	43.75	35.63	48.75	46.79	29.82	32.23
<i>SD</i>	29.68	31.28	30.96	31.12	29.49	26.27	26.73

Table 4. *d-prime Scores on the AGJT at the Immediate Posttest*

	Overall	OSV	OSIV	OSSVV	OSSIVV	Case Mis	Case Mix
Experimental							
<i>Mean</i>	1.94	2.88	2.12	1.75	1.48	1.84	1.55
<i>SD</i>	1.63	1.45	1.73	1.47	1.86	1.56	1.41
<i>Max</i>	4.65	4.65	4.65	4.65	4.65	4.65	4.65
<i>Min</i>	- 2.01	0.17	- 0.58	- 0.67	- 2.01	- 1.15	- 1.76
Control							
<i>Mean</i>	0.50	1.36	0.78	0.14	- 0.17	0.56	0.57
<i>SD</i>	1.43	1.53	1.18	1.30	1.14	1.45	1.50
<i>Max</i>	4.65	4.65	3.00	3.00	2.01	4.65	4.65
<i>Min</i>	- 2.65	- 1.82	- 0.99	- 2.01	- 2.65	- 1.66	- 1.26

experimental group showed a significant test statistic, $W(168) = 0.955, p < .000$, Skewness = 0.11 ($SE = .187$), and Kurtosis = - 0.89 ($SE = .373$), indicating that the data are not exactly following the shape of normal distribution. However, this significant result could have been primarily caused by the large sample size, $df = 168$. Thus, a z-score of skewness and kurtosis was calculated for each and examined whether the values exceeded ± 2 range (Field, Miles, & Field, 2012). The results showed that while the z-score of skewness remained within the range ($z_{\text{skewness}} = 0.588$), that of kurtosis did not ($z_{\text{skewness}} = 2.386$). Thus, the results of the following MANOVA and ANOVA analyses must be interpreted with caution because the distribution did not exactly follow the shape of normal distribution.¹⁰ Figure 3 graphically presents a histogram of the data distribution. Furthermore, the Shapiro-Wilk Test for the control group also reached significance, $W(126) = 0.969, p < .000$, Skewness = 0.53 ($SE = .216$), and Kurtosis = 0.45 ($SE = .428$), but the z-score of skewness and kurtosis remained within ± 2 range this time.

¹⁰ As d-prime score itself is already a standardized index, no transformation of the data could resolve the problem.

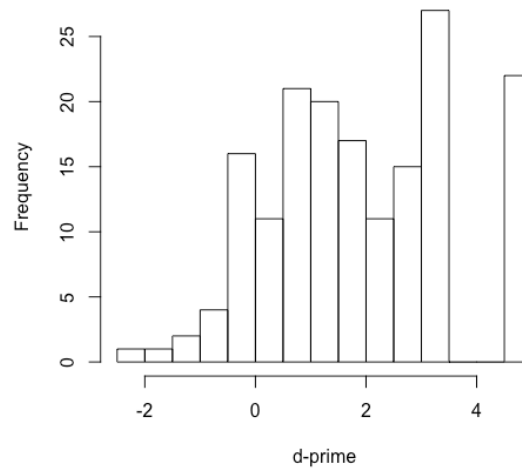


Figure 3. Histogram of d-prime Scores for the Experimental Group at the Immediate Posttest

Furthermore, Levene's Test for homogeneity of variance also showed a significant test statistic, $F(1, 292) = 6.962, p = .008$. However, the difference across the groups was considered to be non-substantial, as the largest standard deviation was not three times larger than the smallest standard deviation between the groups (Houser, 2008). In addition, (M)ANOVA is considered reasonably robust when the sample size is more than 20.

A MANOVA on the d-prime scores showed that there was a significant main effect of Group at the multivariate level, $F(1, 42) = 3.538, p < .000$; Wilk's $\Lambda = .664$, $\eta_p^2 = .336$. Furthermore, subsequent follow-up ANOVAs showed that the experimental group outperformed the control group on all of the construction types, $F(1, 42) = 12.645, p = .001, \eta_p^2 = .212$ for OSV; $F(1, 42) = 9.293, p = .004, \eta_p^2 = .165$ for OSIV; $F(1, 42) = 15.993, p < .000, \eta_p^2 = .254$ for OSSVV; $F(1, 42) = 12.822, p = .001, \eta_p^2 = .214$ for OSSIVV; $F(1, 42) = 8.651, p = .005, \eta_p^2 = .155$ for CaseMis; $F(1,$

42) = 12.645, $p = .024$, $\eta_p^2 = .104$ for CaseMix. Figure 4 graphically summarizes the group differences (red: experimental, blue: control).

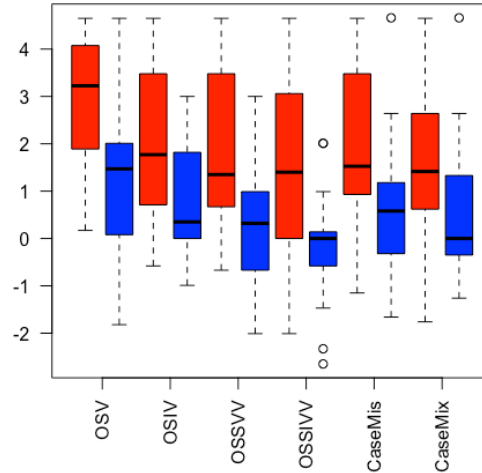


Figure 4. Boxplot of the AGJT at the Immediate Posttest

In order to understand which structures proved to be more difficult than the others, a repeated-measures ANOVA was conducted. This time the analysis was carried out only with the experimental group, and with Construction as a within-subjects factor, followed by a post-hoc analysis with a Fisher's Least Significance Difference (LSD) test. On this occasion, Mauchly Tests for Sphericity did reach significance ($\chi^2(14) = .419$, $p = .084$), and thus Greenhouse-Geisser correction was made to the degree of freedom for the factor. The results of the repeated-measures ANOVA showed that there was a main effect of Construction, $F(5, 135) = 8.299$, $p < .000$, $\eta_p^2 = .188$. Table 5 summarizes results of the post-hoc analysis. The experimental group was most accurate on OSV, for which they scored better than the other construction types. Furthermore, the difference between OSIV and OSSIVV and between OSIV and CaseMix were also significant. Overall, the results demonstrated that the experimental group performed significantly better on the two

Table 5. *Post-hoc Difference between the Construction Types*

	Mean Difference	SE	p
OSV - OSIV	.765	.266	.008
OSV - OSSVV	1.133	.276	.000
OSV - OSSIVV	1.401	.321	.000
OSV - CaseMis	1.042	.282	.001
OSV - CaseMix	1.335	.232	.000
OSIV - OSSIVV	.636	.201	.004
OSIV - CaseMix	.570	.570	.018

simple word order constructions than the others (including the case markings).

However, there was a significant difference between the two simple word order types as well, in that the participants were more accurate on OSV than they were on the other constructions, whereas their scores on OSIV were only better than those of OSSIVV and CaseMix, the most complex word order type and the more difficult case marking violation type, respectively¹¹.

5.1.2 The Role of Cognitive Aptitudes

Table 6 summarizes descriptive statistics for the four cognitive aptitude measures. Before conducting the MANCOVA analysis, additional assumptions specific to MANCOVA were examined. First, the assumption of multicollinearity was checked by correlating all cognitive aptitude measures against one another. Table 7 summarizes results of the correlational analyses. For analytical purposes, they were transformed into z-scores.

¹¹ Comparing the two case marking violation types, the case mixing violation was deemed to be more difficult to detect than the case missing violation, because while the former required knowledge of form-meaning mappings to detect the violation, the latter only required that of forms.

Table 6. *Descriptive Statistics of the Cognitive Aptitude Measures*

	LLAMA-F	LSPAN	NWRT	SRT
Experimental				
<i>Mean</i>	23.50	75.73	14.93	10.28
<i>SD</i>	4.66	7.11	4.29	21.33
<i>Max</i>	30	87.5	23	58.97
<i>Min</i>	13	58	5	- 30.65
Control				
<i>Mean</i>	22.95	72.02	14.71	27.46
<i>SD</i>	3.41	7.06	4.11	23.22
<i>Max</i>	28	89	23	58.95
<i>Min</i>	16	65	9	- 40.01

Table 7. *Correlational Matrix for the Cognitive Aptitude Measures*

	LLAMA-F	LSPAN	NWRT	SRT
LLAMA-F	-	$p = .256$	$p = .636$	$p = .063$
LSPAN	$r = .16$	-	$p = .626$	$p = .305$
NWRT	$r = .07$	$r = .17$	-	$p = .907$
SRT	$r = .28$	$r = .15$	$r = -.02$	-

Note. The lower half lists Pearson Correlation Coefficient and the upper half lists the corresponding p -values.

The results in Table 7 show that none of the aptitude measures significantly correlated with each other. Based on this finding, it was concluded that the present study would use each aptitude measure as a predictor variable on its own, rather than trying to combine them into a smaller number of components. Nevertheless, given that they were hypothesized to lie on a continuum of explicit/controlled and implicit/automatic cognitive processing, the fact that none of them correlated with any other is a striking finding.

Second, the assumption of homoscedasticity was checked by looking at a scatterplot of the residuals from the MANCOVA model and the predicted values that were generated by the same model (Figure 5)¹². The plot exhibited a good dispersion,

¹² Although the residuals and the predicted values are products of the MANCOVA

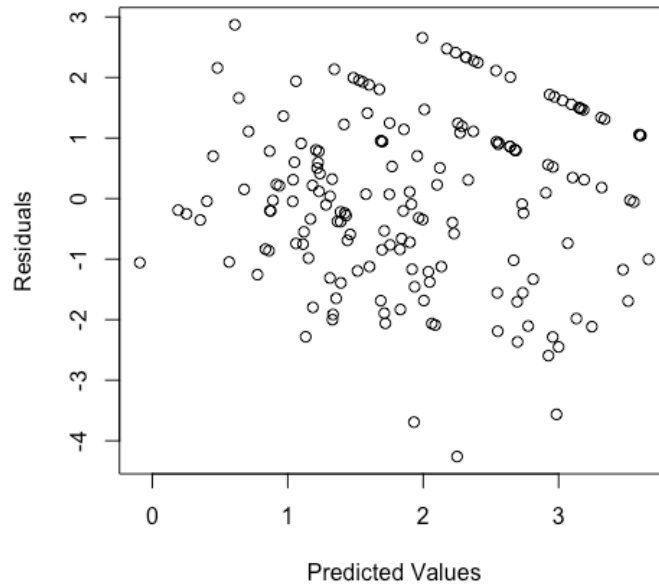


Figure 5. Scatterplot of the Residuals and the Predicted Values on AGJT at the Immediate Posttest

except that the data points seemed to slightly converge at the positive extreme of the predicted values. Hence, it was concluded that that the assumption was satisfied here.

As the main analysis, the results of a MANCOVA showed that a main effect of NWRT as a covariate was significant at the multivariate level, $F(6, 18) = 3.013$, $p = .032$; Wilk's $\Lambda = .499$, $\eta_p^2 = .501$. However, the other three covariates were not significant in the MANCOVA model, $F(6, 18) = .574$, $p = .746$, Wilk's $\Lambda = .574$, $\eta_p^2 = .161$ for LLAMA-F; $F(6, 18) = .790$, $p = .589$, Wilk's $\Lambda = .792$, $\eta_p^2 = .208$ for LSPAN; $F(6, 18) = .482$, $p = .813$, Wilk's $\Lambda = .862$, $\eta_p^2 = .138$ for SRT. These results indicate that individual differences in the participants' PSTM capacity (measured by NWRT) significantly moderated their performance on one or more construction

analysis presented below, it was considered necessary to discuss them beforehand, so as to ensure that the assumption was met.

types. Follow-up ANCOVAs showed that the participants' performance on OSIV and the case-missing violation items was significantly moderated by their PSTM capacity, $F(1, 18) = 4.327, p = .049, \eta_p^2 = .158$ for OSIV, and $F(1, 18) = 6.761, p = .016, \eta_p^2 = .227$ for CaseMis.

To further understand the role of PSTM and their performance on the AGJT, an additional correlational analysis was conducted. Table 8 presents a summary of six pairwise correlation coefficients between NWRT scores and d-prime scores on six construction types. Their corresponding p -values are also presented. The correlational analysis summarized in Table 8 shows that the participants' performance on items that concerned OSIV and Case Missing violation items was significantly and positively related to the their PSTM capacity. Furthermore, the correlation coefficient for OSSVV and OSSIVV items reached significance, if not less than .05. These insignificant results primarily could have been due to the small sample size of the experimental group ($n = 28$). Figure 6 and 7 graphically represent the correlations among the variables.

Table 8. *Correlational Matrix for NWRT and Six Construction Types*

	OSV	OSIV	OSSVV	OSSIVV	CaseMis	CaseMix
NWRT	-.05	.42	.33	.35	.478	.199
p	.792	.024*	.086	.066	.009*	.309

Note. * indicates a significant result. **indicates a significant result even after Bonferroni Correction (.05/6= .008)

5.1.3 The Role of Confidence Ratings and Source Attributions

To investigate the extent to which the level of confidence that the participants reported on each judgment was related to the accuracy on the task, a multiple logistic

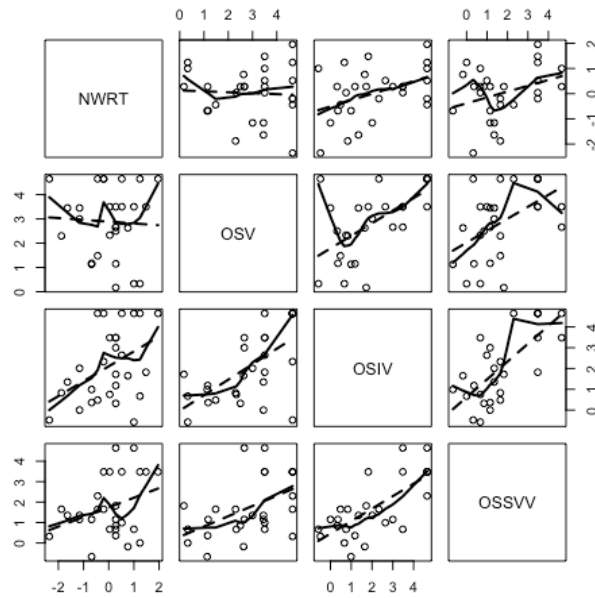


Figure 6. Scatterplot Matrix: NWRT, OSV, OSIV, and OSSVV at the Immediate Posttest

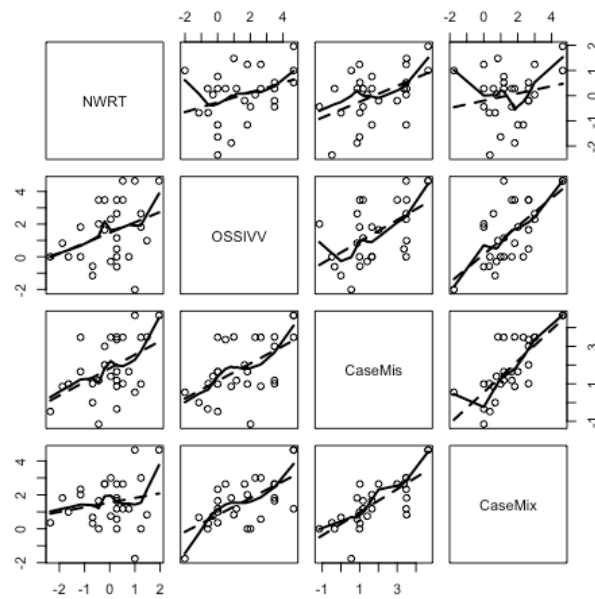


Figure 7. Scatterplot Matrix: NWRT, OSSIVV, CaseMis, and CaseMix at the Immediate Posttest

regression model was built, regressing the binary accuracy scores (i.e., correct or incorrect) on the confidence level. All of the assumptions for logistic regression analyses (i.e., binary outcome score, independence of observation, lack of multicollinearity, and an adequate sample size) were met. Table 9 summarizes standardized coefficients, standard errors, z-values, and significance values in the model.

Table 9. *Results of Logistic Regression on Confidence Ratings at the Immediate Posttest*

	Coefficients	SE	z-value	p
Intercept	- 0.285	.103	- 2.772	.000
Confidence	0.214	.030	7.085	.000*
Confidence for OSV	0.152	.064	2.375	.017*
Confidence for OSIV	0.060	.060	0.990	.322
Confidence for OSSVV	0.073	.063	1.152	.249
Confidence for OSSIVV	0.109	.059	1.848	.064
Confidence for CaseMis	- 0.027	.078	- 0.350	.726
Confidence for CaseMix	- 0.367	.076	- 4.783	.000*
				$R^2 = .077$

Note. * indicates a significant result.

Three coefficients in the model turned out to be significant. First, the overall level of confidence was positively related to the accuracy, and an increase in the confidence level by 1 unit was associated to 5.29% increase in probability of getting an item correct, compared to the grand mean (i.e., intercept). Second, the confidence level was also positively related to the participants' performance on OSV word order items, which was associated to 3.74% increase in probability of getting an item

correct when the confidence level increased by 1 unit. Lastly, and surprisingly, the confidence level was negatively related to the participants' accuracy scores on case mixing violation items. As the participants' confidence level increased by 1 unit, the probability of providing a correct answer on case mixing violation decreased by 8.67%. This suggests that the participants might have developed an incorrect rule about the case marking system in Japlish, which will be further discussed in Section 5.3. Figure 8, 9, and 10 graphically present the relationship between the confidence level and the accuracy score for the significant independent variables.

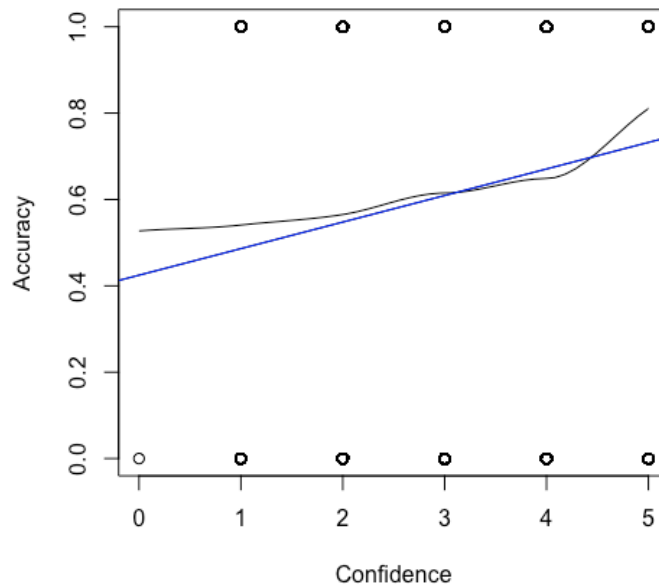


Figure 8. Confidence Level and Overall Accuracy at the Immediate Posttest

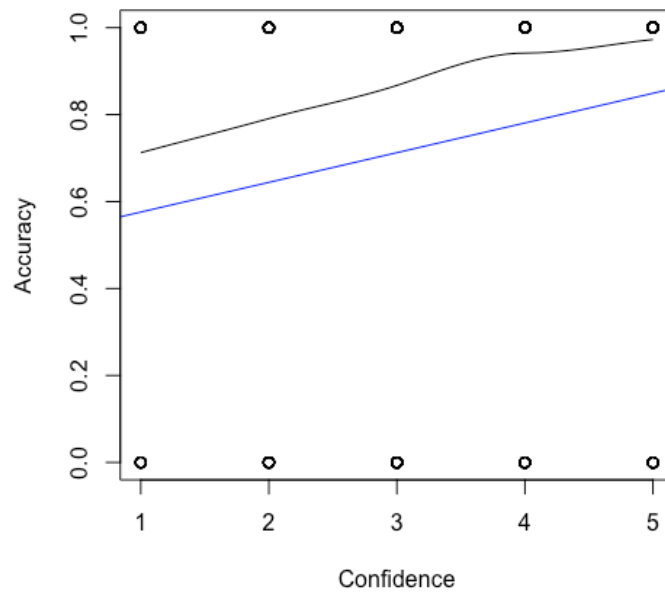


Figure 9. Confidence Level and OSV Accuracy at the Immediate Posttest

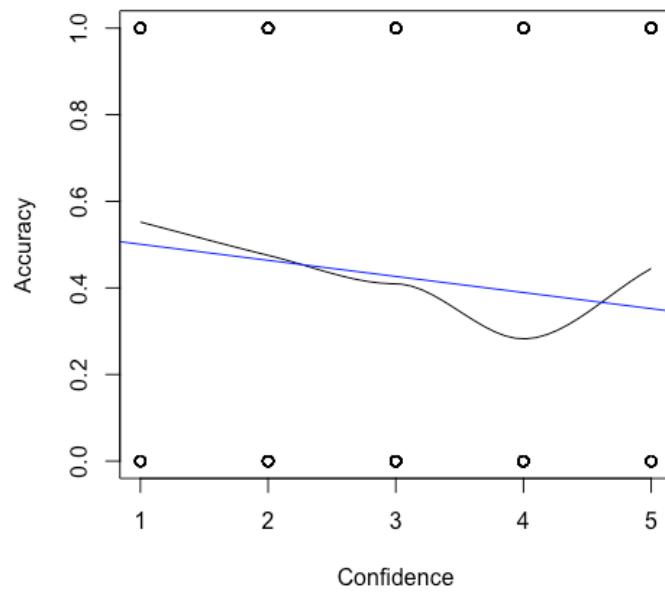


Figure 10. Confidence Level and CaseMix Accuracy at the Immediate Posttest

For the analysis of the source attributions, a mean proportion of correct answers was calculated for each source attribution category on each construction type. Table 10 summarizes the results. The overall picture presented in the table looks quite complex, but seems to show four patterns. First, Rule as the basis of knowledge was robustly correlated with the accuracy scores for most of the structures. Second, Memory (of exposure items) was also related to the above-chance level performance, but seem to be constrained to the word order constructions only. Third, the participants performed above chance when they claimed to be drawing on their intuition, but the results show that this largely worked for the simple word order types. Lastly, when the participants based their responses on a complete guess, they largely performed as well as or lower than the chance level, except OSSVV, for which they performed significantly above chance.

Table 10. *Mean Proportion of Accuracy for Each Source Attribution Category at the Immediate Posttest*

	Overall	OSV	OSIV	OSSVV	OSSIVV	CaseMis	CaseMix
Guess	.53	.51	.51	.61*	.50	.39	.58
SD	.50	.50	.50	.49	.50	.50	.50
N	477	61	80	122	135	31	48
chance	.49	.52	.50	.49	.50	.52	.52
Intuition	.53*	.69**	.60**	.59*	.50	.36	.36*
SD	.50	.49	.49	.49	.50	.48	.48
N	1096	150	220	247	235	118	126
chance	.49	.53	.50	.51	.50	.49	.48
Memory	.61**	.71**	.67**	.67**	.59*	.37*	.37
SD	.49	.46	.47	.47	.49	.49	.49
N	856	168	173	186	175	81	73
chance	.50	.51	.51	.49	.48	.48	.47
Rule	.74**	.85**	.79**	.78**	.73**	.59*	.45
SD	.44	.36	.41	.42	.45	.49	.50
N	1363	380	284	207	211	148	133
chance	.52	.50	.49	.51	.51	.51	.50

Note. * indicates a significant result. ** indicates a significant result even after Bonferroni Correction (.05/28= .0017). *N* indicates the number of data points. *chance* indicates the chance level generated by randomly generated scores.

5.1.4 Word-Monitoring Task

Mean monitoring latencies and GSIs for the experimental and the control groups are summarized in Table 11. Glancing through the mean RTs, the experimental and the control groups seem to differ in terms of the overall monitoring latencies, regardless of grammaticality. In order to investigate the extent to which the groups significantly differed, a MANOVA analysis was conducted. First, RTs that exceeded $\pm 3SD$ of the group mean were excluded¹³. The Shapiro-Wilk Test for normality for the experimental group reached significance, $W(336) = 0.986, p < .002$, Skewness = 0.3 ($SE = .133$), and Kurtosis = - 0.45 ($SE = .266$), indicating that the data are not

Table 11. *Mean Word-Monitoring Latencies at the Immediate Posttest*

	Overall	OSV	OSIV	OSSVV	OSSIVV	Case Mis	Case Mix
Experimental							
Overall							
<i>Mean</i>	571.53	586.01	522.86	558.44	547.45	623.71	590.71
<i>SD</i>	176.71	167.93	188.29	169.94	181.18	151.87	187.73
Grammatical							
<i>Mean</i>	575.17	596.59	536.98	568.67	543.01	608.10	597.63
<i>SD</i>	195.32	175.32	202.68	179.00	210.54	164.85	236.32
Ungrammatical							
<i>Mean</i>	567.89	575.43	508.73	548.21	551.89	639.33	583.78
<i>SD</i>	156.42	162.72	175.30	163.01	150.02	138.94	125.87
Control							
Overall							
<i>Mean</i>	611.38	620.75	563.13	599.32	569.65	701.27	614.18
<i>SD</i>	311.12	297.49	283.69	311.32	300.39	346.87	311.63
Grammatical							
<i>Mean</i>	616.82	651.42	593.91	613.45	556.83	671.23	614.09
<i>SD</i>	317.87	306.46	312.01	342.46	301.08	353.02	315.06
Ungrammatical							
<i>Mean</i>	605.94	590.07	532.36	585.19	582.46	731.31	614.26
<i>SD</i>	305.40	292.46	256.20	285.12	306.58	346.61	335.86

¹³ For the experimental group one data point and for the control group zero data point was excluded from the analysis.

normally distributed. However, this significance result could have been primarily caused by the fact that the sample size was quite large, $df = 336$. Thus, the z-score of skewness and kurtosis was each calculated, and the result showed that neither of them exceeded ± 2 range. Again, the Shapiro-Wilk Test for the control group also reached significance, $W(252) = 0.962, p < .000$, Skewness = 0.64 ($SE = .153$), and Kurtosis ($SE = .306$), and this time z-score of Skewness ($0.64/0.153 = 4.18$) exceeded. Thus, the distribution of the transformed RTs (TransRTs) was positively skewed¹⁴. Figure 11 presents a histogram of the distribution. Second, Levene's Test for homogeneity of variance showed that the group did not significantly differ in their variance ($p = .213$)

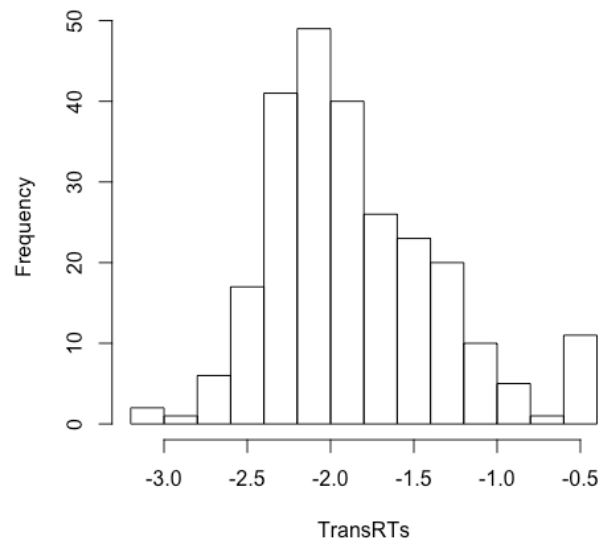


Figure 11. Histogram of TransRTs for the Control Group at the Immediate Posttest

As the main analysis, a MANOVA on the TransRTs showed that there was a significant main effect of Grammaticality at the multivariate level, $F(1, 88) = 2.917, p = .012$; Wilk's $\Lambda = .834, \eta_p^2 = .166$, but the main effect of Group and the interaction

¹⁴ To reach the complete normality was impossible here, as one of the very strong transformation of the data has already been applied (i.e, reciprocal transformation).

of Group and Grammaticality were not significant, $F(1, 88) = .557, p = .763$, Wilk's $\Lambda = .963, \eta_p^2 = .037$; $F(1, 88) = 207, p = .974$, Wilk's $\Lambda = .986, \eta_p^2 = .014$, respectively. This indicated that the experimental group and the control group did not differ in terms of their word-monitoring latencies nor did they differ in terms of the grammaticality effect (i.e., detecting an ungrammatical element in a sentence). This was a surprising result as the two groups seemed to differ significantly in their overall RTs confirmed in Table 11. Although the main effect of Grammaticality was significant, follow-up ANOVAs did not show any significant results.

5.1.5 The Role of Cognitive Aptitudes

Although no significant differences were found between the experimental and the control groups in their monitoring latencies, it is still possible to identify some participants who acquired implicit sensitivity to violation of word orders and case markings at the individual level. A recent line of SLA research claims that the group-level analysis often fails to capture learning at the individual level (e.g., Ellis & Larsen-Freeman, 2006; Lowie & Verspoor, 2015; Murakami, 2016; Murakami & Alexopoulou, 2016; Tanner, Inoue, & Osterhout, 2014). Thus, a MANCOVA analysis was still conducted, only with the experimental group, to investigate whether there was a relationship between the participants' grammatical sensitivity in Japlish and their individual differences in the explicit and implicit cognitive aptitudes. First, data points that exceeded $\pm 3SD$ of the experimental group's mean were excluded. This resulted in the exclusion of two data points (out of 170). Second, the assumption of homoscedasticity was checked by plotting residuals of the model against its predicted values. Figure 12 shows that the data are spread in a wide range. Although there are

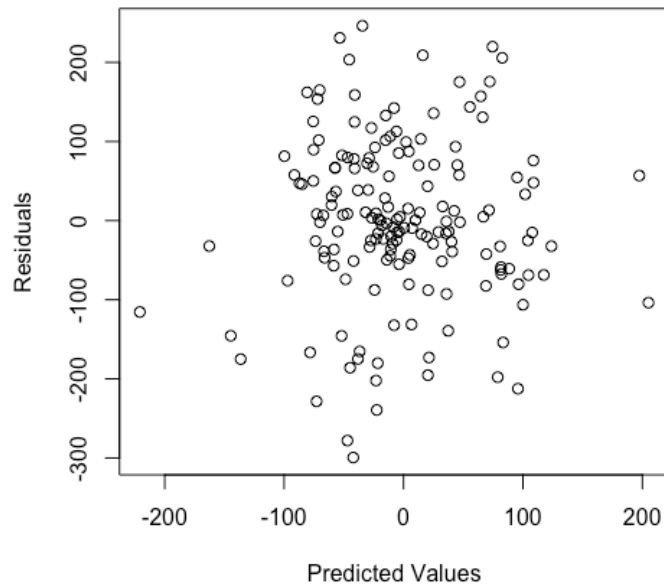


Figure 12. Scatterplot of the Residuals and the Predicted Values on WMT at the Immediate Posttest

some convergences of the data at the extreme ends of the residuals (i.e., observed scores), it was concluded that the assumption was met in this analysis.

Results of the MANCOVA showed that a main effect of SRT as a covariate approached significance at the multivariate level, $F(6, 17) = 2.255, p = .088$; Wilk's $\Lambda = .557, \eta_p^2 = .443$, but no other cognitive aptitude measures reached (or approached) significance, $F(6, 17) = .682, p = .667$, Wilk's $\Lambda = .867, \eta_p^2 = .194$ for LLAMA-F; $F(6, 17) = 1.485, p = .242$, Wilk's $\Lambda = .656, \eta_p^2 = .344$ for LSPAN; $F(6, 17) = 1.848, p = .149$, Wilk's $\Lambda = .605, \eta_p^2 = .395$ for NWRT. Since it was suspected that the small sample size ($n = 28$) could have affected the results, follow-up ANOVAs were conducted, and the results showed that there was a significant relationship between the GSIs on OSSIVV and the scores on NWRT, $F(1, 17) = 4.511, p = .045, \eta_p^2 = .170$, and the GSIs on the case missing violations items and the scores on SRT, $F(1,$

17) = 11.422, $p = .003$, $\eta_p^2 = .342$. The correlation coefficients for those two relationships were $r = .409$, $p = .034$, and $r = -.503$, $p = .006$, respectively. Note that although these two correlations were significant at the univariate level, their overall effect at the multivariate level was not significant, which can run the risk of producing a Type I error. Thus, the results must be interpreted with caution.

5.1.6 Summary of Immediate Posttests

The results from the immediate posttests can be summarized in four main findings. First of all, the experimental group outperformed the control group for all of the construction types on the untimed AGJT. This suggests that the participants acquired conscious, explicit knowledge about the word order and case marking patterns in Japlish and highlights the power of adults' explicit cognitive learning capacity even under an incidental condition, where the participants were inductively exposed to the exemplar sentences. Furthermore, subsequent post-hoc pairwise comparisons of their performance revealed that this learning mechanism was particularly effective for the simple word order types, especially OSV word order, the simplest of all.

Second, the analysis of the confidence ratings and the source attributions attested a facilitative nature of conscious awareness. The confidence levels that the participants reported significantly correlated with the accuracy scores on the AGJT, and this was most effective for the OSV word order type. This finding can be coupled with the source attribution data, which suggested that the participants were more accurate when they were drawing on the rules that they had formulated about the language. However, the participants also performed above chance for those

construction types to which their confidence level was not related. Moreover, they also performed significantly above chance when they claimed to be drawing on their intuition or a complete guessing, satisfying the zero-correlation and the guessing criterion, which implicates an existence of unconscious implicit knowledge (or unconscious structural knowledge, according to Dienes & Scott, 2005).

Third, the experimental and the control groups did not significantly differ in their word-monitoring latencies across the construction types and grammaticality. Although the raw mean RTs seemed to be shorter for the experimental group, this was not borne out in the inferential statistics. More importantly, the experimental did not differ from the control group on the WMT in RT for grammatical vs. ungrammatical items. Hence, the results here differ from attribution data, which suggested some acquisition of implicit knowledge. The disparity between the subjective measures of awareness and the objective measures of knowledge will be further discussed in Chapter 6.

Lastly, the participants' PSTM capacity significantly moderated their performance on both AGJT and WMT. In the former task, which allowed for controlled processing of the language, PSTM was involved with learning of simple construction types, OSIV and case marking system. In the latter task, however, it was with a complex word order type, OSSIVV, that PSTM was significantly and positively correlated. Yet, the other cognitive aptitudes measured were found to be mostly unrelated, which replicated the results of Brooks and Kempe (2013), Grey, Williams, and Rebuschat (2015), and Tagarelli, Borges-Mota, and Rebuschat (2011).

5.2 Delayed Posttests

5.2.1 Untimed Auditory Grammaticality Judgment Task

Mean percentile accuracy and d-prime scores of the experimental and the control groups at the delayed posttest phase are summarized in Table 12 and 13. As was the case in the immediate posttest, the experimental group outperformed the control group in all respects, except the mean percentile accuracy of grammatical OSV items (the experimental group, 85.71, and the control group, 93.45). This was due to the fact that both groups were more inclined to accept sentences than to reject them, and this tendency was especially robust for the control group. The exemplar sentences that the control group were exposed to had word orders and positions of case markers pseudo-randomized. Thus, it was unsurprising for them to incorrectly

Table 12. *Mean Percentile Accuracy for the Untimed AGJT at the Delayed Posttest*

	Overall	OSV	OSIV	OSSVV	OSSIVV	Case Mis	Case Mix
Experimental							
Overall							
<i>Mean</i>	69.12	78.73	74.55	69.61	68.78	74.63	71.68
<i>SD</i>	19.17	28.26	32.33	28.19	33.11	30.41	30.63
Grammatical							
<i>Mean</i>	83.15	85.71	91.51	74.93	81.25	91.51	91.51
<i>SD</i>	17.19	19.75	11.81	26.16	29.76	11.81	11.81
Ungrammatical							
<i>Mean</i>	60.50	71.74	57.58	64.28	56.31	57.75	51.84
<i>SD</i>	24.30	33.70	37.32	29.60	32.01	33.96	30.90
Control							
Overall							
<i>Mean</i>	50.18	66.96	61.61	55.65	50.29	51.79	55.65
<i>SD</i>	16.02	40.84	38.09	33.50	29.67	39.58	38.97
Grammatical							
<i>Mean</i>	77.09	93.45	83.93	72.02	59.52	83.93	83.93
<i>SD</i>	16.61	9.37	23.09	24.01	27.92	23.09	23.09
Ungrammatical							
<i>Mean</i>	32.24	40.47	39.28	39.29	41.07	19.64	27.38
<i>SD</i>	23.46	43.10	37.35	34.07	29.09	22.55	30.00

Table 13. *d*-prime Scores on the Untimed AGJT at the Delayed Posttest

	Overall	OSV	OSIV	OSSVV	OSSIVV	Case Mis	Case Mix
Experimental							
<i>Mean</i>	1.64	2.50	2.09	1.69	1.59	2.03	1.73
<i>SD</i>	1.45	1.88	1.98	1.70	1.89	1.79	1.59
<i>Max</i>	4.06	4.65	4.65	4.65	4.65	4.65	4.65
<i>Min</i>	- 0.51	- 0.67	-1.18	0.88	- 2.33	- 1.18	- 1.94
Control							
<i>Mean</i>	0.27	1.23	1.13	0.51	0.16	0.79	0.87
<i>SD</i>	0.83	1.79	1.90	1.67	1.34	1.26	1.52
<i>Max</i>	1.65	4.65	4.65	4.65	3.48	4.65	4.65
<i>Min</i>	- 0.96	- 1.18	- 1.82	- 1.65	- 2.64	- 0.67	- 1.65

accept the ungrammatical sentences because they should have learned the language to be random.

Before checking the MANOVA assumptions, *d*-prime scores that exceeded $\pm 3SD$ from the corresponding group mean were excluded. This resulted in exclusion of one data point from the control group. First, the Shapiro-Wilk Test for the experimental group's *d*-prime scores showed a significant test statistic, which indicates that the data did not exactly follow the normal distribution, $W(168) = 9.49, p < .000$, Skewness = $-.22$ ($SE = .187$) and Kurtosis = $-.963$ ($SE = .373$). Although the *z*-score of skewness remained ± 2 range, that of Kurtosis did not ($z_{\text{skewness}} = 1.17$ and $z_{\text{kurtosis}} = 2.58$). This suggests that the distribution was flatter than the normal distribution. Furthermore, the Shapiro-Wilk Test for the control group's *d*-prime scores also showed a significant test statistic, $W(168) = 9.35, p < .000$, Skewness = $.755$ ($SE = .216$) and Kurtosis = $.163$ ($SE = .428$). The *z*-score of kurtosis remained

the range but that of skewness did not, indicating a positively skewed nature of the distribution distribution ($z_{\text{skewness}} = 3.49$ and $z_{\text{kurtosis}} = 0.38$, see Figure 13 and 14)

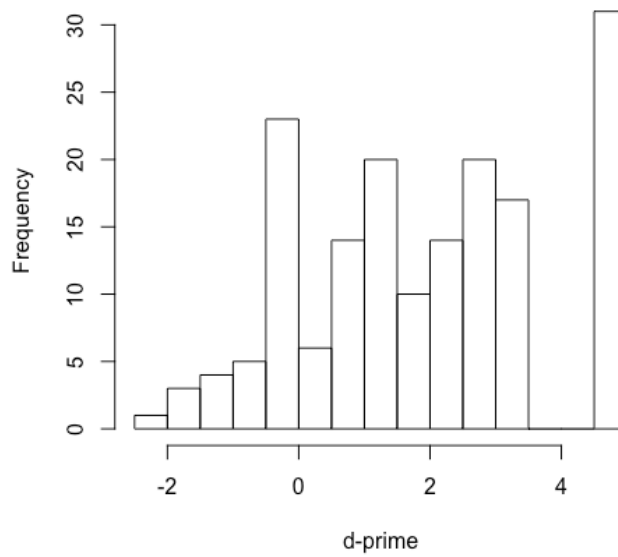


Figure 13. Histogram of d-prime Scores for the Experimental Group at the Delayed Posttest

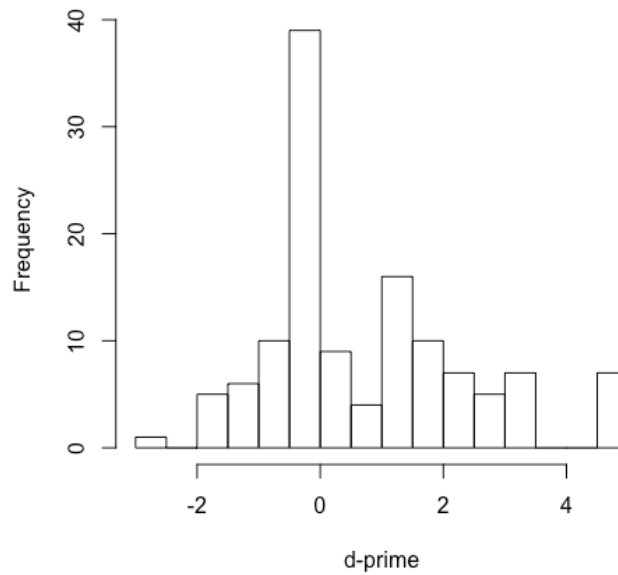


Figure 14. Histogram of d-prime Scores for the Control Group at the Delayed Posttest

Second, Levene's Test for homogeneity of variance reached significance, suggesting the groups differed in terms of their data variance, $F(1, 291) = 7.827, p =$

.005. However, as was the case at the immediate posttest, the difference across the groups was considered to be non-substantial, as the largest standard deviation was not three times larger than the smallest standard deviation between the groups¹⁵.

Results of a MANOVA on the d-prime scores mirrored that of the immediate posttest, indicating that there was a main effect of Group at the multivariate level, $F(6, 41) = 2.549, p = .034$; Wilk's $\Lambda = .728, \eta_p^2 = .272$. Furthermore, follow-up ANOVAs showed that after two weeks of no exposure to the language, the experimental group still outperformed the control group for all of the construction types, $F(1, 41) = 7.643, p = .008, \eta_p^2 = .142$ for OSV; $F(1, 41) = 4.252, p = .045, \eta_p^2 = .085$ for OSIV; $F(1, 41) = 7.962, p = .007, \eta_p^2 = .148$ for OSSVV; $F(1, 41) = 10.693, p = .002, \eta_p^2 = .189$ for OSSIVV; $F(1, 41) = 10.749, p = .002, \eta_p^2 = .189$ for CaseMis; $F(1, 41) = 4.926, p = .031, \eta_p^2 = .097$ for CaseMix.

In order to further examine differences between the experimental group's scores at the immediate posttest and the delayed posttest, a repeated-measures ANOVA was conducted for the experimental group only, with Construction and Time (i.e., immediate vs. delayed) as within-subjects factors¹⁶. The results showed that there was a main effect of Construction, $F(1, 3) = 8.210, p < .000, \eta_p^2 = .233$, and a main effect of Time, $F(1, 27) = 0.001, p = .980, \eta_p^2 = .000$, and the interaction of Construction and Time were not significant, $F(3, 98) = 0.887, p = .467, \eta_p^2 = .032$.

¹⁵ Note that any d-prime scores +/- 3SD of the corresponding group means were excluded. This resulted in an exclusion of two data points from the control group.

¹⁶ Note that Mauchly Tests for sphericity also reached significance for Construction and the interaction of Construction and Time, $\chi^2(5) = .352, p = .025$, and $\chi^2(5) = .364, p = .031$, respectively. Greenhouse-Geisser correction was made to the corresponding degree of freedoms.

Thus, the results confirmed the observation that the experimental group retained their knowledge after two weeks. Figure 15 shows the comparison of the experimental and the control groups (experimental: red, control: blue) and Figure 16 shows that of the immediate and the delayed posttest (immediate: yellow, delayed: green).

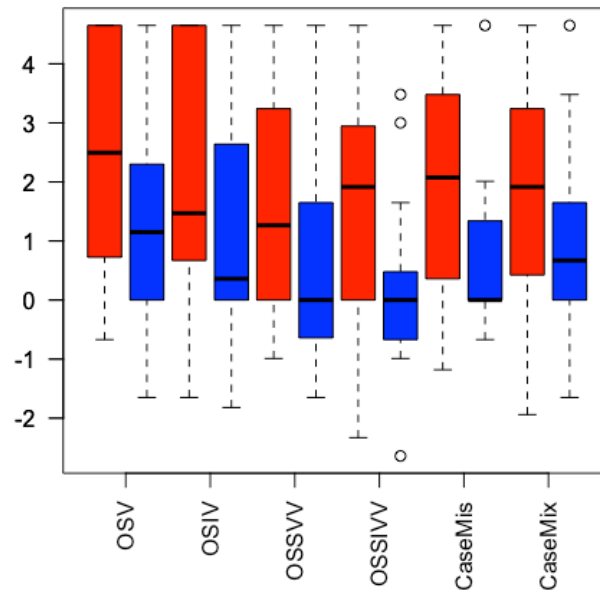


Figure 15. Boxplot of the AGJT at the Delayed Posttest

Lastly, since there was a main effect of Construction in both analyses, another repeated-measures ANOVA was conducted to examine which structures proved to be most difficult for the participants. Results showed that there was a main effect of Construction, $F(4, 96) = 3.076, p < .024, \eta_p^2 = .102$. Table 14 summarizes the results of post-hoc pairwise comparisons with Fisher's LSD test. The participants were more accurate on OSV items than they were on OSSVV, OSSIVV, and CaseMix items.

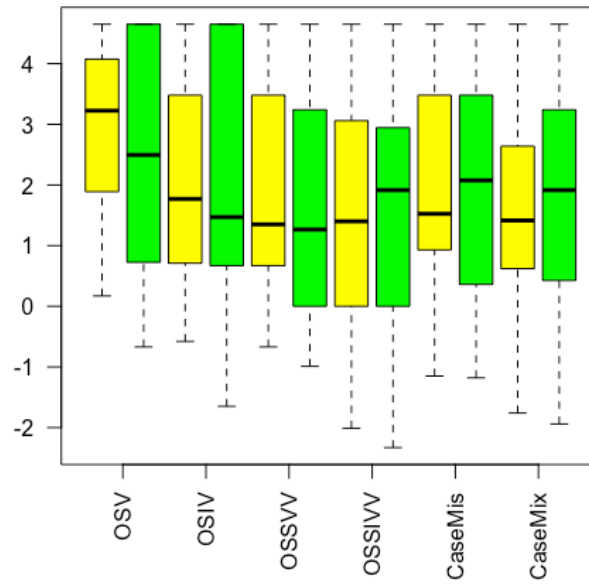


Figure 16. Boxplot for the Immediate and Delayed Posttest

Table 14. *Post-hoc Differences between the Construction Types at Delayed Posttest*

	Mean Difference	SE	p
OSV - OSSVV	.411	.229	.002
OSV - OSSIVV	.806	.364	.018
OSV - CaseMis	.770	.301	.017

5.2.2 The Role of Cognitive Aptitudes

First, the assumption of homoscedasticity was checked by inspecting a scatterplot of residuals and predicted values from the MANCOVA model (Figure 17). The data seem to be spread quite well, and there was no more than a minuscule convergence of the data observed. Results from the MANCOVA showed that in contrast to the results of the immediate posttest, none of the cognitive aptitudes significantly moderated the participants' performance on the AGJT, $F(6, 18) = 2.166$, $p = .095$, Wilk's $\Lambda = .419$, $\eta_p^2 = .419$ for LLAMA-F; $F(6, 18) = 2.166$, $p = .874$,

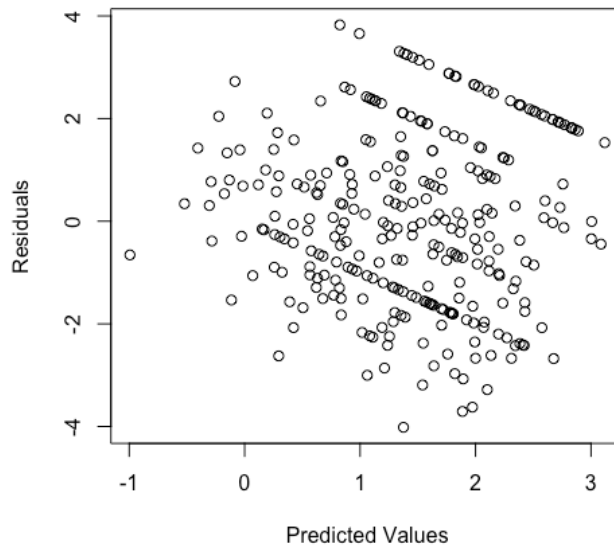


Figure 17. Scatterplot of the Residuals and the Predicted Values on AGJT at the Delayed Posttest

Wilk's $\Lambda = .884$, $\eta_p^2 = .116$ for LSPAN; $F(6, 18) = 0.654$, $p = .687$, Wilk's $\Lambda = .821$, $\eta_p^2 = .179$ for NWRT; $F(6, 18) = 0.710$, $p = .646$, Wilk's $\Lambda = .809$, $\eta_p^2 = .191$ for SRT. This could not have been due to small statistical power, as we have found a significant moderation effect of PSTM at the immediate posttest with the same sample size. Rather, this seems to suggest that retention of explicit knowledge that the participants developed under an incidental exposure condition is not significantly related to any cognitive aptitudes focused upon here.

5.2.3 The Role of Confidence Ratings and Source Attributions

As was done in the analysis of the immediate posttest, a multiple logistic regression model was built to examine the relationship between the participants' confidence level and their accuracy on the AGJT at the delayed posttest. Again, all of the assumptions specific to conducting multiple logistic regression analyses were

satisfied in the analysis here. Corresponding coefficients, standard errors, z-values, and *p*-values of the resulting model are listed in Table 15.

Table 15. *Results of Logistic Regression on Confidence Ratings at Delayed Posttest*

	Coefficients	<i>SE</i>	<i>z-value</i>	<i>p</i>
Intercept	- 0.186	.103	- 1.809	.000
Confidence	0.151	.028	5.249	.000*
Confidence x OSV	0.132	.060	2.188	.028*
Confidence x OSIV	- 0.019	.057	- 0.335	.322
Confidence x OSSVV	0.065	.058	1.115	.249
Confidence x OSSIVV	0.042	.057	0.744	.064
Confidence x CaseMis	- 0.016	.075	- 0.219	.726
Confidence x CaseMix	- 0.046	.017	- 2.677	.007*
				$R^2 = .066$

Note. * indicates a significant result.

The resulting model resembled the one at the immediate posttest quite well. Three independent variables were significantly predictive of the participants' performance on the delayed AGJT. First, the confidence level was positively correlated with the overall accuracy score, associated with a 3.76% increase in probability of getting a given item correct. Second, the confidence level was significantly related to the accuracy scores on the OSV items, which was associated with a 3.28% increase of the probability. Lastly, the level of confidence that the participants reported was negatively predictive of their performance on the case mixing violations, with a 1.19% decrease. Figure 18, 19, and 20, again, graphically presents the correlations.

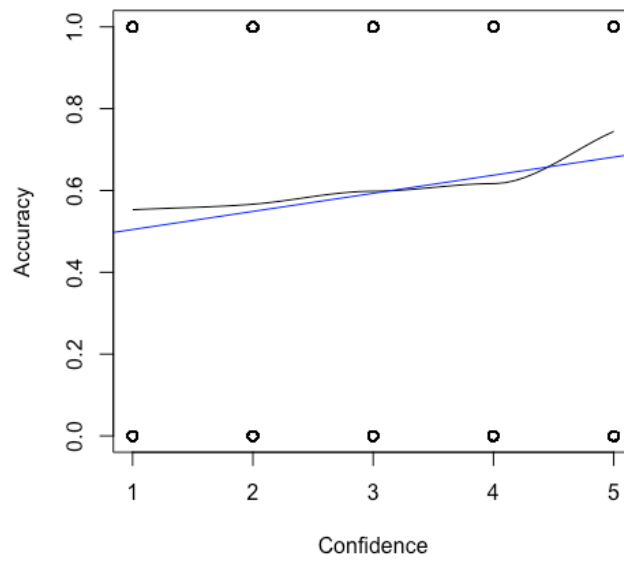


Figure 18. Confidence Level and Overall Accuracy at the Delayed Posttest

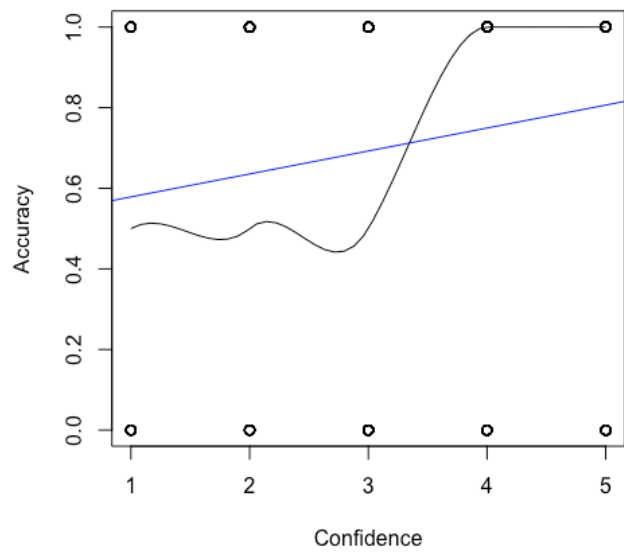


Figure 19. Confidence Level and OSV Accuracy at the Delayed Posttest

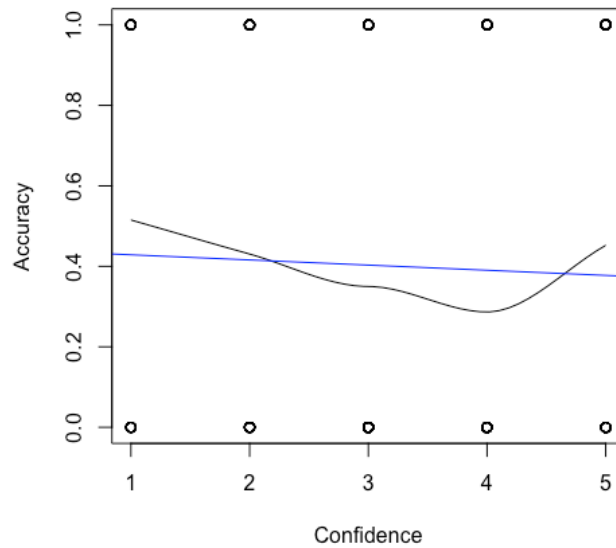


Figure 20. Confidence Level and CaseMix Accuracy at the Delayed Posttest

Further inspection of Figure 19 suggests that although a linear relationship of confidence and accuracy seems to be relatively small, the loess line reveals that most of the positive trend occurred between confidence levels, 3 (70-80%) and 4 (80-90%). Descriptive statistics show that the mean accuracy rate when the participants reported the level of confidence 3 was 69% ($SD = 46$), whereas it was 76% ($SD = 43$) when they reported the confidence level 4. Similarly, further inspection of Figure 20 suggests that while there is a negative trend overall (which even extends below chance level), the participants' accuracy was more or less at chance when they reported confidence level 5 (absolutely certain: 100%).

For the analysis of the source attributions, again, the mean proportion of correct answers was calculated for each source attribution category on each construction type. Table 16 lists the mean proportions and the corresponding result of t-tests. Three main findings stand out. First, the participants were particularly accurate

when they were drawing on rules that they had formulated about the language. However, this was not as effective for case mixing violation items, for which they performed as well with complete guessing as with the rules. Second, memory of exemplar items as the basis of knowledge only worked to some extent, especially for the simple word order types. Lastly, they performed above chance only on OSV items only when they were drawing on their intuitions about the language, but their accuracy was completely at chance or even statistically worse on the others items as well as when they were mere guessing.

Table 16. *Mean Proportion of Accuracy for Each Source Attribution Category at the Delayed Posttest*

	Overall	OSV	OSIV	OSSVV	OSSIVV	CaseMis	CaseMix
Guess	.52	.53	.47	.58	.52	.41**	.52
<i>SD</i>	.50	.50	.50	.50	.50	.50	.50
<i>N</i>	499	57	76	125	138	51	52
<i>chance</i>	.49	.51	.47	.52	.53	.51	.48
Intuition	.51	.70**	.57	.51	.53	.31**	.27**
<i>SD</i>	.50	.46	.50	.50	.50	.46	.45
<i>N</i>	955	136	185	220	214	101	99
<i>chance</i>	.51	.51	.50	.52	.52	.49	.50
Memory	.55*	.58*	.66**	.60*	.55	.33*	.33*
<i>SD</i>	.50	.50	.47	.49	.50	.47	.47
<i>N</i>	886	179	189	184	163	87	87
<i>chance</i>	.50	.49	.50	.51	.52	.48	.50
Rule	.74**	.85**	.79**	.78**	.75**	.85**	.48
<i>SD</i>	.44	.36	.41	.42	.43	.36	.50
<i>N</i>	1545	406	329	249	257	406	152
<i>chance</i>	.48	.50	.51	.53	.50	.49	.50

Note. * indicates a significant result. ** indicates a significant result even after Bonferroni Correction (.05/28= .0017). *N* indicates the number of data points. *chance* indicates the chance level generated by random scores.

5.2.4 Word-Monitoring Task

Mean monitoring latencies for the experimental and the control groups at the delayed posttest are summarized in Table 17. The two groups are comparable in terms

of their monitoring latencies. First, TransRTs that exceeded +/- 3SD of the group mean were excluded. This resulted in exclusion of three data points (out of 588). Second, the Shapiro-Wilk Test was conducted for both groups and neither occasion reached significance, $W(336) = 0.995, p = .42$, Skewness = - 0.160 ($SE = .134$), and Kurtosis = 0.190 ($SE = .266$) for the experimental group, and $W(336) = 0.995, p = .64$, Skewness = - 0.105 ($SE = .153$), and Kurtosis = - 0.317 ($SE = .306$) for the control group. Lastly, Levene's Test for homogeneity of variance also did not reach significance, $F(1, 583) = 0.514, p = .47$.

Table 17. *Mean Word-Monitoring Latencies at the Delayed Posttest*

	Overall	OSV	OSIV	OSSVV	OSSIVV	Case Mis	Case Mix
Experimental							
Overall							
Mean	539.26	545.15	470.40	535.37	500.55	635.41	548.67
SD	170.95	167.18	137.03	171.30	155.47	159.44	190.84
Grammatical							
Mean	536.98	553.34	472.27	542.37	507.89	604.96	541.03
SD	174.28	143.29	138.03	199.91	177.41	153.76	207.93
Ungrammatical							
Mean	541.54	536.96	468.52	528.36	493.20	665.86	556.30
SD	168.03	190.43	138.53	140.38	132.85	161.92	176.65
Control							
Overall							
Mean	541.21	525.31	475.70	509.45	508.54	657.23	571.02
SD	150.68	132.14	116.28	109.98	129.81	167.96	171.38
Grammatical							
Mean	532.00	521.58	487.66	511.67	523.17	604.34	543.54
SD	131.61	119.19	111.47	110.59	152.09	134.13	171.38
Ungrammatical							
Mean	550.43	529.05	463.75	507.22	493.91	710.13	598.50
SD	167.62	146.82	122.45	112.06	104.69	184.30	196.66

Results of the MANOVA analysis showed that a main effect of Grammaticality was significant at the multivariate level, $F(6, 86) = 2.927, p = .012$; Wilk's $\Lambda = .830, \eta_p^2 = .170$, but a main effect of Group and the interaction of Group

and Grammaticality were not significant, $F(6, 86) = .609, p = .723$, Wilk's $\Lambda = .959$, $\eta_p^2 = .041$; $F(6, 86) = .421, p = .863$, Wilk's $\Lambda = .971$, $\eta_p^2 = .029$, respectively. Follow-up univariate ANOVAs further showed that the main effect of Grammaticality was significant on the case missing violation items, $F(1, 86) = 7.787, p = .006, \eta_p^2 = .079$. This suggests that the two groups were comparable in their word-monitoring latencies as well as implicit sensitivity to grammatical violations, and both groups exhibited implicit grammatical sensitivity to case mission violation in the test sentences.

Since there was a significant effect of Grammaticality on the case mission violation items, another repeated-measures ANOVA was conducted for the experimental group and the control group separately. The results for the experimental group showed a significant main effect of Construction, $F(1, 25) = 19.285, p < .000, \eta_p^2 = .435$, but the interaction effect of Construction and Grammaticality only approached significance, $F(5, 125) = 2.075, p = .073, \eta_p^2 = .077$, and the main effect of Grammaticality was not significant, $F(5, 125) = .186, p = .670, \eta_p^2 = .007^{17}$. The same analysis was also conducted for the control group, and the results replicated those of the experimental group; the interaction effect of Construction and Grammaticality was not significant, $F(5, 100) = 2.198, p = .060$. As it was surmised that the non-significant interaction effect of Construction and Grammaticality could have been due to the small sample size for both groups, post-hoc pairwise comparisons were conducted between grammatical and ungrammatical items of each

¹⁷ Mauchly's Test for Sphericity reached significance only for Construction, $\chi^2(5) = .355, p = .048$. Greenhouse-Geisser correction was made to the degree of freedom.

construction type (Table 18 and Table 19). The difference between ungrammatical and grammatical sentences for case missing violation items was significant for both the experimental group and the control group, with a difference of 60.9 ms and 105.7 ms in raw RTs, respectively.

Table 18. *Post-hoc RT Difference between the Construction Types: Experimental*

	Difference (TransRTs)	SE	p
OSV	- 0.114	.076	.13
OSIV	0.005	.105	.95
OSSVV	- 0.023	.081	.76
OSSIVV	- 0.013	.092	.88
CaseMis	0.165	.083	.05*
CaseMix	0.101	.082	.21

Note. Comparison is based on Ungrammatical - Grammatical Items
* indicates a significant result.

Table 19. *Post-hoc RT Difference between the Construction Types: Control*

	Difference (TransRTs)	SE	p
OSV	0.022	.099	.81
OSIV	- 0.132	.108	.21
OSSVV	- 0.021	.080	.77
OSSIVV	- 0.058	.103	.56
CaseMis	0.232	.155	.001*
CaseMix	0.124	.088	.15

Note. Comparison is based on Ungrammatical - Grammatical Items
*indicates a significant result.

This suggests that the participants, although feebly, expressed an implicit sensitivity to a grammatical violation, when one of the case markers of the sentence was completely missing. The fact that not only the experimental group but also the control group showed the sensitivity to the case missing violation items seems puzzling at first, but a careful reflection upon the exposure materials suggested that it is indeed logical. That is, although the control group participants were exposed to the

exemplar sentences whose word orders and positions of case markers pseudo-randomized, it is also true that they never heard a sentence that was lacking a case marker. Hence, it can be expected that they learned the fact that the sentences carried three case markers, even without any understanding of their form-meaning relationships, and still showed the implicit sensitivity to the grammatical violation when the sentences were missing one of the case markers.

5.2.5 The Role of Cognitive Aptitudes

First, GSI data points that exceeded $\pm 3SD$ of the experimental group's mean were excluded. This resulted in exclusion of three data points (out of 171). Again, the assumption of homoscedasticity was first checked by inspecting a scatterplot of residuals and predicted values from the resulting MANCOVA model (Figure 21). The data seem to be spread quite well and unquestionably satisfy the assumption of homoscedasticity. Results of the MANCOVA model showed that none of the cognitive aptitudes significantly moderated the participants' implicit sensitivity to grammatical violations, $F(6, 18) = .576, p = .744$, Wilk's $\Lambda = .839, \eta_p^2 = .161$ for LLAMA-F; $F(6, 18) = .984, p = .465$, Wilk's $\Lambda = .753, \eta_p^2 = .247$ for LSPAN; $F(6, 18) = .994, p = .459$, Wilk's $\Lambda = .751, \eta_p^2 = .249$ for NWRT; $F(6, 18) = .308, p = .924$, Wilk's $\Lambda = .907, \eta_p^2 = .093$ for SRT. In contrast to the results of the immediate posttest, none of the cognitive aptitudes and their interactions with construction types turned out to be significant. These null results could not have been due to the small

sample size, a significant main effect was found for SRT on the GSIs at the multivariate level.

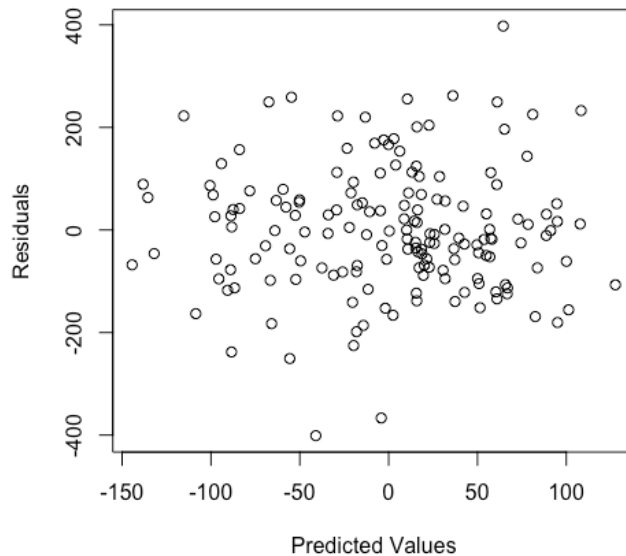


Figure 21. Scatterplot of the Residuals and the Predicted Values on AGJT at the Delayed Posttest

5.2.6 Summary for the Delayed Posttest

The results of the delayed posttest can be summarized in four points. First, as was the case at the immediate posttest, the advantage of the experimental group on the untimed AGJT was robust, surpassing the control group in all respects. This indicates retention of explicit knowledge that the participants developed from the incidental exposure, even after a two-week delay with no exposure to the language. Furthermore, the results also revealed that the participants performed as well at the delayed posttest as at the immediate posttest. Within the construction types, OSV items particularly proved to be more amenable to this learning, on which they performed better than on OSSVV, OSSIVV, and case mixing violation items, the complex word orders and the more difficult case marking violation type.

Second, the analysis of the confidence ratings and the source attributions mirrored that of the immediate posttest, in which explicit, conscious awareness about the language allowed the participants better and above-chance performances. In particular, the participants' confidence level was significantly and positively related to their overall performance and to OSV items. The source attributions showed that the participants performed significantly above chance when they were claiming to be drawing on rules as the basis of their knowledge. However, there were item types that were not related to the confidence level nor claimed to be based on rule knowledge, but rather, where the participants performed above chance-level, meeting the zero-correlation and the guessing criterion. Thus, although the performance on the AGJT suggests the robustness of explicit knowledge, the deeper analysis based on the subjective measures of awareness indicates some acquisition of implicit knowledge.

Third, the experimental group and the control group did not differ in their word-monitoring latencies on the WMT. However, there was a grammaticality effect for both groups on case-missing items, which suggests the emergence of automatic implicit sensitivity to the grammatical violation. The fact that such sensitivity was only found at the delayed posttest suggests that emergence of implicit knowledge (measured by the WMT) requires a certain time delay in order for the memory to consolidate, or that the certain time delay resulted in the decreased competition from explicit processes while performing the task.

Lastly, none of the cognitive aptitudes moderated performance on the AGJT and WMT, and this replicates the results at the immediate posttest.

5.3 Post-Experimental Verbal Reports

Lastly, this section discusses the results of analyzing retrospective verbal reports carried out at the end of the experiment. Table 20 presents the frequency and proportion of participants in each group who were assigned the category of *no report*, *noticing*, and *understanding*. Two simple patterns emerged for the experimental group. First, most of the participants (more than 85%) in the experimental group noticed or understood the simple word orders and the three case markers of Japlish. This was expected as their simple nature was hypothesized to allow the participants to consciously encode them in memory. Furthermore, more than one third of the experimental group participants reached the correct rules for those construction types, except the indirect case marker, “-*ni*”. Second, a smaller proportion of the participants noticed or understood the complex word order types, which were hypothesized to be difficult to encode explicitly. However, 64.3% of the participants still reported or referred to the existence of those structures, although the number of those who figured out the correct rule was much smaller.

Table 20. *Post-Experimental Retrospective Verbal Reports*

	OSV	OSIV	OSSVV	OSSIVV	- <i>ga</i>	- <i>o</i>	- <i>ni</i>
Experimental							
<i>No Report</i>	2 (7.1)	4 (14.2)	10 (35.7)	10 (35.7)	0 (0.0)	4 (14.2)	4 (14.2)
<i>Noticing</i>	11 (39.2)	13 (46.4)	15 (53.5)	13 (46.4)	11 (39.2)	13 (46.4)	19 (67.8)
<i>Understanding</i>	15 (53.5)	11 (39.2)	2 (7.1)	5 (17.8)	17 (60.7)	11 (39.2)	5 (17.8)
Control							
<i>No Report</i>	3 (14.2)	2 (9.5)	16 (76.2)	7 (33.3)	2 (9.5)	6 (28.6)	10 (47.6)
<i>Noticing</i>	10 (47.6)	15 (71.4)	4 (19.0)	13 (61.9)	15 (71.4)	11 (52.4)	7 (33.3)
<i>Understanding</i>	8 (38.1)	4 (19.0)	1 (4.8)	1 (4.8)	4 (19.0)	4 (19.0)	4 (19.0)

Turning to the control group, the pattern is quite similar regarding the simple word orders and the three case markers, except for the indirect case marker, the existence of which 47.6% of them did not report or refer to the construction. As opposed to the experimental group, a much smaller proportion of the control group participants reached the correct rules, as they were exposed to exemplar Japlish sentences whose word orders and positions of the case markers were pseudo-randomized. However, there was at least one person who reached the correct rule in each category. In particular, it was striking to find that one participant in the control group, whose second language was Spanish (with very low proficiency, i.e., a beginner), was assigned to the category of *understanding* for all of the construction types. This participant's overall mean accuracy on the AGJT was 65% at the immediate posttest and 96.25% at the delayed posttest. This suggests that even those participants in the control group were able to learn the language while they were taking the tests. However, as the participant only scored 65% at the immediate posttest, it might also be the case that he/she noticed that the language sounded like Japanese and learned about the word orders and case markings outside the study during the delay period. In any case, however, the results underscore the necessity to include a control group in study designs, in order to investigate the extent to which the experimental group truly learns a language from exposure.

In order to examine if the control group performed above chance, one-sample t-test was conducted on their d-prime scores, with results summarized in Table 21. Significantly above-chance performances of the control group were confirmed for their overall performance and the simple word orders at the immediate posttest, and

for their overall performance, the simple word orders, and the two case marking violation item types at the delayed posttest.

Table 21. *Results of One-Sample t-test for the Control Group*

	Overall	OSV	OSIV	OSSVV	OSSIVV	CaseMis	CaseMix
Immediate							
<i>M</i>	0.54**	1.36**	0.78*	0.14	- 0.17	0.56	0.57
<i>SD</i>	1.42	1.53	1.18	1.30	1.14	1.45	1.50
Delayed							
<i>M</i>	0.78**	1.23*	1.13*	0.51	0.16	0.79*	0.87*
<i>SD</i>	1.61	1.79	1.90	1.67	1.34	1.26	1.52

Note. * indicates a significant result. **indicates a significant result even after Bonferroni Correction (.05/14= .003). The chance level here was defined to be a d-prime score of 0.

Chapter 6: Discussion

6.1 Research Questions and Hypotheses

This chapter discusses eight research questions that were posited in Chapter 3. First, the corresponding hypotheses will be discussed with regard to whether they were supported by the results of the present study, and if they are, whether they can be explained by the rationales mentioned in Chapter 3.

- RQ 1 asked whether adult learners of L2 are able to learn simple word order types of Japlish, measured by two different types of L2 outcome measures at an immediate posttest, AGJT and WMT. Below are the three hypotheses posited in the current study and their status based on the results (notated as “A”):

H1a: For the simple word order types, the experimental group will significantly outperform the control group on the AGJT.

A1a: Hypothesis 1a is supported as the experimental group outperformed the control group for OSV and OSIV items. The hypothesis was confirmed as it was predicted in the reasoning presented, that for the simple types of word orders, participants notice the patterns and engage in explicit and intentional learning. The post-experimental retrospective verbal reports confirm the rationale, as most of the participants in the experimental group noticed their existence.

H1b: For the simple word order types, the subjective measures of awareness

on the AGJT will show that those who are aware of the knowledge they have acquired show better performance on the task.

A1b: Hypothesis 1b is mostly supported, as the confidence level that the participants reported significantly correlated with the accuracy on the task, and rule as the basis of knowledge contributed to the participants' significantly above-chance level performances. As with Hypothesis 1a, the rationale behind this hypothesis was the same, namely that the participants engage in explicit learning for these construction types.

H1c: For the simple word order types, the experimental group will *not* significantly differ from the control group on the WMT.

A1c: Hypothesis 1c is supported by the data. The two groups were comparable with regard to their monitoring latencies as well as grammatical sensitivity. This can be explained by the fact that the simple nature of the constructions allowed the participants in the experimental group to explicitly encode them in memory and also by the fact that the amount of exposure to the language was insufficient for implicit knowledge to emerge.

- RQ2 inquired if the same adult learners of L2 are able to learn the complex word order types of Japlish, measured by the two different types of L2 outcome measures at an immediate posttest:

H2a: For the complex word order types, the experimental group will significantly outperform the control group on the AGJT.

- A2a: Hypothesis 2a is supported by the results of the experiment. The experimental group significantly outperformed the control group for OSSVV and OSSIVV items. Originally, the hypothesis was posited based on the rationale that the task is not exclusive to measuring explicit knowledge alone and the participants thus can perform the task with implicit knowledge as well. It was expected that the complex nature of the two word order types do not enable the participants to consciously and actively encode them in memory, leading to implicit knowledge. However, an inspection of the entire data set suggests that the experimental group developed explicit knowledge, not implicit knowledge, and outperformed the control group with this type of knowledge.
- H2b: For the complex word order types, subjective measures of awareness suggest that there will be *no* systematic correlation between the level of confidence and the basis of knowledge the participants report.
- A2b: Hypothesis 2b is not supported by the data. Although the confidence level and the accuracy on the AGJT did not significantly correlate with one another (but reached significance for OSSIVV items), the results of the source attributions identified a robust connection between rule as the basis of knowledge and the accuracy scores on the task. With regard to the retrospective verbal reports, 64.3% of the participants in the experimental group at least consciously noticed the constructions. The fact that there was a strong connection between conscious

awareness and the accuracy score, revealed by the subjective measures, but that only one-third of the participants reported the existence of the complex word order types, indicated that the retrospective verbal reports underestimated the explicit knowledge that the participants developed of the complex patterns.

H2c: For the complex word order types, the experimental group will significantly differ from the control group on the WMT.

A2c: Hypothesis 2c is not supported by the results, as the experimental and the control groups did not differ in their word-monitoring latencies as well as grammatical sensitivity. As was the case for Hypothesis 1c, the amount of exposure sentences that the participants experienced was not sufficient for them to develop unconscious, implicit knowledge of the complex word order patterns. In light of this, the four methodologies used to detect explicit and implicit knowledge seem to differ in their sensitivity. In particular, the confidence ratings did not detect a facilitative nature of conscious awareness and the retrospective verbal reports showed that some portion of the participants did not notice the target constructions. While these two suggest acquisition of implicit knowledge, the results of the WMT obviously argue against that.

- RQ3 asked if adult learners of L2 are able to learn the morphological case markings of Japlish, measured by two types of L2 outcome measures:

- H3a: For the case markings, the experimental group will significantly outperform the control group on the AGJT.
- A3a: Hypothesis 3a is supported. The experimental group significantly outperformed the control group for the case missing and case mixing violation items. The rationale behind the hypothesis was the same as that of the simple word order types: that the participants would engage in explicit and intentional learning due to the salience of the constructions. Again, the retrospective verbal reports substantiate the rationale, as they suggest that most of the participants noticed the case markers (100%, 85.8%, 85.8%, for *-ga*, *-o*, and *-ni*, respectively).
- H3b: For the case markings, subjective measures of awareness on the AGJT suggest that those who are aware of the knowledge they have acquired will show better performance on the task; thus the zero-correlation and the guessing criterion will *not* be satisfied.
- A3b: Hypothesis 3b is partially supported by the data presented in Chapter 5. While the confidence level that the participants reported did not positively correlate with the accuracy of their performance, the responses from the source attributions clearly suggest the facilitative role of rule as the basis of knowledge, at least for case missing violation items. An interesting finding was the negative relationship that was observed for the case mixing violation items. One plausible explanation can be that the participants developed incorrect rules of the Japlish case markings. In any case, there is a notable difference

between the case missing and the case mixing violation items in the degree of learnability under an incidental exposure condition, which will be elaborated more in *General Discussion*.

H3c: For the case markings, the experimental group will not significantly differ from the control group on the WMT.

A3c: Hypothesis A3c is supported, as the experimental and the control groups did not differ in their word-monitoring latencies as well as grammatical sensitivity. As was the case above, the amount of exposure that the participants experienced was not sufficient for them to develop implicit knowledge of the case marking systems.

- RQ4 investigated which explicit and implicit cognitive aptitudes significantly moderate the learning of the simple word order types, measured by the two L2 outcome measures:

H4: The learning of the simple word order constructions indexed by the AGJT (but not the WMT) will be significantly moderated by subjects' WM capacity (i.e., the listening span) and their language analytic ability (i.e., LLAMA-F).

A4: Hypothesis 4 is not supported in the present study. A repeated-measures ANOVA on the participants' d-prime scores with the four cognitive aptitudes detected a significant interaction effect of NWRT and Construction. A subsequent correlational analysis identified a positive relationship between the participants' PSTM capacity and their performance on OSIV word order items. However, no other

cognitive aptitudes significantly moderated the learning of the simple word order types on the AGJT as well as WMT, which replicates results of the previous studies that learning under incidental conditions is not related to learner individual differences.

- RQ5 investigated which explicit and implicit cognitive aptitudes significantly moderate the learning of the complex word order types, measured by the two L2 outcome measures:

H5: The learning of the complex word order types indexed by the AGJT and WMT will be significantly moderated by subjects' implicit sequence-learning ability (i.e., SRT) and PSTM capacity (i.e., the nonword repetition).

A5: Hypothesis 5 is partially supported by the data. A repeated-measures ANOVA on the participants' GSIs detected a significant main effect of NWRT and an interaction of SRT and Construction. A subsequent correlational analysis identified a positive relationship between the participants' implicit grammatical sensitivity on OSSIVV items and their PSTM capacity. However, no other cognitive aptitudes significantly moderated the participants' learning of the complex word order types on the WMT as well as AGJT. It can be expected that for learning of the complex word orders, especially for OSSIVV, PSTM capacity was crucial as it affects the quality and quantity of phonological sequences that are held in the phonological loop, which eventually affects the quality of their mental representation. Although

it is yet unclear as to why NWRT was more predictive of GSIs than SRT, the nature of the two cognitive aptitudes provides a possible explanation. As reviewed in Chapter 2, PSTM is crucial for both implicit and explicit processes, whereas implicit sequence-learning ability is exclusively implicit. Given that the knowledge that was measured at the immediate posttest was mostly explicit, NWRT should be more correlated with the learning outcomes than SRT. However, as the WMT is a measure of implicit knowledge, it is yet unknown why the learning outcomes measured by the task correlated with the more explicit aptitude (i.e., NWRT).

- RQ6 asked which explicit and implicit cognitive aptitudes significantly moderate the learning of the case marking systems, measured by the two L2 outcome measures:

H6: The learning of case markings assessed by the untimed AGJT (but not the WMT) will be significantly moderated by WM capacity and LAA.

A6: Hypothesis 6 is not supported by the results of the experiment. As was the case with Hypothesis 4, a repeated-measures ANOVA on the d-prime scores showed a significant interaction effect of NWRT and Construction. Although a subsequent correlational analysis showed that the participants' performances on case missing violation items and their PSTM capacity was positively related, no other cognitive aptitude measures including LLAMA-F and LSPAN significantly moderated learning of the Japlish case markings, measured by the AGJT. Another

observation was that a repeated-measures ANOVA on the participants' GSIs showed a significant interaction of SRT and Construction. Again, a subsequent correlational analysis was conducted, and this revealed a negative relationship between the implicit sequence-learning ability and the grammatical sensitivity on case missing violation items.

- RQ7 was an exploratory question which investigated the extent to which results of the delayed posttest parallel or diverge from those of the immediate posttest. In particular, the question focused upon whether the findings on Research Question 1, 2, and 3 would still be observed for the delayed posttest. There were no prior experimental hypotheses posited for RQ7, but three general findings can be extrapolated:

A7a: On the AGJT, the participants performed as well at the delayed posttest as at the immediate posttest. At the delayed posttest, the experimental group still outperformed the control group in all of the construction types, and a subsequent ANOVA analysis showed that there was no significant difference between their performance at the immediate posttest and the delayed posttest (no interaction with Construction either). This suggests that knowledge learned under an incidental condition can be durable enough to be maintained over two weeks with no exposure to the language. Hence, the results replicate the previous research that found the durability of knowledge acquired in such conditions (Grey, et al., 2015; Robinson, 2002a). Furthermore, the comparison of the participants' performances at the immediate and

the delayed posttest (Figure 16) showed that for some construction types, they actually did better on the delayed posttest (i.e., OSSIVV, CaseMis, and CaseMix), which also replicates results of several studies on incidental/implicit learning (Ellis, Loewen, & Erlam, 2006; Li, 2010; Mackey, 1999; Mackey & Goo, 2007; Morgan-Short & Bowden, 2006; Morgan-Short, Finger, Grey, & Ullman, 2012).

- A7b: The analysis of the confidence ratings and the source attributions showed that the facilitative nature of conscious awareness was still associated with the accuracy of performance on the AGJT. The results at the delayed posttest almost completely mirrored those at the immediate posttest. The confidence level that the participants reported was positively predictive of their performance for overall performance and for OSV items. Furthermore, it was negatively related to accuracy scores on case mixing items, indicating that the participants carried over the incorrect rules of the case marking system to the delayed posttest. Lastly, there was a continuing effect of rule as the basis of knowledge, which, taken together, shows that the knowledge that the participants maintained over a two-week delay was mostly explicit.
- A7c: The analysis of the participants' word-monitoring latencies at the delayed posttest showed that they developed implicit sensitivity to grammatical violation of case missing. Interestingly, this was only observed for the delayed posttest, which suggests that development of implicit knowledge requires some time of memory consolidation or at

least time for the competition from explicit processes to diminish.

Thus, the results here underscore the necessity for any studies of incidental learning to have a delayed posttesting opportunity, so as to examine the delayed effects of incidental exposure. Yet, it must also be noted that such implicit sensitivity was only observed for the case missing violation items that tap into knowledge of pure form-form relationships.

- RQ8 was also an exploratory question, which asked whether the findings on Research Question 4, 5, and 6 (about individual differences) would parallel or diverge from results of the delayed posttest. One general finding can be extrapolated for RQ8:

A8: The results of the delayed posttest replicate those of the immediate posttest that learner individual differences, in particular cognitive aptitudes in the current study, might not be related to learning under incidental conditions. Although there was a significant positive relationship between PSTM capacity and the learning of Japlish word orders and case markings at the immediate posttest, this diminished to be non-significant at the delayed posttest.

6.2. General Discussion

The present study investigated the acquisition of explicit and implicit knowledge under an incidental condition, using various methodological approaches that have been adopted in the SLA literature (e.g., DeKeyser, 2012; Ellis, 2005; Rebuschat, 2013; Vafaei, Suzuki, & Kachinske, 2017; Williams, 2005). A synthesis

of the results in the present experiment casts light on four important and long-lasting debates in the field of SLA, and this last section of Chapter 6 serves to discuss any contribution the present study can make to those areas.

First of all, the present study identified clear discrepancies among the four methodological approaches in terms of their sensitivity to acquisition of explicit and implicit knowledge. The study has shown that the retrospective verbal reports tend to underestimate the explicit knowledge that the participants developed from incidental exposure, indicating some extent of implicit knowledge, while the WMT revealed no development of or very limited implicit grammatical sensitivity (i.e., only for case missing items at the delayed posttest). This confirms the claims by Hama and Leow (2010) and Shanks and St. John (1994) that offline verbal reports are an insensitive measure of conscious awareness in that they are subject to memory decay and fabrication and that participants do not report noticing or understanding of target features just because they are unable to verbalize them. In contrast, the WMT in the present study turned out to be a very strict measure of implicit knowledge, showing no or very little implicit acquisition.

Furthermore, there was a clear disagreement between the subjective measures of awareness and the objective measures of explicit and implicit knowledge. While the confidence ratings and the source attributions detected acquisition of both explicit and implicit knowledge for some construction types, this was not the case for the AGJT and WMT, in that they differed in terms of what they detected to be explicit or implicit knowledge. In particular, an interesting question can be asked as to why the two approaches diverged on the degree and the extent to which the participants

developed unconscious implicit knowledge. The present investigation provides two possible explanations. The first one is to posit that a WMT is a rigorous measure of pure implicit knowledge, whereas the subjective measures of awareness overestimate its acquisition (but see Rebuschat, 2013 on this issue). The second is to posit that the two approaches tap into different levels of implicit knowledge. The critical difference between the two tasks here was that while the participants afforded an unlimited time to perform the confidence ratings and the source attributions on the AGJT, their processing of the Japlish sentences had to be as fast as possible during the WMT. It is possible, then, to postulate that the two approaches measure implicit knowledge that differs in the degree of automaticity and memory entrenchment, and indeed, research in both cognitive psychology and SLA has long examined the development of automaticity in implicit learning, characterized as a data-driven memory-based process (e.g., Logan, 1988, 1992; Robinson, 1996, 1997; Perruchet & Pacteau, 1990). In this light, there needs to be future research which examines the veridicality of the two possibilities posited here and further investigates the validity of each approach.

Second, the results from the WMT at the delayed posttest demonstrated that implicit learning of linguistic constructions is indeed feasible for adult learners of L2, measured by a task which required automatic (and possibly, implicit) processing of incoming auditory stimuli. Word-monitoring latencies of both the experimental and the control groups slowed when they heard sentences with one of the case markers completely missing. The fact that the control group also showed such implicit sensitivity to grammatical violation can be explained by the design of the training material, in which they never heard sentences with any of the case markers being

omitted. Thus, it can be expected that not only the experimental group but also the control group can learn that nouns bear some kind of ending in the language, and they react to its absence based on the learning experience from the exposure phase.

If it is so then, the question must be asked as to the extent to which implicit learning from brief incidental exposure is feasible. In SLA literature, there has been an accumulating volume of research that investigated implicit learning by adults. This includes learning of syntax (Kachinske, Osthus, Solovyeva, & Long, 2015; Rebuschat, 2008; Rebuschat & Williams, 2009, 2012), morphology (Leung & Williams, 2011, 2012; Rogers, Révész, & Williams, 2016; Williams, 2005), morphosyntax (Grey, Williams, & Rebuschat, 2014), semantics (Paciorek & Williams, 2015), and suprasegmental stress regularity (Graham & Williams, 2016). The present study attempted to add an innovative investigation to the literature by examining whether adult learners of L2 are able to learn word orders and case marking system of a semi-artificial language, utilizing various methodological approaches to measurement of explicit and implicit knowledge. In the study, the analysis of the immediate and the delayed posttest found that the participants exhibited implicit grammatical sensitivity to the case-missing violation only. If this is the case, it must be concluded that implicit learning from brief incidental exposure is limited to acquisition of purely linguistic form knowledge. Indeed, the study is not the first one to demonstrate such results, as DeKeyser (1995), with 20 learning sessions, showed the limitation of implicit learning to form-form associations only (see also Godfroid, 2016 for similar results). The difference between the current study and the previous research lies in the fact that the present study is the first to adopt a WMT as a measure

of an automatic and implicit application of grammatical knowledge in an artificial or a semi-artificial language. To this end, the study calls for new studies of explicit and implicit learning under incidental conditions that attempt to triangulate explicit and implicit knowledge through various methodological approaches, including the use of WMT.

At the same time, SLA researchers have acknowledged for a long time that implicit unconscious L2 learning is likely to be a slow process and automatization of implicit knowledge takes a considerably longer time of language exposure than that of explicit rule instructions (Ellis, 1994; DeKeyser, 1995, 2003; Krashen, 1982; Robinson, 1996, 1997). Yet, almost all of the previous implicit learning studies (including the study reported here) have implemented their training sessions within a brief period of time. What is now crucial is to conduct a line of research that analyzes acquisition of explicit and implicit knowledge in a more longitudinal time frame. It comes as no surprise that many SLA researchers have already called for longitudinal analysis of language development, as ‘time’ is one of the most central variables in language learning research (Ortega & Iberri-Shea, 2005; Ortega & Byrnes, 2008). Indeed, these longitudinal analyses of developing implicit knowledge will likely to unearth the true potential of adults’ implicit learning capacity.

Third, the present investigation demonstrated that explicit learning is quite common under an incidental exposure condition, and moreover, the direct comparison of the immediate and the delayed posttest showed that the participants retained explicit knowledge that they developed of the language across the two testing sessions. This is clearly consistent with the previous research of explicit L2

instruction, which has repeatedly attested the powerful effect of explicit learning mechanisms (e.g., Goo, Granena, Yilmaz, & Novella, 2015; Norris & Ortega, 2000; Spada & Tomita, 2010). On the other hand, the results are also inconsistent with those of previous studies, in that there were no differences between the experimental group's performance at the immediate posttest and the delayed posttest, whereas previous studies often found that explicit knowledge declined after a delay with no exposure or no training.

All of this does not mean that the two learning modes, explicit and implicit, are mutually exclusive (see Bell, 2017 for an experimental study). While learners process the meaning of a L2 sentence and consciously reflect upon its meaning and structure, they also gain an instance of experience using the L2. Conceivably then, they can cultivate both explicit and implicit knowledge under incidental conditions, though the rate of development dramatically differs between the two. In actual classroom practice though, L2 teachers can act upon linguistic environments that surround learners, intending to affect learning *conditions*, but it is always beyond their capability to guarantee that the intended learning processes would happen. In order to make suggestions for improvements in L2 pedagogy possible, there needs to be an area of L2 classroom research that investigates the acquisition of explicit and implicit knowledge under various learning conditions that have been already put in practice.

Last but not least, the study demonstrated that explicit and implicit cognitive aptitudes did not significantly moderate learning of Japlish constructions under an incidental condition. Although the study replicates results of Brooke and Kempe

(2013), Grey, Williams, and Rebuschat (2015), and Tagarelli, Borges-Mota, and Rebuschat (2011) that learner individual differences are not related to acquisition of explicit and implicit knowledge from brief incidental exposure, it is yet to be explained as to why this is the case. SLA research has surely witnessed opposite cases where cognitive aptitudes were found to be clearly related to outcomes (Brooks, Kwoka, & Kempe, 2017; Robinson, 2005a), and this is a fruitful area of future research.

Chapter 7: Conclusions and Limitations

The study reported here attempted to explore one of the most complex, and rather complicated, areas of SLA, explicit and implicit learning and knowledge. Despite the insights that the study can contribute to the current literature, there are also some limitations of the experiment that must be explicitly recognized. First of all, it must be recognized that the sample size in the experiment is less than ideal. It might be true that some of the independent variables, especially the cognitive aptitudes, were in fact predictive of the participants' performances, but they failed to reach significance due to the limitation in statistical power. On the other hand, it should be emphasized that the sample size recruited here was noticeably more than that of previous studies, and the research design itself was an improvement, as the past studies rarely incorporated a trained control group in their design.

Second, it is admitted that the use of a semi-artificial language as of one of its limitations. The validity of a semi-artificial language has been already questioned by some researchers (Godfroid, 2016, but also cf. Rogers, Révész, & Rebuschat, 2016), namely that the conditions of learning semi-artificial languages might differ from those of natural L2 learning due to the increased saliency. The study employed a semi-artificial language, *Japlish*, because it allowed the experimental control of the participants' experience with the language. However, its use is also a double-edged sword, and it is fully admitted that the artificiality of the language could have affected the results of the experiment.

In the current literature, there are three possible options that L2 researchers can choose from as the target in a study of explicit and implicit learning: (a) natural

languages, (b) full artificial languages, and (c) semi-artificial languages. The decision of which option a researcher should go with depends on the nature and aims of the study to be conducted. In experimental settings, where participants' prior experiences with the target language needs to be carefully controlled, full and semi-artificial languages might be favored over natural languages. In the case of semi-artificial languages, in particular, researchers can bypass the need to train their participants on vocabulary items (but see 2.1.4 for inherent problems associated with the use of semi-artificial languages). It is a limitation of using artificial languages, however, that they can take learners only through the very initial stages of language learning (Ellis & Schmidt, 1998). In classroom settings, on the other hand, where ecological validity is prioritized over experimental control, natural languages might be more favored, and researchers can investigate learning at various levels, depending on learner proficiency.

Third, it is also worth indicating that the study was an experimental study conducted in a laboratory setting. It is known that language learning takes place in a dynamic environment, cognitively, linguistically, and socially situated. Language use is an embodied phenomenon, grounded in our cognitive capacities, highly social proclivities, and unique ways to experience and interact with the world (Tyler, 2012). Again, more ecologically sound research of explicit and implicit learning should be carried out in a classroom environment embedded in a particular context. Furthermore, the mode of language processing in the study was exclusively auditory, and every language test was carried out with the participants listening to Japlish sentences. However, language learning is a multimodal process (e.g., Vigliocco,

Perniss, & Vinson, 2014) and it is of interest to any SLA researcher to see a study that combines the two processing modes in explicit and implicit learning under incidental conditions.

Fourth but not least, the amount of exposure to the language that the participants received was far less than the ideal for them to develop implicit knowledge. In addition, the intensity of exposure to the target constructions was also unnatural in the sense that, in normal L2 learning settings, exposure is randomly distributed over a large period of time. To reiterate, the literature is in an acute need of research that conducts longitudinal and situated analyses of developing explicit and implicit knowledge under incidental conditions. It is a well-known fact that SLA studies are typically short and adopt testing materials that require conscious judgments; these study qualities are likely to bias the effectiveness of explicit learning and instruction (e.g., Doughty, 2003; Norris & Ortega, 2000). Unless the literature accumulates empirical research that spells out longitudinal development of explicit and implicit knowledge, the true potential of implicit learning cannot be accurately evaluated.

Having admitted these limitations, the last paragraph of this thesis again emphasizes the contribution that the present investigation can make to the advancement of knowledge in the field. Distinguishing explicit and implicit learning of L2 has been central to the understanding of learning processes underlying L2 acquisition. Therefore, this is the area of SLA that deserves most attention (and it probably does), and this is the area that we most urgently need to get a solid grip on. The study has revealed clear discrepancies of methodological approaches to

measurement of explicit and implicit knowledge and demonstrated the extent to which explicit and/or implicit learning would work under a brief incidental condition. Now that the study has identified gaps and issues in the explicit and implicit learning literature, it is the responsibility of future research to better the understanding of L2 learning processes and further examine the potential of explicit and implicit learning in SLA.

Appendices

Appendix A: Exposure Task Stimuli

Experimental Group

O-S-V word order (1-13: plausible, 14-25: implausible), 4 words

- (1) This bike-o John-ga bought
- (2) This wall-o Mary-ga painted
- (3) The cake-o Mike-ga ate
- (4) The ink-o Stacey-ga spilled
- (5) Those songs-o James-ga sang
- (6) The door-o Karen-ga broke
- (7) A homework-o Tim-ga finished
- (8) The fire-o Angela-ga lighted
- (9) That pen-o Tom-ga used
- (10) The video-o Cathy-ga watched
- (11) This chair-o Dwight-ga carried
- (12) A key-o Linda-ga dropped
- (13) Those words-o Steve-ga learned
- (14) A door-o Tom-ga drank
- (15) The dish-o Cathy-ga spoke
- (16) The girl-o Dwight-ga built
- (17) A phone-o Linda-ga wrote
- (18) The schools-o Steve-ga played
- (19) Those shoes-o Pamela-ga tasted
- (20) The book-o Jeff-ga inspired
- (21) A room-o Nicole-ga warned
- (22) The guitar-o John-ga peeled
- (23) A year-o Mary-ga moved
- (24) A letter-o Mike-ga offended
- (25) This bucket-o Stacey-ga soothed

O-S-IO-V word order (1-13: plausible, 14-25: implausible), 6 words

- (1) The picture-o John-ga his friends-ni sent
- (2) A letter-o Mary-ga her boss-ni faxed
- (3) This language-o Mike-ga his students-ni taught
- (4) A sweater-o Stacey-ga her husband-ni presented
- (5) His license-o James-ga the police-ni showed
- (6) The salt-o Karen-ga her brother-ni passed
- (7) A bone-o Tim-ga the dog-ni gave
- (8) A peace-o Angela-ga the government-ni demanded
- (9) A restaurant-o Tom-ga his parents-ni recommended
- (10) A letter-o Cathy-ga her boyfriend-ni wrote

Appendix A: Continued

- (11) This present-o Dwight-ga his friend-ni bought
- (12) Some money-o Linda-ga a bank-ni deposited
- (13) That computer-o Steve-ga his son-ni bought
- (14) A question-o Tom-ga the book-ni asked
- (15) Some money-o Cathy-ga a cat-ni lent
- (16) The secret-o Dwight-ga this table-ni told
- (17) A job-o Linda-ga that rocket-ni offered
- (18) A ball-o Steve-ga a stone-ni threw
- (19) This cookie-o Pamela-ga her bag-ni baked
- (20) The medal-o Jeff-ga the towel-ni awarded
- (21) A song-o Nicole-ga a tax-ni sang
- (22) A story-o John-ga the town-ni recited
- (23) A deal-o Mary-ga a virus-ni proposed
- (24) This flower-o Mike-ga the quality-ni brought
- (25) This food-o Stacey-ga a ladder-ni fed

O-S-[S-V]-V word order (1-8: plausible, 9-16: implausible), 7 words

- (1) The diamond-o John-ga this man-ga stole thought
- (2) The tuition-o Mary-ga her school-ga raised mentioned
- (3) The donuts-o Mike-ga his dog-ga ate realized
- (4) That vase-o Stacey-ga her spouse-ga broke said
- (5) His car-o James-ga a cyclist-ga damaged thought
- (6) That girl-o Karen-ga the police-ga found mentioned
- (7) His health-o Tim-ga the smoking-ga hurt realized
- (8) A gun-o Angela-ga a robber-ga shot said
- (9) Some milk-o Tom-ga his mother-ga added thought
- (10) The governor-o Cathy-ga her colleague-ga abused mentioned
- (11) Those kids-o Dwight-ga his friend-ga scared realized
- (12) This sushi-o Linda-ga that chef-ga made said
- (13) The job-o Steve-ga his father-ga quit thought
- (14) An emphasis-o Tom-ga the train-ga hit mentioned
- (15) The theater-o Cathy-ga a singer-ga played realized
- (16) An ability-o Dwight-ga the fire-ga burned said
- (17) An outcome-o Linda-ga her boss-ga fired thought
- (18) Those books-o Steve-ga the waitress-ga spilled mentioned
- (19) This paper-o Pamela-ga her colleague-ga drove realized
- (20) Those plants-o Jeff-ga the boy-ga drank said
- (21) That news-o Nicole-ga her teacher-ga reprimanded thought
- (22) A blanket-o John-ga his friend-ga flew mentioned
- (23) A book-o Mary-ga an elephant-ga drank realized
- (24) This lamp-o Mike-ga his teacher-ga started said
- (25) The fear-o Stacey-ga her daughter-ga loaded thought

Appendix A: Continued:

O-S-[S-IO-V]-V word order (1-13: plausible, 14-25: implausible), 9 words

- (1) Some food-o John-ga his wife-ga their dog-ni brought thought
- (2) The documents-o Mary-ga her workmate-ga their boss-ni faxed mentioned
- (3) This coffee-o Mike-ga a student-ga the professor-ni brought realized
- (4) A present-o Stacey-ga the boy-ga his mother-ni sent said
- (5) The book-o James-ga his sister-ga her friend-ni lent thought
- (6) A bribe-o Karen-ga her colleague-ga their boss-ni offered mentioned
- (7) A tip-o Tim-ga his friend-ga the driver-ni gave realized
- (8) A letter-o Angela-ga her husband-ga the mayor-ni wrote said
- (9) A necklace-o Tom-ga his friend-ga his wife-ni gave thought
- (10) That car-o Cathy-ga her husband-ga the neighbor-ni sold mentioned
- (11) The secret-o Linda-ga her boyfriend-ga his mother-ni told realized
- (12) Some money-o Dwight-ga his uncle-ga a man-ni spared said
- (13) The pizza-o Steve-ga his friend-ga a roommate-ni saved thought
- (14) A car-o Tom-ga his parents-ga the beauty-ni bought mentioned
- (15) A bike-o Cathy-ga her aunt-ga a taste-ni presented realized
- (16) A waitress-o Dwight-ga his wife-ga his son-ni read said
- (17) A medal-o Linda-ga the president-ga the future-ni awarded thought
- (18) An honor-o Pamela-ga the king-ga a relief-ni bestowed said
- (19) The sorrow-o Steve-ga his manager-ga his friend-ni proposed realized
- (20) A misery-o Jeff-ga the student-ga his teacher-ni asked said
- (21) A victory-o Nicole-ga her father-ga the police-ni showed thought
- (22) A number-o John-ga the waitress-ga a customer-ni offered mentioned
- (23) A letter-o Mary-ga her colleague-ga the patience-ni wrote realized
- (24) The book-o Mike-ga his professor-ga the movement-ni lent said
- (25) A conflict-o Stacey-ga her spouse-ga her son-ni bought thought

Control Group

O-S-V word order (1-13: plausible, 14-25: implausible), 4 words

- (1) John-ga this bike-o bought (S-ga O-o V)
- (2) Mary-ga painted this wall-o (S-ga V O-o)
- (3) Ate Mike-ga the cake-o (V S-ga O-o)
- (4) Spilled the ink-o Stacey-ga (V O-o S-ga)
- (5) Songs-o James-ga sang (O-o S-ga V)
- (6) The door-o broke Karen-ga (O-o V S-ga)
- (7) Tim-o a homework-ga finished (S-o O-ga V)
- (8) Angela-o lighted the fire-ga (S-o V O-ga)
- (9) Used Tom-o that pen-ga (V S-o O-ga)
- (10) Watched the video-ga Cathy-o (V O-ga S-o)
- (11) This chair-ga Dwight-o carried (O-ga S-o V)
- (12) A key-ga dropped Linda-o (O-ga V S-o)
- (13) Steve-ga those words-o learned (S-ga O-o V)

Appendix A: Continued:

- (14) Tom-ga drank a door-o (S-ga V O-o)
- (15) Spoke Cathy-ga the dish-o (V S-ga O-o)
- (16) Built the girl-o Dwight-ga (V O-o S-ga)
- (17) A phone-o Linda-ga wrote (O-o S-ga V)
- (18) The schools-ga played Steve-o (O-ga V S-o)
- (19) Pamela-o those shoes-ga tasted (S-o O-ga V)
- (20) Jeff-o inspired the book-ga (S-o V O-ga)
- (21) Warned Nicole-o a room-ga (V S-o O-ga)
- (22) Pealed the guitar-ga John-o (V O-ga S-o)
- (23) A year-ga Mary-o moved (O-ga S-o V)
- (24) A letter-ga offended Mike-o (O-ga V S-o)
- (25) Stacey-ga this bucket-o soothed (S-ga O-o V)

Frequency: SOV (5), SVO (4), VSO (4), VOS (4), OSV (4), OVS(4)
S-ga (12), S-o (13), O-ga (13), O-o (12)

O-S-IO-V word order (1-13: plausible, 14-25: implausible), 6 words

- (1) John-ga his friends-ni the picture-o sent (S-ga I-ni O-o V)
- (2) Mary-o her boss-ga faxed a letter-ni (S-o I-ga V O-ni)
- (3) Mike-ni this language-ga his students-o taught (S-ni O-ga I-o V)
- (4) Stacey-ga a sweater-o her presented husband-ni (S-ga O-o V I-ni)
- (5) James-o showed the police-ga his license-ni (S-o V I-ga O-ni)
- (6) Karen-ni passed her brother-ga the salt-o (S-ni V O-ga I-o)
- (7) The dog-ni Tim-ga a bone-o gave (I-ni S-ga O-o V)
- (8) The government-ga Angela-o demanded a peace-ni (I-ga S-o V O-ni)
- (9) His parents-o a restaurant-ga Tom-ni recommended (I-o O-ga S-ni V)
- (10) Her boyfriend-ni a letter-o wrote Cathy-ga (I-ni O-o V S-ga)
- (11) His friend-ga bought Dwight-o this present-ni (I-ga V S-o O-ni)
- (12) A bank-o deposited some money-ga Linda-ni (I-o V O-ga S-ni)
- (13) That computer-o Steve-ga his son-ni bought (O-o S-ga I-ni V)
- (14) A question-ni Tom-o asked the book-ga (O-ni S-o V I-ga)
- (15) Some money-ga a cat-o Cathy-ni lent (O-ga I-o S-ni V)
- (16) The secret-o this table-ni told Dwight-ga (O-o I-ni V S-ga)
- (17) Linda-o that rocket-ga a job-ni offered (S-o I-ga O-ni V)
- (18) Steve-ni a stone-o threw a ball-ga (S-ni I-o V O-ga)
- (19) Pamela-ga this cookie-o her bag-ni baked (S-ga O-o I-ni V)
- (20) Jeff-o the medal-ni awarded the towel-ga (S-o O-ni V I-ga)
- (21) Nicole-ni sang a tax-o a song-ga (S-ni V I-o O-ga)
- (22) John-ga recited a story-o the town-ni (S-ga V O-o I-ni)
- (23) A virus-ga Mary-o a deal-ni proposed (I-ga S-o O-ni V)
- (24) The quality-o Mike-ni brought this flower-ga (I-o S-ni V O-ga)
- (25) A ladder-ni this food-o Stacey-ga fed (I-ni O-o S-ga V)

Appendix A: Continued:

Frequency: SIOV(2), SIVO (2), SOIV (2), SOVI (2), SVIO (2), SVOI (2), ISOV (2), ISVO (2), IOSV (2), IOVS (1), IVSO (1), IVOS (1), OSIV (1), OSVI (1), OISV (1), OIVS (1).
S-ga (9), S-o (8), S-ni (8), I-ga (8), I-ni (9), I-o (8), O-ga (8), O-ni (8), O-o (9).

O-S-[S-V]-V word order (1-8: plausible, 9-16: implausible), 7 words

- (1) John-ga this man-ga the diamond-o stole thought (S-ga S-ga O-o V V)
- (2) Mary-o her school-o raised the tuition-ga mentioned (S-o S-o V O-ga V)
- (3) Mike-ga his dog-ga ate realized the donuts-o (S-ga S-ga V V O-o)
- (4) Stacey-o that vase-ga her spouse-o broke said (S-o O-ga S-o VV)
- (5) James-ga his car-o a damaged cyclist-ga thought (S-ga O-o V S-ga V)
- (6) Karen-o that girl-ga found mentioned the police-o (S-o O-ga V V S-o)
- (7) Tim-ga hurt the smoking-ga his health-o realized (S-ga V S-ga O-o V)
- (8) Angela-o shot a robber-o said a gun-ga (S-o V S-o V O-ga)
- (9) Tom-ga added some milk-o his mother-ga thought (S-ga V O-o S-ga V)
- (10) Cathy-o abused the governor-ga mentioned her colleague-o (S-o V O-ga V S-o)
- (11) His kids-o Dwight-ga a friend-ga scared realized (O-o S-ga S-ga V V)
- (12) This sushi-ga Linda-o made this chef-o said (O-ga S-o V S-o V)
- (13) The job-o Steve-ga quit thought his father-ga (O-o S-ga V V S-ga)
- (14) An emphasis-ga hit Tom-o the train-o mentioned (O-ga V S-o S-o V)
- (15) The theater-o played Cathy-ga realized a singer-ga (O-o V S-ga V S-ga)
- (16) An ability-ga burned said Dwight-o the fire-o (O-ga V V S-o S-o)
- (17) Fired Linda-ga her boss-ga an outcome-o thought (V S-ga S-ga O-o V)
- (18) Spilled Steve-o the waitress-o mentioned those books-ga (V S-o S-o V O-ga)
- (19) Drove Pamela-ga this paper-o her colleague-ga realized (V S-ga O-o S-ga V)
- (20) Drank Jeff-o those plants-ga said the boy-o (V S-o O-ga V S-o)
- (21) Reprimanded that news-o Nicole-ga her teacher-ga thought (V O-o S-ga S-ga V)
- (22) Flew a blanket-ga John-o mentioned his friend-o (V O-ga S-o V S-o)
- (23) Drank a book-o realized Mary-ga an elephant-ga (V O-o V S-ga S-ga)
- (24) Started said Mike-o his teacher-o this lamp-ga (V V S-o S-o O-ga)
- (25) Loaded thought Stacey-ga the fear-o her daughter-ga (V V S-ga O-o S-ga)

Frequency: SSOVV (1), SSVOV (1), SSVVO (1), SOSVV (1), SOVSV (1), SOVVS (1), SVSOV (1), SVSVO (1), SVOSV (1), SVOVS (1), OSSVV (1), OSVSV (1), OSVVS (1), OVSSV (1), OVSVS (1), OVVSS (1), VSSOV (1), VSSVO (1), VSOSV (1), VSOVS (1), VOSSV (1), VOSVS (1), VOVSS (1), VVSSO (1), VVSOS (1).
S-ga (26), S-o (24), O-ga (12), O-o (13)

O-S-[S-IO-V]-V word order (1-13: plausible, 14-25: implausible), 9 words

- (1) John-ga his wife-ga a dog-ni some food-o brought thought (S-ga S-ga I-ni O-o V V)
- (2) Mary-ni her workmate-ni her boss-o faxed the documents-ga mentioned (S-ni S-ni

Appendix A: Continued:

- I-o V O- ga V)
- (3) Mike-o a student-o the professor-ga this brought realized this coffee-ni (S-o S-o I-gaVVO-ni)
 - (4) Stacey-ga the boy-ga sent said his mother-ni a present-o (S-ga S-ga V V I-ni O-o)
 - (5) James-ni his sister-ni the book-ga lent thought her friend-o (S-ni S-ni O-ga V V I-o)
 - (6) Her boss-ga Karen-o her colleague-o a bribe-ni offered mentioned (I-ga S-o S-o O-ni V V)
 - (7) The driver-ni Tim-ga his friend-ga gave a tip-o realized (I-ni S-ga S-ga V O-o V)
 - (8) The mayor-o Angela-ni a letter-ga wrote said her husband-ni (I-o S-ni O-ga V V S-ni)
 - (9) His wife-ga Tom-o his A necklace-ni gave friend-o thought (I-ga S-o O-ni V S-o V)
 - (10) The neighbor-ni Cathy-ga sold mentioned her husband-ga that car-o (I-ni S-ga V V S-ga O- o)
 - (11) His mother-o Linda-ni told her boyfriend-ni the secret-ga realized (I-o S-ni V S-ni O-ga V)
 - (12) Some money-ni Dwight-o his uncle-o a man-ga spared said (O-ni S-o S-o I-ga V V)
 - (13) The pizza-o Steve-ga his friend-ga saved a roommate-ni thought (O-o S-ga S-ga V I-ni V)
 - (14) A car-ga Tom-ni the beauty-o bought mentioned his parents-ni (O-ga S-ni I-o V V S-ni)
 - (15) A bike-ni Cathy-o taste-ga a presented her aunt-o realized (O-ni S-o I-ga V S-o V)
 - (16) A waitress-o Dwight-ga read said his wife-ga his son-ni (O-o S-ga V V S-ga I-ni)
 - (17) A medal-ga Linda-ni awarded the president-ni the future-o thought (O-ga S-ni V S-ni I-o V)
 - (18) Bestowed Pamela-o the king-o a relief-ga an honor-ni said (V S-o S-o I-ga O-ni V)
 - (19) Proposed Steve-ga his friend-ni the sorrow-o realized his manager-ga (V S-ga I-ni O-o V S- ga)
 - (20) Asked Jeff-ni his teacher-o a misery-ga said the student-ni (V S-ni I-o O-ga V S-ni)
 - (21) Showed Nicole-o a victory-ni thought her father-o the police-ga (V S-o O-ni V S-o I-ga)
 - (22) Offered John-ga mentioned the waitress-ga a customer-ni a number-o (V S-ga V S-ga I-ni O-o)
 - (23) Wrote Mary-ni the patience-o realized her colleague-ni a letter-ga (V S-ni I-o V S-ni O-ga)
 - (24) Lent Mike-o the book-ni his professor-o the movement-ga said (V S-o O-ni S-o I-ga V)

Appendix A: Continued

(25) Bought Stacey-ga thought her son-ni a conflict-o her spouse-ga (V S-ga V I-ni O-o S-ga)

Frequency: SIOVV (1), SSIVOV (1), SSIVVO (1), SSVVIO (1), SSVIOV (1), SSOVVI (1), ISSOVV (1), ISSVOV (1), ISOVVS (1), ISOVSV (1), ISVVSO (1), ISVSOV (1), OSSIVV (1), OSSVIV (1), OSIVVS (1), OSIVSV (1), OSVVSI (1), OSVSIV (1), VSSIOV (1), VSIOVS (1), VSOVSI (1), VSVSIO (1), VSIVSO (1), VSOSI (1), VSVIOS (1)
S-ga (18), S-o (16), S-ni (16), O-ga (8), O-o (9), O-ni (9), I-ga (8), I-o (8), I-ni (9)

Appendix B: AGJT Stimuli

List 1

Grammatical

O-S-V (4 words)

- (1) This cake-o Daniel-ga favored
- (2) The table-o Sarah-ga wiped
- (3) The admission-o Nick-ga secured
- (4) A parcel-o Emma-ga received
- (5) The street-o Chris-ga crossed
- (6) These lessons-o Alison-ga studied
- (7) A dog-o Ben-ga had
- (8) The hairstyle-o Rachel-ga changed

O-S-I-V (6 words)

- (1) The application-o Bruce-ga a university-ni mailed
- (2) A party-o Vivian-ga her mother-ni threw
- (3) A lesson-o Richard-ga his students-ni gave
- (4) That ball-o Ellie-ga a player-ni tossed
- (5) An autograph-o Elvis-ga his fan-ni signed
- (6) This computer-o Daisy-ga her father-ni chose
- (7) That money-o Tony-ga a church-ni donated
- (8) The truth-o Lyla-ga her professor-ni told

O-S-[S-V]-V (7 words)

- (1) The exam-o Bill-ga his friend-ga passed said
- (2) These toys-o Claire-ga her child-ga wanted realized
- (3) Those kids-o Boris-ga that man-ga helped mentioned
- (4) Her heart-o Lydia-ga her boyfriend-ga destroyed thought
- (5) The war-o Scott-ga the soldiers-ga survived said
- (6) The professor-o Laura-ga her colleague-ga upset realized
- (7) His paper-o Will-ga the journal-ga accepted mentioned
- (8) Those students-o Patty-ga the dean-ga warned thought

O-S-[S-I-V]-V (9 words)

- (1) A story-o Phillip-ga his wife-ga their son-ni narrated said
- (2) A check-o Holly-ga her friend-ga the salesman-ni wrote realized
- (3) The law-o Allan-ga the congress-ga the President-ni proposed mentioned
- (4) This document-o Judie-ga her colleague-ga their boss-ni faxed thought
- (5) A scarf-o Ethan-ga his brother-ga their mother-ni gave said
- (6) The flowers-o Nora-ga her neighbor-ga her sister-ni sent realized
- (7) A deal-o Brad-ga his colleague-ga a customer-ni offered mentioned
- (8) Her secret-o Olivia-ga her brother-ga their parents-ni told thought

Appendix B: Continued

Ungrammatical Items

O !V S (4 words)

- (1) A car-o washed Daniel-ga
- (2) A boat-o rowed Sarah-ga
- (3) The dinner-o cooked Nick-ga
- (4) The baby-o fed Emma-ga
- (5) A ladder-o climbed Chris-ga
- (6) This computer-o used Alison-ga
- (7) The radio-o fixed Ben-ga
- (8) An apple-o peeled Rachel-ga

O S V !I (7 words)

- (1) A latte-o Bruce-ga brewed his wife-ni
- (2) An email-o Vivian-ga forwarded her colleague-ni
- (3) The profit-o Richard-ga shared his employees-ni
- (4) A legacy-o Ellie-ga left her children-ni
- (5) A dollar-o Elvis-ga paid the cashier-ni
- (6) A lullaby-o Daisy-ga sang her child-ni
- (7) Some money-o Tony-ga bet his friend-ni
- (8) A dress-o Lyla-ga made her daughter-ni

O S V !S V (7 words)

- (1) The promise-o Bill-ga broke his wife-ga said
- (2) A vase-o Claire-ga broke her son-ga realized
- (3) Those men-o Boris-ga stopped the police-ga mentioned
- (4) The tree-o Lydia-ga decorated her father-ga thought
- (5) A monster-o Scott-ga released the captain-ga said
- (6) That computer-o Laura-ga fixed her boyfriend-ga realized
- (7) The country-o Will-ga liberated the military-ga mentioned
- (8) Five languages-o Patty-ga spoke her professor-ga thought

O S S V !I V (9 words)

- (1) An advice-o Phillip-ga his wife-ga gave a boy-ni said
- (2) Those wounds-o Holly-ga her husband-ga showed a doctor-ni realized
- (3) The grades-o Allan-ga her teacher-ga reported the principal-ni mentioned
- (4) A car-o Judie-ga her parents-ga loaned their son-ni thought
- (5) His debt-o Ethan-ga his brother-ga paid the bank-ni said
- (6) Her symptoms-o Nora-ga her sister-ga explained a nurse-ni realized
- (7) An iPhone-o Brad-ga his parents-ga presented his brother-ni mentioned
- (8) An access-o Olivia-ga the government-ga granted her colleague-ni thought

Case Missing (6 words: !-ga, !-o, !-ni)

- (1) The information-o Daniel his colleague-ni provided (!-ga)
- (2) The rule Sarah-ga her friend-ni described (!-o)

Appendix B: Continued

- (3) A bike-o Nick-ga his son bought (!-ni)
- (4) A success-o Emma her boyfriend-ni wished (!-ga)
- (5) This notebook Chris-ga his sister-ni returned (!-o)
- (6) Her transcript-o Alison-ga her parents showed (!-ni)
- (7) A plan-o Ben his boss-ni offered (!-ga)
- (8) The incident Rachel-ga her neighbor-ni described (!-o)

Case Switching (6 words: ga-o, ga-ni, o-ni)

- (1) An invitation-ga Bruce-o his colleagues-ni sent (ga-o)
- (2) The injection-o Vivian-ni her dog-ga gave (ga-ni)
- (3) A package-ni Richard-ga his parents-o delivered (o-ni)
- (4) A reform-ga Ellie-o the council-ni demanded (ga-o)
- (5) His time-o Elvis-ni his child-ga devoted (ga-ni)
- (6) Two tickets-ni Daisy-ga her parents-o reserved (o-ni)
- (7) A cappuccino-ga Tony-o his customer-ni brewed (ga-o)
- (8) A baguette-o Lyla-ni her husband-ga baked (ga-ni)

List 2

Grammatical

O-S-V (4 words)

- (1) The key-o Daniel-ga found
- (2) Some food-o Sarah-ga shopped
- (3) His car-o Nick-ga repaired
- (4) These shoes-o Emma-ga brushed
- (5) A window-o Chris-ga closed
- (6) The carpet-o Alison-ga cleaned
- (7) A chair-o Ben-ga made
- (8) Her voice-o Rachel-ga recorded

O-S-I-V (6 words)

- (1) A quiz-o Bruce-ga his son-ni gave
- (2) The information-o Vivian-ga her lawyer-ni requested
- (3) An collaboration-o Richard-ga his colleague-ni offered
- (4) A plan-o Ellie-ga her boss-ni proposed
- (5) The reason-o Elvis-ga his friend-ni explained
- (6) The schedule-o Daisy-ga her co-workers-ni emailed
- (7) A motorbike-o Tony-ga his brother-ni lent
- (8) A sweater-o Lyla-ga her daughter-ni knitted

OSSVV (7 words)

- (1) His cloth-o Bill-ga his wife-ga reused said
- (2) The TV-o Claire-ga his son-ga broke realized
- (3) This topic-o Boris-ga the committee-ga discussed mentioned
- (4) Her picture-o Lydia-ga that stranger-ga took thought

Appendix B: Continued

- (5) Some money-o Scott-ga his boss-ga paid said
- (6) Her song-o Laura-ga that singer-ga copied realized
- (7) A monkey-o James-ga a scientist-ga studied mentioned
- (8) That cheese-o Patty-ga a mouse-ga stole thought

O-S-[S-I-V]-V (9 words)

- (1) A bicycle-o Phillip-ga his wife-ga their son-ni presented said
- (2) The document-o Holly-ga her husband-ga the police-ni showed realized
- (3) Ten dollars-o Allan-ga his colleague-ga a stranger-ni gave mentioned
- (4) A house-o Judie-ga her boyfriend-ga her brother-ni found thought
- (5) Extra fees-o Ethan-ga the police-ga his wife-ni charged said
- (6) The file-o Nora-ga her colleague-ga their boss-ni sent realized
- (7) A song-o Brad-ga his child-ga his wife-ni sang mentioned
- (8) Their research-o Olivia-ga her colleague-ga the committee-ni presented thought

Ungrammatical Items

O !V S (4 words)

- (1) A microwave-o used Daniel-ga
- (2) This wallet-o dropped Sarah-ga
- (3) The rumor-o spread Nick-ga
- (4) New neighbors-o welcomed Emma-ga
- (5) His friend-o deceived Chris-ga
- (6) That restaurant-o disliked Alison-ga
- (7) Hot coffee-o ordered Ben-ga
- (8) The future-o glimpsed Rachel-ga

O S V !I (7 words)

- (1) English literature-o Bruce-ga taught his students-ni
- (2) A book-o Vivian-ga read her kids-ni
- (3) The dinner-o Richard-ga prepared his wife-ni
- (4) The seats-o Ellie-ga reserved her parents-ni
- (5) A door-o Elvis-ga held a woman-ni
- (6) An issue-o Daisy-ga described her student-ni
- (7) A fee-o Tony-ga charged his friend-ni
- (8) That wine-o Lyla-ga ordered her boyfriend-ni

O S V !S V (7 words)

- (1) The problem-o Bill-ga avoided his friend-ga said
- (2) Their mother-o Claire-ga hugged her sister-ga realized
- (3) His proposal-o Boris-ga accepted his girlfriend-ga mentioned
- (4) A guy-o Lydia-ga slapped her friend-ga thought
- (5) That man-o Scott-ga employed his boss-ga said
- (6) A news-o Laura-ga preached the church-ga realized

Appendix B: Continued

- (7) That kid-o Will-ga recognized his friend-ga mentioned
- (8) The citizen-o Patty-ga united the president-ga thought

O S S V !I V (9 words)

- (1) Those clothes-o Phillip-ga his wife-ga chose their daughter-ni said
- (2) A secret-o Holly-ga her husband-ga whispered their son-ni realized
- (3) The revenge-o Allan-ga his colleague-ga urged their boss-ni mentioned
- (4) A longevity-o Judie-ga her sister-ga wished their parents-ni thought
- (5) That vehicle-o Ethan-ga his wife-ga sold a neighbor-ni said
- (6) A mercy-o Nora-ga her brother-ga begged their enemy-ni realized
- (7) A pie-o Brad-ga his mother-ga baked his father-ni mentioned
- (8) The saving-o Olivia-ga her husband-ga deposited a bank-ni thought

Case Missing (6 words: !-ga, !-o, !-ni)

- (1) An appointment-o Bill his student-ni promised (!-ga)
- (2) A lunch Claire-ga her colleagues-ni brought (!-o)
- (3) Confidential information-o Boris-ga his friend revealed (!-ni)
- (4) The t-shirt-o Lydia her husband-ni bought (!-ga)
- (5) The truth Scott-ga his psychiatrist-ni disclosed (!-o)
- (6) A friend-o Laura-ga his parents introduced (!-ni)
- (7) The food-o Will his dog-ni fed (!-ga)
- (8) A song Patty-ga her boyfriend-ni wrote (!-o)

Case Switching (6 words: ga-o, ga-ni, o-ni)

- (1) A word-ga Phillip-o his teacher-ni spelled (ga-o)
- (2) Her collection-o Holly-ni her friend-ga showed (ga-ni)
- (3) A ring-ni Allan-ga his wife-o sent (o-ni)
- (4) A news-ga Judie-o the public-ni reported (ga-o)
- (5) A marriage-o Ethan-ni his girlfriend-ga promised (ga-ni)
- (6) Strong patience-ni Nora-ga her parents-o demanded (o-ni)
- (7) The data-ga Brad-o his colleague-ni shared (ga-o)
- (8) Her number-o Olivia-ni a guy-ga gave (ga-ni)

Appendix C: WMT Stimuli

List 1

Grammatical Items (Target: underlined)

O-S-V (Before target words: 7 words)

- (1) To earn money, the drug-o Daniel-ga tested several months ago.
- (2) During the day, her car-o Sarah-ga washed very carefully.
- (3) In the morning, a homework-o Nick-ga finished in a special hurry.
- (4) A month ago, a celebrity-o Emma-ga visited with an excitement.
- (5) For some money, a book-o Chris-ga wrote during the summer break.
- (6) In midnight yesterday, her boyfriend-o Alison-ga called with a mounting anger.
- (7) During his study, biology lectures-o Ben-ga hated more than anything else.
- (8) An hour ago, her child-o Rachel-ga scolded for a discipline.

O-S-I-V (Before target word: 9 words)

- (1) For a surprise, a dog-o Bruce-ga his daughter-ni bought yesterday night.
- (2) For her birthday, a package-o Vivian-ga her mother-ni sent a couple months ago.
- (3) With a gratitude, a job-o Richard-ga his friend-ni offered in the last month.
- (4) With some hesitance, her bicycle-o Ellie-ga her brother-ni lent in yesterday evening.
- (5) During the meeting, his opinions-o Elvis-ga the professor-ni told with true honesty.
- (6) With her strategies, an insurance-o Daisy-ga her customer-ni sold a few minutes ago.
- (7) At eleven o'clock, the document-o Tony-ga his boss-ni faxed as soon as possible.
- (8) For their anniversary, a restaurant-o Lyla-ga her parents-ni reserved at yesterday's celebration party.

Case Missing

- (1) After the ride, fifty dollars-o Dwight-ga the driver-ni handed in front of his house.
- (2) An hour ago, a seat-o Linda-ga her husband-ni saved at a movie theater.
- (3) Before the exam, his computer-o Steve-ga his brother-ni lent for a few days.
- (4) At the work, nice lunch-o Pamela-ga her colleague-ni brought during a lunch break.
- (5) At the party, his wife-o Jeff-ga his parents-ni introduced for the first time.
- (6) With some reluctance, her condominium-o Nicole-ga a businessman-ni sold a week before the move-out.
- (7) At his house, some pictures-o Eric-ga his girlfriend-ni showed before dinner today.
- (8) At 12 o'clock, a lunch-o Anna-ga her husband-ni cooked as usual.

Case Mixing

- (1) At the stadium, a ball-o Dwight-ga a player-ni tossed during the yesterday's game.
- (2) In the meeting, her point-o Linda-ga her boss-ni explained very succinctly.
- (3) From his dormitory, his transcript-o Steve-ga his parents-ni faxed at the end of the semester.

Appendix C: Continued

- (4) As a present, a necklace-o Pamela-ga her daughter-ni gave yesterday night.
- (5) To make money, his car-o Jeff-ga his friend-ni sold in the afternoon yesterday.
- (6) In the morning, a letter-o Nicole-ga her parents-ni wrote as soon as arriving on campus.
- (7) After the exam, his score-o Eric-ga his friends-ni boasted repeatedly.
- (8) In yesterday night, some sushi-o Anna-ga her husband-ni ordered for his dinner.

O-S-[S-V]-V (Before target word, 9 words)

- (1) In the court, his mind-o Bill-ga his company-ga broke said very allegedly.
- (2) The last month, the grasses-o Claire-ga her husband-ga cut realized very quickly.
- (3) For a lunch, a sandwich-o Boris-ga his brother-ga ate mentioned in the afternoon.
- (4) The last year, a restaurant-o Lydia-ga her friend-ga ran thought as a business.
- (5) After the accident, their son-o Scott-ga his wife-ga missed said for multiple times.
- (6) For no reason, the money-o Laura-ga her colleague-ga wasted realized in the meeting.
- (7) After the incident, all citizen-o Will-ga the president-ga disturbed mentioned in an interview.
- (8) In the discussion, a fact-o Patty-ga his husband-ga denied thought to disagree with her.

O-S-[S-I-V]-V (Before target word: 10 words)

- (1) Two days ago, a dress-o Phillip-ga his wife-ga their daughter-ni ordered said for a surprise.
- (2) Without her knowing, a ticket-o Holly-ga her husband-ga his friend-ni sold realized last night.
- (3) Yesterday at work, the salary-o Allan-ga his company-ga the employees-ni paid mentioned in contrast to his expectation.
- (4) At her work, a report-o Judie-ga her colleague-ga their boss-ni submitted thought without consulting her.
- (5) More than anything, a happiness-o Ethan-ga his mother-ga his brother-ni wished said at the party.
- (6) With a piffany, an idea-o Nora-ga her friend-ga their professor-ni suggested realized after the meeting.
- (7) After the success, her article-o Brad-ga his girlfriend-ga a journal-ni submitted mentioned during a conversation.
- (8) With her appeals, a raise-o Olivia-ga her boss-ga all employees-ni promised thought at her work.

Ungrammatical Items (ungrammatical element: bold & target: underlined)

O !V S (Before target word: 7 words)

- (1) In yesterday morning, Mt. Everest-o **climbed Daniel-ga** for his lifelong dream.
- (2) An hour ago, her husband-o **called Sarah-ga** with some tears.
- (3) On his computer, several sentences-o **typed Nick-ga** during a conversation yesterday.

Appendix C: Continued

- (4) An hour ago, her clothes-o **packed Emma-ga** for her trip to Hawaii.
- (5) In this morning, his flight-o **caught Chris-ga** at the very last seconds.
- (6) For her research, an insect-o **observed Alison-ga** very closely in the lab.
- (7) For his audience, three languages-o **spoke the governor-ga** in his speech today.
- (8) As her masterpiece, that mural-o **painted Rachel-ga** in the last month.

O S V !I (Before target word: 9 words)

- (1) As a present, a computer-o Bruce-ga bought **his daughter-ni** for her birthday.
- (2) On the internet, a hotel-o Vivian-ga booked **her husband-ni** yesterday night.
- (3) After the class, a favor-o Richard-ga asked **his teacher-ni** on his way home.
- (4) As a researcher, her potential-o Ellie-ga showed **her boss-ni** during the meeting.
- (5) Through an email, his paper-o Elvis-ga submitted his **instructor-ni** very late yesterday.
- (6) At a hospital, comic books-o Daisy-ga brought **her child-ni** five minutes ago.
- (7) During a break, a lunch-o Tony-ga shared **his colleague-ni** at the work yesterday.
- (8) During the class, a clue-o Lyla-ga gave **her students-ni** very briefly today.

O S V !S V (Before target words: 9 words)

- (1) On Sunday night, the schedule-o Bill-ga monitored **his friend-ga** said very convincingly.
- (2) With a good mood, a music-o Claire-ga played **her friend-ga** realized at the party.
- (3) In the news, a missile-o Boris-ga lunched **the air-force-ga** mentioned a few hours ago.
- (4) On the article, the information-o Lydia-ga fabricated **her colleague-ga** thought before submission.
- (5) During a conversation, a stranger-o Scott-ga helped **his son-ga** said with a pride.
- (6) During a conference, the revenue-o Laura-ga tripled **her company-ga** realized in that quarter.
- (7) Yesterday, at last, a contract-o Will-ga reviewed **his team-ga** mentioned during the interview.
- (8) After her graduation, a business-o Patty-ga started **her friend-ga** thought in this economy.

O S S V !I V (Before target word: 9 words)

- (1) In the meeting, the issue-o Phillip-ga his colleague-ga explained **their customer-ni** said a few hours ago.
- (2) For his birthday, a boat-o Holly-ga her husband-ga bought **their son-ni** realized yesterday night.
- (3) In the class, the rules-o Allan-ga his teacher-ga explained **the students-ni** mentioned very carefully.
- (4) At an intersection, a map-o Judie-ga her husband-ga showed **a traveler-ni** thought in this afternoon.

Appendix C: Continued

- (5) After the class, a homework-o Ethan-ga his professor-ga gave **the students-ni** said with a disappointment.
- (6) During a battle, his weakness-o Nora-ga her husband-ga revealed **their enemy-ni** realized very anxiously.
- (7) After the closing, a task-o Brad-ga his boss-ga assigned **his co-workers-ni** mentioned very furiously.
- (8) After the recession, some money-o Olivia-ga her husband-ga transferred **her cousin-ni** thought at the bank.

Case Missing (!-ga, !-o, !-ni)

- (1) At the park, five dollars-o **Dwight** a boy-ni gave for his help. (!-ga)
- (2) An hour ago, **a message** Linda-ga her colleagues-ni relayed through the email. (!-o)
- (3) For his birthday, his car-o Steve-ga **his son** loaned in the afternoon today. (!-ni)
- (4) At a restaurant, this pasta-o **Pamela** her husband-ni ordered with no doubt. (!-ga)
- (5) On her bed, **a story** Jeff-ga his daughter-ni read at 9 o'clock. (!-o)
- (6) Two hours ago, a hotel-o Nicole-ga **her friend** reserved for her visit. (!-ni)
- (7) At the door, a gift-o **Eric** his friend-ni delivered a few minutes ago. (!-ga)
- (8) About the product, **a detail** Anna-ga her customer-ni provided over the phone. (!-o)

(F) Case Mixing (ga-o, ga-ni, o-ni)

- (1) At the store, **a bill-ga Dwight-o** the cashier-ni handed for his diet coke. (ga-o)
- (2) At the school, some sandwiches-o **Linda-ni her colleague-ga** shared during a lunch break. (ga-ni)
- (3) Today, at last, **a house-ni** Steve-ga **his family-o** built on his own. (o-ni)
- (4) In the library, **this book-ga Pamela-o** her friend-ni kept for her study. (ga-o)
- (5) Two days ago, a bonus-o **Jeff-ni his employees-ga** paid for the first time in years. (ga-ni)
- (6) In this winter, **a sweater-ni** Nicole-ga **her husband-o** knitted for his outing. (o-ni)
- (7) At the store, **a skirt-ga Eric-o** his daughter-ni bought with no hesitance. (ga-o)
- (8) At the work, the regulation-o **Anna-ni her friend-ga** explained very carefully. (ga-ni)

List 2

Grammatical (target: underlined)

O-S-V (Before target words: 7 words)

- (1) Before moving out, his moped-o Daniel-ga sold at a garage sale.
- (2) An hour ago, the professor-o Sarah-ga called in the middle of the night.

Appendix C: Continued

- (3) On his phone, the price-o Nick-ga calculated for a new computer.
- (4) To the audience, the issue-o Emma-ga addressed in yesterday's news.
- (5) In the morning, a pizza-o Chris-ga ordered for a breakfast.
- (6) At nine o'clock, her house-o Alison-ga left in a hurry yesterday.
- (7) After the flight, his watch-o Ben-ga adjusted as soon as arriving in Tokyo.
- (8) In the meeting, the problem-o Rachel-ga identified very cleverly.

O-S-I-V (Before target word: 9 words)

- (1) At a dinner, a friend-o Bruce-ga his wife-ni introduced very briefly yesterday.
- (2) Five minutes ago, her debts-o Vivian-ga the bank-ni paid finally after her bankruptcy.
- (3) On her birthday, a flight-o Richard-ga his mother-ni booked very secretly.
- (4) On the street, her clothes-o Ellie-ga a woman-ni sold at a free market.
- (5) At a bar, a beer-o Elvis-ga his customer-ni brewed for free yesterday.
- (6) For his retirement, a cake-o Daisy-ga her father-ni baked in the kitchen yesterday.
- (7) In the meeting, his plan-o Tony-ga his mentor-ni described a few hours ago.
- (8) With a chat, a message-o Lyla-ga her boyfriend-ni sent a few minutes ago.

Case Missing

- (1) For a present, golf clubs-o Dwight-ga his father-ni bought on his way home.
- (2) At a school, English literature-o Linda-ga her children-ni taught on every Tuesday.
- (3) An hour ago, his blood-o Steve-ga Red Cross-ni donated for the first time.
- (4) After the work, a lunch-o Pamela-ga her brother-ni offered for his help.
- (5) During a class, a pen-o Jeff-ga his friend-ni lent out of his kindness.
- (6) For her car, a price-o Nicole-ga her neighbor-ni proposed with some reluctance.
- (7) At 9 o'clock, an adventure-o Eric-ga his son-ni narrated for a bedtime story.
- (8) In the meeting, an answer-o Anna-ga the committee-ni demanded very persistently.

Case Mixing

- (1) In his speech, a wisdom-o Dwight-ga graduating students-ni imparted yesterday.
- (2) From a movie, a line-o Linda-ga her boyfriend-ni quoted on their date.
- (3) During the gamble, his car-o Steve-ga his colleague-ni bet with his folly.
- (4) At a conference, the research-o Pamela-ga her colleagues-ni presented very quickly.
- (5) After the championship, a trophy-o Jeff-ga his student-ni awarded for his first place.
- (6) Ten minutes ago, those seeds-o Nicole-ga her parrot-ni fed for the first time today.
- (7) In a hurry, the reports-o Eric-ga his boss-ni faxed after the meeting.
- (8) For some money, those books-o Anna-ga her friend-ni sold at her house.

O-S-[S-V]-V (Before target words: 9 words)

- (1) In the afternoon, green tee-o Bill-ga his wife-ga enjoyed said at a tee party.

Appendix C: Continued

- (2) In the meeting, their plan-o Claire-ga her colleague-ga finalized realized without any careful consideration.
- (3) At the work, new interns-o Boris-ga his boss-ga hired mentioned in the beginning of the year.
- (4) With careful thinking, many problems-o Lydia-ga today's economy-ga caused thought in a lecture today.
- (5) After the dinner, the restaurant-o Scott-ga his wife-ga rated said when leaving.
- (6) On a computer, a software-o Laura-ga her husband-ga programmed realized without her knowing.
- (7) At 10am today, his power-o Will-ga the President-ga exercised mentioned in a broadcast news.
- (8) About the incident, a story-o Patty-ga her colleague-ga recalled thought at a police station.

O-[S-S-I-V]-V (Before target word: 10 words)

- (1) Two years ago, a trip-o Phillip-ga his wife-ga her parents-ni planned said for their anniversary.
- (2) At a shop, those clothes-o Holly-ga her husband-ga their kids-ni bought realized without any consultation.
- (3) At the party, a song-o Allan-ga his brother-ga his girlfriend-ni made mentioned with a jealousy.
- (4) During the presentation, a rationale-o Judie-ga her colleague-ga the audience-ni explained thought very beautifully.
- (5) Five minutes ago, a pill-o Ethan-ga his wife-ga their son-ni gave said for his awful cold.
- (6) In a class, a handout-o Nora-ga his professor-ga his students-ni brought realized for the first time.
- (7) An hour ago, term paper-o Brad-ga his colleague-ga their instructor-ni submitted mentioned two hours after the due.
- (8) In the midnight, a singal-o Olivia-ga her husband-ga their neighbor-ni communicated thought with Morse codes.

Ungrammatical Items (ungrammatical: bold & target: underlined)

O !V S (Before target words: 7 words)

- (1) Before the due, his assignment-o **finished Daniel-ga** with no difficulty.
- (2) As an architect, that bridge-o **designed Sarah-ga** a few years ago.
- (3) At a concert, that singer-o **hated Nick-ga** after his terrible performance.
- (4) A year ago, the championship-o **won Emma-ga** with her effortless practice.
- (5) With his friends, that window-o **broke Chris-ga** with his baseball bat.
- (6) In the morning, her clothes-o **packed Alison-ga** for her trip to U.K.
- (7) After the practice, a shower-o **took Ben-ga** quickly in his room.
- (8) On every Sunday, a movie-o **watched Rachel-ga** with her daughter.

Appendix C: Continued

O S V !I (Before target word: 9 words)

- (1) In Barcelona, yesterday, Spanish words-o Bruce-ga taught **his son-ni** in a very clever way.
- (2) At a shop, a t-shirt-o Vivian-ga selected **her boyfriend-ni** with her taste of clothes.
- (3) After his success, life lessons-o Richard-ga lectured **his mentee-ni** for two consecutive days.
- (4) With her connections, an apartment-o Ellie-ga found **her brother-ni** for a very cheap price.
- (5) At a garage, his bike-o Elvis-ga gave **his brother-ni** before moving out yesterday.
- (6) During their visit, her room-o Daisy-ga showed **her parents-ni** with a little embarrassment.
- (7) At 12 o'clock, vegetable soup-o Tony-ga cooked **his daughter-ni** for her lunch today.
- (8) At her house, five dollars-o Lyla-ga handed **a boy-ni** for his tireless help.

O S V !S V (9 words)

- (1) In the interview, their child-o Bill-ga protected **his wife-ga** said at the horrible car accident.
- (2) For no reasons, his friends-o Claire-ga hit **her son-ga** realized at the kindergarten.
- (3) With unnecessary questions, his presentation-o **Boris-ga** ruined his colleague-ga mentioned at a conference today.
- (4) On the article, his theory-o Lydia-ga corroborated **her professor-ga** thought with a new methodology.
- (5) Without any consultation, his guitar-o Scott-ga sold **his wife-ga** said at the free market.
- (6) A year ago, a company-o Laura-ga started **her colleague-ga** realized as a lifelong dream.
- (7) On December 20th, the Christmas-o Will-ga celebrated **his family-ga** mentioned a little bit early than usual.
- (8) Three minutes ago, his wallet-o Patty-ga dropped **that man-ga** realized at an intersection.

O S S V !I V (Before target: 10 words)

- (1) At the party, the piano-o Phillip-ga his son-ga played **his wife-ni** said very happily yesterday.
- (2) After the walk, some milk-o Holly-ga her husband-ga **brought their dog-ni** realized a few minutes ago.
- (3) In the kitchen, those cookies-o Allan-ga her daughter-ga baked **her mother-ni** mentioned very proudly.
- (4) In the meeting, several questions-o Judie-ga her colleague-ga asked **their professor-ni** thought without any thoughts.
- (5) At the ceremony, an honor-o Ethan-ga the President-ga bestowed **his father-ni** said for his contribution to the country.

Appendix C: Continued

- (6) After the incident, the issue-o Nora-ga her colleague-ga explained **their client-ni** realized with a true honesty.
- (7) At the school, his son-o Brad-ga his teacher-ga recommended **a fellowship-ni** mentioned with an astonishment.
- (8) At a bar, a drink-o Olivia-ga that man-ga bought **her friend-ni** thought with a jealousy.

Case Missing (!-ga, !-o, !-ni)

- (1) In a class, an advice-o **Dwight** his students-ni gave a day before their graduation. (!-ga)
- (2) At a store, **a computer** Linda-ga her father-ni chose with her experience. (!-o)
- (3) Through a mail, an invitation-o Steve-ga **his friends** sent two months before his wedding. (!-ni)
- (4) In the speech, an assertion-o **Pamela** her voter-ni made very passionately. (!-ga)
- (5) For a day, **his cottage** Jeff-ga his friend-ni lent for free. (!-o)
- (6) In the meeting, **a break** Nicole-ga her boss proposed after an hour of discussion. (!-ni)
- (7) In a reply, an extension-o **Eric** his students-ni granted for one more week. (!-ga)
- (8) In the night, **a story** Anna-ga her children-ni told before going to bed. (!-o)

Case Switching (ga-o, ga-ni, o-ni)

- (1) During a conversation, **his feelings-ga** **Dwight-o** his friend-ni revealed honestly.
- (2) After a purchase, a receipt-o **Linda-ni** **her customer-ga** faxed within an hour.
- (3) After the injury, **his wounds-ni** Steve-ga **the doctor-o** showed very briefly.
- (4) An hour ago, **a detail-ga** **Pamela-o** her customer-ni provided over the phone.
- (5) In the morning, some coffee-o **Jeff-ni** **his wife-ga** brewed with some sugar and milk.
- (6) With some hesitance, **fifty dollars-ni** Nicole-ga **her brother-o** loaned yesterday.
- (7) On a paper, **his name-ga** **Eric-o** a woman-ni wrote at a restaurant.
- (8) At a store, a cappuccino-o **Anna-ni** **her husband-ga** ordered out of her kindness.

Grammatical Fillers (target: underlined)

O-S-V

- (1) During a lecture, that word-o John-ga defined instead of the professor.
- (2) In the meeting, the possibilities-o Mary-ga listed as many as possible.
- (3) After the ceremony, his son-o Mike-ga congratulated for his graduation.
- (4) At the discussion, her friend-o Stacey-ga supported no matter what.
- (5) In the morning, the weather-o James-ga checked for his trip to Canada.
- (6) On the article, her critics-o Karen-ga attacked very severely.
- (7) At a store, a computer-o Tim-ga purchased for his research.
- (8) A month ago, her PhD-o Angela-ga completed at the age of 22.

Appendix C: Continued

O-S-I-V

- (1) In the meeting, the expense-o John-ga his boss-ni reported as the end-of-year report.
- (2) For her cold, a soup-o Mary-ga her daughter-ni cooked at 12pm today
- (3) In the afternoon, the propaganda-o Mike-ga the citizen-ni promoted in front of a station.
- (4) A year ago, an apartment-o Stacey-ga her parents-ni bought to live together.
- (5) After the incident, the news-o James-ga the station-ni transmitted as soon as possible.
- (6) For many years, her life-o Karen-ga her parents-ni dedicated very vigorously.
- (7) During a meeting, a joke-o Tim-ga his friend-ni told for many times.
- (8) At 9 o'clock, a story-o Angela-ga her son-ni read before going to bed.

O-S-[S-V]-V

- (1) During the lecture, the military-o John-ga his ancestor-ga led said with a pride.
- (2) Without her knowing, her picture-o Mary-ga her husband-ga took realized at her birthday party.
- (3) During the conversation, their son-o Mike-ga his wife-ga convinced mentioned very perseveringly.
- (4) In the laboratory, chemical compounds-o Stacey-ga her colleague-ga analyzed thought right before her presentation.
- (5) Several days ago, the club-ni James-ga his friend-ga joined said after his persistent persuasion.
- (6) During the speech, her language-o Karen-ga his husband-ga corrected realized very covertly.
- (7) At the conference, the theory-o Tim-ga his professor-ga advocated mentioned in his presentation.
- (8) In the meeting, a decision-o Angela-ga her boss-ga made thought after an hour of discussion.
- (9) After the fire, the damage-o Tom-ga his lawyer-ga estimated said over the phone.

O-[S-S-I-V]-V

- (1) In the report, most expenditure-o John-ga the mayor-ga the infrastructures-ni invested said instead of education.
- (2) On a street, several sentences-o Mary-ga her husband-ga the traveler-ni interpreted realized in the afternoon today.
- (3) For their research, a license-o Mike-ga the commissioner-ga his colleague-ni authorized mentioned after an hour of consultation with some professionals.
- (4) During the meeting, his viewpoint-o Stacey-ga her boss-ga all attendants-ni articulated thought very neatly.
- (5) For a surprise, a gift-o James-ga his wife-ga his parents-ni sent said with a gratitude.
- (6) Given her potential, an approval-o Karen-ga her professor-ga her colleague-ni gave realized during a conversation.

Appendix C: Continued

- (7) Given his success, the knighthood-o Tim-ga the Queen-ga his father-ni awarded mentioned on the national TV.
- (8) At the store, soy milk-o Angela-ga her husband-ga their children-ni bought thought for their lactose intolerance.
- (9) Without consulting him, their cottage-o Tom-ga his wife-ga the neighbor-ni sold said with a rage.

Appendix D: Biographical Questionnaire

Biographical Questionnaire

1. Participant Code: _____
2. Gender: Female _____
 Male _____
3. Age: Age _____
4. Email: _____@_____
5. Major: _____
6. Year: _____

7. Do you have any problems with hearing or sight? If yes, please:

8. What is your native language?

9. What other languages do you know? (Please include proficiency, choosing from the options below).

[beginner, low intermediate, upper intermediate, advanced]
10. Can you also tell how much experience you have in those languages?

[Example: one semester course, two semester courses, and a year of immersion, etc.]
11. What other countries other than the United States have you ever been to? How long did you stay there? (Example answer: Mexico, 2 months).

THANK YOU!

Appendix E: Post-Experimental Questionnaire

Participant Code

Post-experiment Survey

Thanks so much for your participation! Before ending, I would like you to take about 10 mins to respond to a survey. Your responses to the survey are very important to the research questions in this study. So please answer to survey questions with your complete honesty. Your responses will be kept completely anonymous and used only for the purpose of this study.

1. Could you tell me everything that you noticed or understood about the language used in the study? Did you notice any particular rule or regularity?

2. If you did notice anything, can you tell "when" you came to notice it (or them)?—Was that when you were listening to the examples of the language? Or when you were taking the tests? Or even now?

Appendix E: Continued

3. This study focused on some word orders. Did you notice any word order patterns or rules in the language? If you did, can you describe what they are (anything!)?

4. This study focused on case-markers in the language. Did you notice anything about the word endings in the language? If you did, can you describe what they are (anything please!)?

5. Also, can you tell when you noticed them?

Appendix E: Continued

6. In languages of the world, there are fixed patterns of word orders such as SVO (John reads a book), SVIO (John buys me a book), and SVSVO (John said that Mary bought the book): S=subject, V=verb, I=indirect object, and O=direct object. Are you aware of any patterns in the language in this framework as far as you know (was it VSO, SOV, SIOV, or any others)?

7. In the language you saw, there are simple word order patterns. Are you aware of any patterns? Also, which ones were they as far as you know? (Were they SVO, SOV, OSV, OVS, VSO, SIOV, SIVO, SOIV or any others).

8. In the language you heard, there are complex word order patterns. Are you aware any of them? Also, which do you think they were? (SSVVO, SSIOVV, SSVOV, SSIVOV, SVISOV or any others)?

Appendix E: Continued

9. In the language, there were three word endings, "-ga", "--o", and "--ni". Are you aware of any patterns or rules regarding with them?

THANK YOU!

Bibliography

- Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, 30(4), 481-509
- Baddeley, A. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.
- Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology*, 63, 1-29.
- Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105(1), 158-173.
- Baddeley, A. D. & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation*, Vol. 8 (pp. 47-89). New York, NY: Academic Press.
- Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, 27(5), 586-595.
- Bell, P. (2009). Le cadeau or la cadeau? The role of aptitude in learner awareness of gender distinctions in French. *The Canadian Modern Language Review / La revue canadienne des langues vivantes*, 65(4), 615-643.
- Bell, P. K. (2017). Explicit and implicit learning: Exploring their simultaneity and immediate effectiveness. *Applied Linguistics*, 38(3), 297-317.
- Bergsleithner, J. M. (2011). The role of noticing and working memory capacity in L2 oral performance. *Organon*, 26(51), 217-243.

- Berry, D. C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Hove: Lawrence Erlbaum
- Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41-68). New York: Cambridge University Press.
- Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, 33(2), 247-271.
- Brooks, P. J., Kwoka, N., & Kempe, V. (2017). Distributional effects and individual differences in L2 morphology learning. *Language Learning*, 67(1), 171-207.
- Carroll, J. B. (1973). Implications of aptitude test research and psycholinguistic theory for foreign language teaching. *International Journal of Psycholinguistics*, 2, 5-14.
- Carroll, J. B. (1981). Twenty-five years of research in foreign language aptitude. In K. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83-118). Rowley, MA: Newbury House.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Carroll, J. B. & Sapon, S. M. (1959). *Modern language aptitude test*. New York: Psychological Corporation.
- Çeçen S., & Erçetin, G. (2016). Interrelationships among L2 linguistic knowledge, working memory functions, and L2 reading. In G. Granena, D. O. Jackson, & Y. Yılmaz (Eds.), *Cognitive individual differences in second language processing and acquisition*. (pp. 223-248). Amsterdam: John Benjamins.

- Chan, C. (1992). *Implicit cognitive processes: Theoretical issues and applications in computer systems design*. Unpublished Ph.D. dissertation, University of Oxford, Oxford, United Kingdom.
- Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, 2(10), 406-416.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). New York, NY: Cambridge University Press.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450-466.
- DeKeyser, R.M. (1994). Implicit and explicit learning of L2 grammar: A pilot study. *TESOL Quarterly*, 28, 188-94.
- DeKeyser, R. M. (1995). Learning second language grammar rules. *Studies in Second Language Acquisition*, 17(3), 379-410.
- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22(4), 499-533.
- DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (eds.), *The handbook of second language acquisition*. (pp. 313-348). Malden, MA: Wiley- Blackwell.

- DeKeyser, R. M. (2005). What makes second-language grammar difficult? A review of issues. *Language Learning*, 55(s1), 1-25.
- DeKeyser, R. M. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, 62(s2), 189-200.
- DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 94–112). London: Routledge.
- DeKeyser, R. M. (2016). Of moving targets and chameleons: Why the concept of difficulty is so hard to pin down, *Studies in Second Language Acquisition*, 38(2), 353-363.
- DeKeyser, R. M., & Koeth, J. (2011). Cognitive aptitudes for second language learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 395-406). London: Routledge.
- Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review*, 8(2), 343-350.
- Dienes, Z. (2008). Subjective measures of unconscious knowledge. *Progress in Brain Research*, 168, 49-64.
- Dienes, Z. (2012). Conscious versus unconscious learning of structure. In P. Rebuschat & J.N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 337-364). Berlin: De Gruyter.

- Dienes, Z., Altmann, G., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(5), 1322-1338.
- Dienes, Z. & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research*, 69(5-6), 338-351.
- Dörnyei Z. & Skehan, P. (2003). Individual differences in second language learning. In Doughty and M. H. Long (Eds.), *The handbook of second language acquisition*. Malden, MA: Wiley-Blackwell. (pp. 589-630).
- Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty & M. H. Long (Eds.), *Handbook of Second Language Acquisition*. (pp. 256-310). Oxford, UK: Wiley-Blackwell.
- Doughty, C., Campbell, S., Mislevy, M., Bunting, M., Bowles, A., & Koeth, J. (2010). Predicting near-native ability: The factor structure and reliability of Hi-LAB. In M. Prior, Y. Watanabe, & S. Lee (Eds.), *Selected proceedings of the 2008 Second Language Research Forum* (pp. 10-31). Somerville, MA: Cascadilla Proceedings Project.
- Ellis, N. C. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology*, 5, 289-318.
- Ellis, N. C. (1994). *Implicit and explicit learning of languages*. San Diego, CA: Academic Press.

- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18(01), 91-126.
- Ellis, N.C. (1999). Cognitive approaches to SLA. *Annual Review of Applied Linguistics*, 19, 22-42.
- Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27(2), 305-352.
- Ellis, N. C. (2015). Implicit and explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Investigating implicit and explicit language learning* (pp. 3-24). Amsterdam: John Benjamins.
- Ellis, N. C., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics—Introduction to the Special Issue. *Applied Linguistics*, 27(4), 558-589.
- Ellis, N. C., Römer, U. & O'Donnell, M. B. (2016). *Usage-based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar*. Language Learning Monograph Series. Hoboken, NJ: Wiley-Blackwell.
- Ellis, N. C., & Schmidt, R. (1997). Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition*, 19, 145-171.
- Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *Quarterly Journal of Experimental Psychology*, 49, 234-250.

- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in second language acquisition*, 27(2), 141-172.
- Ellis, R. (2015). Form-focused instruction and the measurement of implicit and explicit L2 knowledge. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages*. (pp 417-442). Amsterdam: John Benjamins.
- Ellis, R., & Loewen, S. (2007). Confirming the operational definitions of explicit and implicit knowledge in Ellis (2005): Responding to Isemonger. *Studies in Second Language Acquisition*, 29(01), 119-126.
- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, 28(2), 339-369.
- Engle, R., Kane, M. J. & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active maintenance and executive control* (pp.102-131). New York: Cambridge University Press.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35(1), 116-124.
- French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, 29(3), 463-487.

- Godfroid, A. (2016). The effects of implicit instruction on implicit and explicit knowledge development. *Studies in Second Language Acquisition*, 38(2), 177-215.
- Goo, J., Granena, G., Novella, M., & Yilmaz, Y. (2015). Implicit and explicit instruction in L2 learning: Norris and Ortega (2000) revisited and updated. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443-482). Amsterdam: John Benjamins.
- Graham, C. R., & Williams, J. N. (2016). Implicit learning of Latin stress regularities. *Studies in Second Language Acquisition*, 38(4), 1-27.
- Granena, G. (2012). *Age differences and cognitive aptitudes for implicit and explicit learning in ultimate second language attainment*. Unpublished Ph.D. Dissertation. University of Maryland, College Park, MD.
- Granena, G. (2013a). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, 63(4), 665-703.
- Granena, G. (2013b). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment*. (pp. 105-129).
- Granena, G. (2013c). Reexamining the robustness of language aptitude in SLA. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment*. (pp. 179-204). Amsterdam: John Benjamins.

- Granena, G. (2016a). Cognitive aptitudes for implicit and explicit learning and information-processing styles: An individual differences study. *Applied Psycholinguistics*, 37(3), 577-600.
- Granena, G. (2016b). Part 1: Age differences, maturational constraints, and implicit and explicit L2 learning. *Brill Research Perspectives in Multilingualism and Second Language Acquisition*, 1(1), 9-44.
- Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, 29(3), 311-343.
- Grey, S., Cox, J. G., Serafini, E. J., & Sanz, C. (2015). The role of individual differences in the study abroad context: Cognitive capacity and language development during short-term intensive language exposure. *Modern Language Journal*, 99(1), 137-157.
- Grey, S., Williams, J. N., & Rebuschat, P. (2014). Incidental exposure and L3 learning of morphosyntax. *Studies in Second Language Acquisition*, 36(4), 611-645.
- Grey, S., Williams, J. N., & Rebuschat, P. (2015). Individual differences in incidental language learning: Phonological working memory, learning styles, and personality. *Learning and Individual Differences*, 38(, 44-53.
- Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition*, 35(3), 423-449.

- Hama, M., & Leow, R. P. (2010). Learning without awareness revisited. *Studies in Second Language Acquisition*, 32(3), 465-491.
- Hamrick, P. (2015). Declarative and procedural memory abilities as individual differences in incidental language learning. *Learning and Individual Differences*, 44, 9-15.
- Hamrick, P., & Rebuschat, P. (2013). How implicit is statistical learning? In P. Rebuschat & J.N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 365-382). Berlin: de Gruyter.
- Hamrick, P., & Sachs, R. (2017). Establishing evidence of learning in experiments employing artificial linguistic systems. *Studies in Second Language Acquisition*, 1-17.
- Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition*, 19(3), 379-400.
- Hassin, R. R., Bargh, J. A., Engell, A. D., & McCulloch, K. C. (2009). Implicit working memory. *Consciousness and Cognition*, 18(3), 665-678.
- Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty an complexity in L2 acquisition. *Studies in Second Language Acquisition*, 38(2), 163-175.
- Houser, R. (2008). *Counseling and educational research: Evaluation and application*. Thousand Oaks, CA: Sage.

- Hulstijn, J. H. (1989). Implicit and incidental second language learning: Experiments in the processing of natural and partly artificial input. In H.W. Dechert & M. Raupach (Eds.), *Interlingual processing* (pp. 49-73). Tübingen: Gunter Narr.
- Hulstijn, J. H. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction*. (pp. 258-286). Cambridge: Cambridge University Press.
- Hulstijn, J. H. (2003). Incidental and Intentional Learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349-381). Oxford: Wiley- Blackwell.
- Hulstijn, J. H. (2013). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. (pp. 2632-2640). Oxford: Blackwell- Publishing.
- Jackson, D. O. (2013). Implicit second language learning and individual differences. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka, (Eds.). *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 277-293). Honolulu: University of Hawai'i, National Foreign Language Resource Center.
- Jackson, D. O. (2014). *The relative contribution of input modification, learner awareness, and individual differences to second language construction learning*. Unpublished Ph.D. Dissertation. University of Hawai'i at Mānoa, Honolulu, HI.

- Jackson, D. O. (2016a). An empirical study of working memory, personality, and second language construction learning. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition*. (pp. 157-184).
- Jackson, D. O. (2016b). Working memory and second language acquisition: Theory and findings. *The Journal of Kanda University of International Studies*, 28, 21-47.
- Janacsek, K., & Nemeth, D. (2013). Implicit sequence learning and working memory: Correlated or complicated?. *Cortex*, 49(8), 2001-2006.
- Janacsek, K., & Nemeth, D. (2015). The puzzle is complicated: When should working memory be related to implicit sequence learning, and when should it not? (Response to Martini et al.). *Cortex*, 64, 411-412.
- Jiang, N. (2012). *Conducting reaction time research in second language studies*. New York: Routledge.
- Jiménez, L., Vaquero, J. M., & Lupiáñez, J. (2006). Qualitative differences between implicit and explicit sequence learning. *Journal of experimental psychology: Learning, Memory, and Cognition*, 32(3), 475.
- Juffs, A., & Harrington, M. (2011). State-of-the-art article: Aspects of working memory in L2 learning. *Language Teaching*, 44, 137-166.
- Kachinske, I., Osthus, P., Solovyeva, K., & Long, M. (2015). Implicit learning of a L2 morphosyntactic rule, and its relevance for language teaching. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages*. (pp. 385-416). Amsterdam: John Benjamins.

- Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, 116(3), 321-340.
- Kerz, E., Wiechmann, D., & Riedel, F. (2017). Implicit learning in the crowd: Investigating the role of awareness in the acquisition of L2 knowledge. *Studies in Second Language Acquisition*, 39(4), 711-734.
doi:10.1017/S027226311700002X.
- Kilborn, K., & Moss, H. (1996). Word monitoring. *Language and Cognitive Processes*, 11(6), 689-694.
- Kormos, J. (2013). New conceptualizations of language aptitude in second language development. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment*. (pp. 131-152).
- Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, 11(2), 261-271.
- Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.
- Krashen, S. D. (1994). The Input Hypothesis and its rivals. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45-77). London: Academic Press.

- Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition*, 10(3), 294-340.
- Lado, B. (2008). *The role of bilingualism, type of feedback, and cognitive capacity in the acquisition of non-primary languages: A computer-based study*. Unpublished Ph.D. dissertation. Georgetown University, Washington, D.C.
- Lado, B. (2017). Aptitude and pedagogical conditions in the early development of a nonprimary language. *Applied Psycholinguistics*, 38(3), 679-701.
- Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning*, 47(3), 467-505.
- Leow, R. P. (2000). A study of the role of awareness in foreign language behavior. *Studies in Second Language Acquisition*, 22(04), 557-584.
- Leow, R. P. (2015). Implicit learning in SLA: Of processes and products. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages*. (pp. 47-66). Amsterdam: John Benjamins.
- Leow, R. P., & Hama, M. (2013). Implicit learning in SLA and the issue of internal validity: A response to Leung and Williams's (2011) "The implicit learning of mappings between forms and contextually derived meanings". *Studies in Second Language Acquisition*, 35(3), 545-557.
- Leung, J. H., & Williams, J. N. (2011). The implicit learning of mappings between forms and contextually derived meanings. *Studies in Second Language Acquisition*, 33(1), 33-55.

- Leung, J. H., & Williams, J. N. (2012). Constraints on implicit learning of grammatical form-meaning connections. *Language Learning*, 62(2), 634-662.
- Leung, J. H., & Williams, J. N. (2014). Cross-linguistic differences in implicit language learning. *Studies in Second Language Acquisition*, 36(4), 733-755.
- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60(2), 309-365.
- Li, S. (2016). The construct validity of language aptitude: A Meta-Analysis. *Studies in Second Language Acquisition*, 38(4), 801-842.
- Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., ... & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, 63(3), 530-566.
- Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, 21(4), 861-883.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95(4), 492-527.
- Logan, G. D. (1992). Shapes of reaction time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18(5), 883-914.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.). *Handbook of second language acquisition* (pp.413-468). San Diego, CA: Academic Press.

- Long, M. H. (2013). Maturational constraints on child and adult SLA. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment*. (pp. 3-42). Amsterdam: John Benjamins.
- Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Malden, MA: Wiley-Blackwell.
- Long, M. H. (2017). Instructed second language acquisition (ISLA): Geopolitics, methodological issues, and some major research questions. *Instructed Second Language Acquisition*, 1(1), 7-44.
- Lowie, W., & Verspoor, M. (2015). Variability and variation in second language acquisition orders: A dynamic reevaluation. *Language Learning*, 65(1), 63-88.
- Mackey, A. (1999). Input, interaction, and second language development. *Studies in Second Language Acquisition*, 21(4), 557-587.
- Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning*, 60(3), 501-533.
- Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 408-452). Oxford, UK: Oxford University Press.
- Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed second language learning* (pp. 181-209). Amsterdam: John Benjamins.

- Marsden, E., Mackey A., & Plonsky, L. (2016). The IRIS Repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS Repository of Instruments for Research into Second Languages* (pp. 1-21). New York: Routledge.
- Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, 34(3), 379-413.
- Martini, M., Sachse, P., Furtner, M. R., & Gaschler, R. (2015). Why should working memory be related to incidentally learned sequence structures. *Cortex*, 64, 407-410.
- McLaughlin, B. (1980). On the use of miniature artificial languages in second-language research. *Applied Psycholinguistics*, 1(4), 357-369.
- Meara, P. (2005). *LLAMA language aptitude tests: The manual*. Lognosites: University of Swansea.
- Miyake, A. & Friedman, D. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. Hearly & L. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention*. (pp. 339-364). Mahwah, NJ: Erlbaum Associates.
- Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.
- Morgan-Short, K., Finger, I., Grey, S., & Ullman, M. T. (2012). Second language processing shows increased native-like neural responses after months of no exposure. *Plos One*, 7(3), e32974.

- Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K.A., Carpenter, H., & Wong, P. C. M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition*, 17(1), 56-72.
- Murakami, A. (2016). Modeling systematicity and individuality in nonlinear second language development: The case of English grammatical morphemes. *Language Learning*, 66(4), 834-871.
- Murakami, A., & Alexopoulou, T. (2016). Longitudinal L2 development of the English article in individual learners. In A. Papafragou, D. Grodner, J. Trueswell, & D. Mirman (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 1050-1055). Philadelphia, PA: Cognitive Science Society.
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cognitive Psychology*, 19(1), 1-32.
- Norouzian, R., & Plosnky, L. (2017). Eta- and partial eta-squared in L2 research: A cautionary review and guide to more appropriate usage. *Second Language Research*, 34(2), 1-15.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417-528.
- Norris, J. M., & Ortega, L. (2003). Defining and measuring SLA. In C. J. Doughty, & M. H. Long (Eds.). *Handbook of second language acquisition* (pp. 717-761). Malden, MA: Blackwell.

- Onnis, L., Waterfall, H. R., & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition*, 109(3), 423–430.
- Ortega, L. & Byrne, H. (2008). *The longitudinal study of advanced L2 capacities*. London, UK: Routledge.
- Ortega, L. & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, 25, 26-45.
- Paciorek, A., & Williams, J.N. (2015). Implicit learning of semantic preferences of verbs. *Studies in Second Language Acquisition*, 37(2), 359-382.
- Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins
- Perruchet, P. (2008). Implicit learning. In J. Byrne & H. L. Roediger, III (Eds.), *Learning and memory : A comprehensive reference*. (Vol. 2: Cognitive psychology of memory, pp.597- 621), Oxford : Elsevier.
- Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or fragmentary knowledge? *Journal of Experimental Psychology: General*, 119(3), 264-275.
- Pimsleur, P. (1966). *Pimsleur language aptitude battery*. New York, NY: Harcourt Brace Jovanovich
- Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, 133(2), 227-244.

- R Core Team (2017). R: A language and environment for statistical computing
(Version 3.3.3). R Foundation for Statistical Computing, Vienna, Austria.
URL <http://www.R-project.org/>.
- Ranta, L. (2002). The role of learners' language-analytic ability in the communicative classroom. In P. Robinson (Ed.), *Individual differences and instructed second language learning*. (pp. 159-180). Amsterdam: John Benjamins.
- Ranta, L. (2005). Language analytic ability and oral production in a second language: Is there a connection? In A. Housen & M. Pierrald (Eds.), *Investigations in instructed second language acquisition*. (pp. 99-130). Berlin: de Gruyter.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6(6), 317-327.
- Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, 81(1), 115-119.
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118(3), 219.
- Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford: Oxford University Press.
- Reber, A. S. (2011). An epitaph for grammar: An abridged history. In C. Sanz & R. Leow (Eds.), *Implicit and explicit language learning*. (pp. 23-34). Washington D. C.: Georgetown University Press.
- Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(5), 888-896.

- Rebuschat, P. (2008). *Implicit learning of natural language syntax*. Unpublished Ph.D. Dissertation. University of Cambridge, Cambridge, UK.
- Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, 63(3), 595-626.
- Rebuschat, P. (2015). *Implicit and explicit learning of languages*. Amsterdam: John Benjamins.
- Rebuschat, P., Hamrick, P., Riestenberg, K., Sachs, R., & Ziegler, N. (2015). Triangulating measures of awareness: A contribution to the debate on learning without awareness. *Studies in Second Language Acquisition*, 37(2), 299-334.
- Rebuschat, P., & Williams, J. (2009). Implicit learning of word order. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society* (p. 1031). Austin, TX: Cognitive Science Society.
- Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, 33(4), 829-856.
- Rebuschat, P., & Williams, J. N. (2013). Implicit learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (Vol. 20). Oxford: Wiley- Blackwell.
- Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics*, 44, 563-575.

- Reingold, E. M., & Merikle, P. M. (1990). On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind & Language*, 5(1), 10-28.
- Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*, 62(1), 93-132.
- Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition*, 18(1), 27-67.
- Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition*, 19(2), 223-247.
- Robinson, P. (2002a). Learning conditions, aptitude complexes, and SLA. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 113-133). Amsterdam: John Benjamins.
- Robinson, P. (2002b). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA: A replication and extension of Reber, Walkenfield and Hernstadt (1991). In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211-266). Amsterdam: John Benjamins.
- Robinson, P. (2005b). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, 25, 46-73.

- Robinson, P. (2005a). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition*, 27(2), 235-268.
- Robinson, P. (2005b). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, 25, 46-73.
- Robinson, P. (2007). Aptitudes, abilities, contexts, and practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. (pp. 256-286). Cambridge: Cambridge University Press.
- Robinson, P. (2012). Individual differences, aptitude complexes, SLA processes, and aptitude test development. In M. Pawlak (Ed.), *New perspectives on individual differences in language learning and teaching*. (pp. 57-75). New York, NY: Springer.
- Roehr, K., & Gánem-Gutiérrez, G. A. (2009). The status of metalinguistic knowledge in instructed adult L2 learning. *Language Awareness*, 18(2), 165-181.
- Rogers, J. (2016). *Developing implicit knowledge of L2 case marking under incidental learning conditions*. Unpublished Ph.D. Dissertation. University College London, London, UK.
- Rogers, J., Révész, A., & Rebuschat, P. (2015). Challenges in implicit learning research: Validating a novel artificial language. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages*. (pp. 275-299). Amsterdam: John Benjamins.

- Rogers, J., Révész, A., & Rebuschat, P. (2016). Implicit and explicit knowledge of inflectional morphology. *Applied Psycholinguistics*, 37(4), 1-32.
- Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence: Quantitative and qualitative analysis*. New York: Peter Lang.
- Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences, and instructional design. In P. Robinson (Ed.), *Cognition and second language instruction*. (pp. 319-353). Cambridge: Cambridge University Press.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics* 11(2), 129-158.
- Schmidt, R. W. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11, 237-326.
- Schmidt, R. W. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. W. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1-63). Honolulu, HI: National Foreign Language Resource Center.
- Schmidt, R.W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.
- Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology*, 45(1), 21-50.
- Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, 16(2), 155-172.

- Shanks, D. R. (2005). Implicit learning. In K. Lambert & R. Goldstone (Eds.), *Handbook of cognition* (pp. 202-220). London: Sage.
- Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, 17(3), 367–447.
- Sheen, Y. (2007). The effects of corrective feedback, language aptitude, and learner attitudes on the acquisition of English articles. In A. Mackey (Eds.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 301-322). Oxford: Oxford University Press.
- Siegelman, N., & Forst, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, 81, 105-120.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Skehan, P. (2002). Theorizing and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning*. (pp. 69-94). Amsterdam: Benjamins.
- Skehan, P. (2012). Language aptitude. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition*. (pp. 381-395).
- Skehan, P. (2015). Foreign language aptitude and its relationship with grammar: A critical overview. *Applied Linguistics*, 36(3), 367-384.

- Skehan, P. (2016). Foreign language aptitude, acquisitional sequences, and psycholinguistic processes. In G. Granena, D. O. Jackson & Y. Yilmaz (Eds.), *Cognitive individual differences in L2 processing and acquisition*. Amsterdam: John Benjamins (pp. 15-38).
- Snow, R. E. (1987). Aptitude complexes. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning and instruction, Vol. 3: Conative and affective process analysis* (pp. 11-34). Hillsdale, NJ: Erlbaum.
- Snow, R. E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational Psychologists*, 27(1), 5-32.
- Snow, R. E. (1994). Abilities in academic tasks. In R.J. Sternberg & R K. Wagner (Eds.), *Mind in context: Interactionist perspectives on human intelligence* (pp. 3-37). New York: Cambridge University Press.
- Soto, D., & Silvanto, J. (2014). Reappraising the relationship between working memory and conscious awareness. *Trends in Cognitive Sciences*, 18(10), 520-525.
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60(2), 263-308.
- Suzuki, Y. (2015). *Using new measure of implicit L2 knowledge to study the interface of explicit and implicit knowledge*. Unpublished Ph.D. dissertation. University of Maryland, College Park, MD, USA.
- Suzuki, Y. (2017). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics*, 38(5), 1229-1261.

- Suzuki, Y., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning*, 65(4), 860-895.
- Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, 67(4), 747-790.
- Tagarelli, K. M., Borges-Mota, M., & Rebuschat, P. (2011). The role of working memory in implicit and explicit language learning. In L. Carlson, C. Hoelscher & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*.
- Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, 17(2), 277-293.
- Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytic ability. In A. Mackey (Ed.): *Conversational interaction in second language acquisition*. (pp. 171-95). Oxford: Oxford University Press.
- Turner, M. L., & Engle, R. W. (1989). Is working capacity task dependent? *Journal of Memory and Language*, 28(2), 127-154.
- Tyler, A. (2012). *Cognitive linguistics and second language learning: Theoretical basics and experimental evidence*. London, UK: Routledge.

- Vafaei, P., Suzuki, Y., & Kachinske, I. (2017). Validating grammaticality judgment tests: Evidence from two new psycholinguistic measures. *Studies in Second Language Acquisition*, 39(1), 59–95.
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1-67.
- Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130292.
- Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, 49A(1), 51-79.
- Wen, Z. E., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. *Language Teaching*, 50(1), 1-31.
- Wen, Z., & Skehan, P. (2011). A new perspective on foreign language aptitude: Building and supporting a case for ‘working memory as language aptitude’. *Ilha Do Desterro: A Journal of English Language, Literatures and Cultural Studies*, 60, 15–44.
- Williams, J. N. (2004). Implicit learning of form-meaning connections. In J. Williams, B. VanPatten, S. Rott & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 203-218). Mahwah, NJ: Lawrence Erlbaum.

- Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, 27(2), 269.
- Williams, J. N. (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 319-353). Bingley, UK: Emerald Group.
- Williams, J. N. (2012). Working memory and SLA. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition*. (pp. 427-441). London: Routledge.
- Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning*, 53(1), 67-121.
- Williams, J. N., & Kuribara, C. (2008). Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua*, 118(4), 522-553.
- Williams, J. N., & Rebuschat, P. (2012). Statistical learning and syntax: What can be learned and what difference does meaning make. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 237-264). Berlin: de Gruyter.
- Winke, P. (2013). An investigation into L2 aptitude for advanced Chinese language learning. *Modern Language Journal*, 97(1), 109-130.
- Woltz, D. J. (1999). Individual differences in priming: The roles of implicit facilitation from prior processing. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and*

- content determinants* (pp. 135-156). Washington, DC: American Psychological Association.
- Woltz, D. J. (2003). Implicit cognitive processes as aptitudes for learning. *Educational Psychologist*, 38, 95–104.
- Yi, W. (in press). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition*.
- Yilmaz, Y. (2013). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics*, 34(3), 344-368.
- Yilmaz, Y., & Granena, G. (2016). The role of cognitive aptitudes for explicit language learning in the relative effects of explicit and implicit feedback. *Bilingualism: Language and Cognition*, 19(1), 147-161.