Abstract

Title of dissertation: AUTOMATIC ACTIVATION OF SEMANTIC REPRESENTATION DURING SECOND LANGUAGE PROCESSING

Written by: Sun Young Ahn, Doctor of Philosophy, 2015

Directed by: Professor Nan Jiang
Department of Second Language Acquisition

The present study is motivated by two questions. First, can late learners of a second language (L2), who begin learning after puberty and are unbalanced bilinguals, activate or visualize the meaning of an L2 word or sentence as quickly as do first language (L1) speakers? Second, if so, what factors—such as L2 proficiency and the amount of its use—contribute to developing native–like efficient processing in L2?

To address these questions, the degrees of automatic semantic activation were compared between L1 and L2 speakers through emotional involvement during word recognition and mental imagery generation during sentential reading. To this end, a total of 60 late–advanced L2 Korean speakers participated in the emotional Stroop Task and the sentence–based picture recognition task along with 36 L1 Korean speakers.

The results revealed that the emotional Stroop effect was not statistically significant in the late L2 group but was significant in the L1 group; whereas the sentence–picture congruency effect was significant in both L2 and L1 groups with similar degrees. This means that late L2 Korean speakers could activate sentence meaning during L2 sentential reading as automatically as L1 speakers but could not activate word meaning as efficiently as L1 speakers. Different degrees of semantic activation among the L2 group across experiments compared to L1 speakers can be considered as cross–task variation; that is, L2 speakers exhibited native–like patterns

when semantic activation was promoted but did not when constrained in the tasks (in a sentence–based picture recognition task and an emotional Stroop task, respectively).

Furthermore, the results showed that the effect of L2 use was positively significant both on the emotional Stroop effect and the sentence–picture congruency effect. These findings suggest that the degree of automatic semantic activation during L2 word recognition, as well as sentence reading can be improved with increased L2 use, despite the late starting age of L2 acquisition. Overall, the present study found positive evidence that late L2 speakers may achieve native–like efficiency in reading comprehension in L2, assisted with the extensive L2 use in addition to high proficiency in L2.

AUTOMATIC ACTIVATION OF SEMANTIC REPRESENTATION DURING SECOND LANGUAGE PROCESSING

by

SUN YOUNG AHN

Dissertation 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
Doctor of Philosophy
2015

Advisory Committee:
Professor Nan Jiang, Chair
Professor Robert DeKeyser
Professor Kira Gor
Professor Steven Ross
Professor Min Wang, Dean's Representative

Acknowledgement

My journey at UMD has been filled with excitement, challenges, and amazing people. I would like to thank all those I met and worked with during my doctoral program for their dedicated assistance, which has made it possible me to successfully complete my journey.

First, I would like to thank all my committee members. I cannot thank my advisor, Dr. Nan Jiang, enough for his advice from the earliest stage of developing ideas for this dissertation and designing the experiments to the last stage of revising the manuscript. I was happy to lean on his insight, experience, and wisdom. I also would like to thank Dr. Robert DeKeyser for inspiring and supporting me throughout my doctoral program and Dr. Steven Ross and Dr. Kira Gor for taking their valuable time to provide insightful comments that upgraded my dissertation.

My sincere thanks also go to my advisor for my master's degree, Dr. Youngkyu Kim at the Ewha Womans' University, for introducing me to this great program and his long–lasting support. Also, my special thanks go to Dr. Diane Larsen–Freeman at the University of Michigan for inviting me to her renowned classes which inspired me to pursue a Ph.D in this field. Dr. I extend my gratitude to Dr. Seonhwa Jeong at the

Chung–Ang University, Dr. Youngjoo Kim at the Kyung Hee University, and Dr. Mark Peterson at the Brigham Young University who provided me with crucial help during the data collection process. With their help, I was able to efficiently recruit non–heritage, advanced learners of Korean. In addition, I am truly appreciative of all the volunteer participants in my study. Further, I want to thank M.A. Mihwa Song, my friend, whose assistance made the data collection more efficient and smooth.

My deepest thanks go to my family. When I started my doctoral program, I was 6 months pregnant with my daughter, Dahnbi. It took as many years as her current age for me— for us—to complete the degree. Her very existence has been the very source of my energy, courage, and motivation to overcome all difficulties. Also, I cannot thank my husband, Hee Mok, enough. Because of him, I was able to dream and finally make it come true. My thanks also go to my mom, Yeonsoo Jeong, and dad, Jonghwa Ahn, and elder brother, Jaeyoung Ahn, for always believing in me, letting me explore to my heart's content, and praying for me at all times. Last but not least, I am grateful for the endless love and support of my parents–in–law, Kyungil Shin and Junshik Park. I love you all.

Table of Contents

Acknowledgement .... **ii**
Table of Contents .... **iv**
List of Tables .... **vi**
List of Figures .... **vii**
Chapter 1. Introduction **1**
Chapter 2. Literature Review **5**
2.1. Current Status of Investigating Automatic Semantic Activation in L2 Processing 5
2.1.1. Weak Activation of Semantic Representation in L2 Processing 5
2.1.2. Comparable Activation of Semantic Representation in L1 and L2 14
2.1.3. Limitations of Previous Studies and Remaining Questions 19
2.2. Examining Semantic Activation through Emotional Involvement 26
2.2.1. The Unique Properties of Emotional Words 26
2.2.2. The L1–L2 Difference in Processing Emotional Words 29
2.2.3. Implications of the Emotion Word Processing for Examination of Semantic Activation 33
2.3. Examining Semantic Activation through Mental Imagery 37
2.3.1. Theoretical Issues on How Meaning is Represented during Language Comprehension: Embodied Cognition 38
2.3.2. Empirical Evidence of the Embodiment Theory 42
2.3.3. Implications of Mental Imagery for Examination of Automatic Semantic Activation 44
2.4. Motivation of the Present Study 45
2.4.1. Research Questions 49
Chapter 3. Pilot Studies **50**
3.1. Overview 50
3.2. Methods 50
3.3. Results and Discussion 52
Chapter 4. Emotional Involvement in L2 Processing (Experiment 1) **55**
4.1. Study Overview 55
4.2. Participants 57
4.3. Methods 61
4.3.1. Materials 61
4.3.2. Procedures 62

4.4. Data Preparation and Analyses....63
4.5. Results....65
4.5.1. Emotional Involvement in L1 Word Recognition....65
4.5.2. Emotional Involvement in L2 Word Recognition....70
4.5.3. Differences in Emotional Involvement between L1 and L2 Word Recognition 75
4.5.4. L2 Proficiency and L2 Use in Emotional Involvement during L2 Word Recognition....76
4.6. Section Discussion....78
Chapter 5. Mental Imagery in L2 Processing (Experiment 2)....82
5.1. Study Overview....82
5.2. Participants....83
5.3. Methods....83
5.3.1. Materials....83
5.3.2. Procedures....86
5.4. Data Preparation and Analyses....88
5.5. Results....90
5.5.1. Mental Imagery in L1 Sentence Reading....90
5.5.2. Mental Imagery in L2 Sentence Reading....96
5.5.3. Differences in Mental Imagery between L1 and L2 Sentence Reading....102
5.5.4. L2 Proficiency and L2 Use in Mental Imagery during L2 Sentence Reading....104
5.6. Section Discussion....107
Chapter 6. General Discussion....110
6.1 Automatic Semantic Activation in Late L2 Speakers....110
6.1.1. Automatic Semantic Activation in L2 Word Recognition....111
6.1.2. Automatic Semantic Activation in L2 Sentence Reading....115
6.1.3. Differences in Automatic Semantic Activation between L2 Word and Sentence Processing....118
6.2 Factors Contributing to Automatic Semantic Activation in L2 Processing....122
Chapter 7. Conclusion....128
Appendices....134
Appendix 1: Stimuli for Experiment 1 (Emotional Stroop Task)....134
Appendix 2: Stimuli for Experiment 2 (Sentence-based Picture Recognition Task)....135
References....140

List of Tables

Table 1. Results of Piloting Emotional Involvement in L2 (English) Processing..........53
Table 2. Results of Piloting Mental Imagery in L2 (English) Processing......................53
Table 3. Results of Piloting with Korean Stimuli..54
Table 4. Advanced Korean L2 Speakers' Language Backgrounds................................60
Table 5. Summary of Remaining Data for Experiment 1..65
Table 6. Summary Statistics of Model 1 on L1 Speakers' Emotional Involvement.......67
Table 7. Summary Statistics of Model 2 on L1 Speakers' Emotional Involvement.......68
Table 8. Summary Statistics of Model Comparison for L1 Speakers' Emotional Involvement..69
Table 9. Summary Statistics of ANOVAs on L1 Speakers' Emotional Involvement.....70
Table 10. Summary Statistics of Model 1 on L2 Speakers' Emotional Involvement.....72
Table 11. Summary Statistics of Model 2 on L2 Speakers' Emotional Involvement.....73
Table 12. Summary Statistics of Model Comparison for L2 Speakers' Emotional Involvement..74
Table 13. Summary Statistics of ANOVAs on L2 Speakers' Emotional Involvement...75
Table 14. Summary Statistics of Interaction between Contributing Factors and Emotional Involvement in L2...77
Table 15. Summary of Remaining Data for Experiment 2..89
Table 16. Summary Statistics of Model 1 on L1 Speakers' Mental Imagery.................91
Table 17. Summary Statistics of Model 2 on L1 Speakers' Mental Imagery.................92
Table 18. Summary Statistics of Model 3 on L1 Speakers' Mental Imagery.................94
Table 19. Summary Statistics of Model Comparison for L1 Speakers' Mental Imagery...95
Table 20. Summary Statistics of ANOVAs on L1 Speakers' Mental Imagery...............96
Table 21. Summary Statistics of Model 1 on L2 Speakers' Mental Imagery.................97
Table 22. Summary Statistics of Model 2 on L2 Speakers' Mental Imagery.................98
Table 23. Summary Statistics of Model 3 on L2 Speakers' Mental Imagery...............100
Table 24. Summary Statistics of Model Comparison for L2 Speakers' Mental Imagery.101
Table 25. Summary Statistics of ANOVAs on L2 Speakers' Mental Imagery.............102
Table 26. Summary Statistics of Interaction between Group and Mental Imagery......103
Table 27. Summary Statistics of Interaction between Contributing Factors and Mental Imagery in L2..106
Table 28. The Findings of the Present Study...111

List of Figures

Figure 1. Schematic of Theories' Position along the Continuum from Amodal to Modal
....40

Figure 2. One Example of L2 Use Data in Language Background Questionnaire.........61
Figure 3. The Stimulus Sequence within a Trial of the Emotional Stroop Task.............63
Figure 4. The Results of Experiment 1..76
Figure 5. An Example of an Image..84
Figure 6. The Stimulus Sequence within a Trial of the Sentence-based Picture Recognition Task...87
Figure 7. The Results of Experiment 2..104

Chapter 1. Introduction

Language comprehension is highly efficient in first language (L1) processing. When listening or reading, L1 speakers can immediately visualize, understand, and perceive what a word or a sentence denotes by routinely applying it to its real-world referent. This is referred to as embodied cognition in language comprehension (for a review, see Pecher & Zwaan, 2005). As bilingual environments become more common, people have more opportunities to read newspapers, novels, academic journals, emails, text messages, product instructions, or advertisements in a second language (L2). To what extent, however, can L2 comprehension be as efficient as L1 comprehension? Can L2 speakers immediately visualize or perceive what a word or a sentence represents in L2 processing as efficiently and completely as L1 processing?

The efficiency of language comprehension may depend on the level of automaticity with which semantic representation is accessed and processed from memory during language processing.[1] If the level of semantic activation in L2 is lower than in L1, L2 reading will take more time and energy than L1 reading. Consequently, within equivalent time constraints, L2 speakers may not understand, visualize, or perceive language as completely as do L1 speakers. This will be the case for most typical late L2 learners who

[1]When evaluating the efficiency of language comprehension in a larger unit (e.g., a sentence or discourse), other factors will affect efficiency, including syntactic complexity, skill in semantic integration, world background knowledge, and so on (Grabe, 1991). The focus of the present paper, however, is the level of overall semantic activation in a given linguistic unit (regardless if it is a word or a sentence) but not the components that attribute more or less to its level. Thus, in the present paper, the term *semantic activation* indicates the overall outcome of semantic processing as an integrative and the highest concept of these components, if any.

are unbalanced bilinguals, with their L1 as their dominant language. Given that native-like efficient language comprehension is the primary goal of L2 learning, however, it is important to understand whether typical late L2 learners can activate semantic representation as automatically as do L1 speakers. If so, they will be able to generate emotions and imagery in L2 reading comparable to those in L1 speakers' minds.

It is not straightforwardly predictable whether late L2 learners will show efficient comprehension in L2 to similar degree with L1 speakers'. Considering that age of onset of L2 acquisition plays a crucial role in ultimate L2 attainment (for a review of age effects, see DeKeyser, 2012), late L2 learners, defined as those who began learning their L2 after puberty, may be less likely to obtain native-like automatic semantic activation in L2 processing. It is possible, however, that late L2 learners will reach a native-like level in L2 semantic processing because semantics have been shown to be least affected by age of onset of L2 acquisition (Slabakova, 2006; Weber-Fox & Neville, 1996).

Although either case is possible, neither has been investigated directly; as such, this is an important issue to pursue for both research and practice. This is especially true when it concerns L2 learners of less commonly-taught languages, such as Korean, who usually begin L2 learning in college and are thereby late learners in most cases (an exception here is heritage speakers). Answering the questions of whether L2 learners can activate semantic representation in L2 as automatically as do L1 speakers and how the level of semantic activation changes as proficiency and use of L2 increase have crucial implications for classroom learning, such as establishing separate tracks for early or heritage learners and late or nonheritage speakers. Furthermore, if late L2 learners ultimately can reach automatic semantic activation at a native-like level, it would be significant to know what factors contribute to this activation or eventually lead to native-like efficient comprehension in L2.

Whereas many studies examining semantic representation have focused on the relative strengths of word-and-concept connections in L1 and L2, a great deal of inconsistency has emerged among the findings on this topic. Automatic semantic activation in L2 needs further research qualitatively for several reasons. First, the tasks used in the previous literature tend to allow strategic processing rather than *automatic* processing (e.g., a translation task in the study of Finkbeiner and Nicol, 2003). Even if strategic processing cannot be prohibited completely, it should be prevented as much as possible; indeed, automatic processing must be the main mechanism to access semantic information. If both strategic and automatic processing occur, the result (semantic representation) is likely to be confounded with explicit knowledge, memory capacity, or other factors such as lexical-level representation activation. In future studies, ways to assess automatic semantic activation should be improved.

Second, most previous studies have given less consideration to potentially important factors for semantic system development. In many studies, neither the individuals' L2 proficiency nor L2 use were manipulated systematically or adequately controlled. Instead, participants' L2 proficiency has been determined based on self-reports, which are subjective and nonstandard, and L2 use has barely been considered as a factor affecting the degree of automatic semantic activation. In the sense that L2 proficiency (e.g., Kroll & de Groot, 1997) and L2 use (e.g., Duyck & Warlop, 2009) may play a critical role in semantic development, we must account for these two factors.

The aims of the present study, therefore, are twofold. First, the present study attempted to investigate to what extent semantic representation can be activated automatically in L2 speakers compared to L1 speakers. To assess the level of semantic activation more directly than in previous studies, the emotional involvement and mental imagery that are generated automatically in L2 speakers' minds were examined and compared to those in L1 speakers' minds. Second, the present study aimed to observe how the level of semantic activation changed according to L2 proficiency and use. L2 proficiency was scrutinized using an objective language assessment rather than subjective self–report, and the accumulation of L2 use was calculated in details. This calculation was based on hours of L2 use through interactions in L2 (i.e., L2 is used in situations beyond the classroom), media (television, movies, L2 reading for pleasure), and so on since the late L2 Korean speakers began learning Korean.

In the next chapter, previous studies on the L1–L2 differences in semantic activation are reviewed along with the methodological drawbacks embedded in these studies. Then, the theoretical and empirical backgrounds that support Experiment 1 (emotional involvement) and Experiment 2 (mental imagery) are discussed as alternatives to that which has been researched to date.

Chapter 2. Literature Review

2.1. Current Status of Investigating Automatic Semantic Activation in L2 Processing

It has often been assumed that semantic information is more accessible in L1 than in L2. However, prior studies that employed diverse ways to assess semantic activation in L1 and L2 processing have produced discrepant results. In this chapter, these studies are reviewed according to their findings.

2.1.1. Weak Activation of Semantic Representation in L2 Processing

The hypothesis that semantic representation is activated less automatically in L2 than in L1 has been inferred through L2 speakers' lower sensitivity to semantic manipulation than that of L1 speakers. During L2 processing, L2 speakers have shown weak or no semantic priming effects on word recognition (e.g., Sholl et al., 1995; Zhao et al., 2011); difficulty with semantic integration on completing L2 sentences compared to L1 speakers (Hu & Jiang, 2011; Weber-Fox & Neville, 1996); and delayed recognition of semantic anomalies in L2 sentential reading (Alvarez et al., 2003; Ardal et al., 1990; Weber-Fox & Neville, 1996).

Accumulated research on cross-language semantic priming effects provides support for weak semantic activation in L2 word processing. One of the approaches used in prior studies is investigating semantic priming effects. For example, L2 speakers were asked to perform a lexical decision task on word targets preceded by a semantically related prime (either a semantic associate or a translation in the participants' other language) or a semantically unrelated prime. The common finding among the studies using this approach was that L1 prime words produced a stronger semantic priming effect on L2 targets than the reverse (for more detailed explanations of the asymmetric priming according to language directions, see Jiang, 1999). This suggests more automatic semantic activation through L1 words than L2 words. The basic concept behind this paradigm is that when a word is presented, the activation spreads to those words that are semantically related or associated to the presented word (Collins & Loftus, 1975). Therefore, studies showing less semantic priming effects by L2 primes than by L1 primes suggest that activation can occur through L1 words more automatically than through L2 words.

For example, Zhao and colleagues (2011) showed an asymmetric semantic priming effect (stronger effect in L1–L2 than L2–L1) on a lexical decision task, which differed according to participants' L2 proficiency and the L2 learning context. The participants included three groups of L1 Chinese–L2 English speakers with different L2 learning backgrounds and L2 proficiency levels. One group was proficient in English as a second language (ESL), another group proficient in English as a foreign language (EFL), and a third group was less proficient in EFL. In this study, the three groups displayed different patterns in semantic priming effects according to the priming directions and their L2 proficiency and learning contexts. The two EFL groups, both high-proficiency and low-proficiency, produced significant semantic priming in the L1–L2 direction but not in L2–L1 direction. This implies lower semantic activation in L2 than in L1 among EFL groups regardless of the participant's L2 proficiency. In contrast, the high-proficiency ESL group produced significant semantic priming effects in both directions, although the priming effects were smaller in the direction of L2–L1 (25.35 ms) than in the direction of L1–L2

(95.91 ms). Notably, even when both directions produced significant semantic priming effects through translation equivalents, the priming effect was still larger in the L1–L2 direction than in the L2–L1 direction, implying more automatic semantic activation in L1 than L2.

When the primes were semantic associates rather than translations in the participants' other language, however, even the high-proficiency ESL group, who produced significant priming effects in both directions, did not show semantic priming effects in a L2–L1 direction (Zhao et al., 2011). It is known that priming is usually stronger when primes are translation equivalents than when they are semantically related words (for a review, see Altarriba & Basnight–Brown, 2008). Indeed, translations in two languages exhibit more overlap in semantic information across languages than do semantically related words, such that when the meaning activation of a word in one language spreads to words in the other language, this can activate a translation equivalent more so than the semantic associate. For example, in De Groot and Nas' study (1991) where L1 Dutch–L2 English speakers judged whether letter strings were words in L2 or not, their responses were sped up by translation equivalents more than by semantic associative words in L2, regardless of whether the primes were cognate or noncognate and masked or unmasked.

These findings based on L2 speakers even at a later stage of L2 development are consistent with the tendency that semantic priming is not as reliable in a L2–L1 direction as in a L1–L2 direction (Zhao et al., 2011). Therefore, it can be inferred that semantic activation in L2 is neither as strong nor as automatic as it is in L1 even among high–proficiency L2 speakers.

In the same vein, Sholl and colleagues (1995) reported weak semantic activation among L2 speakers through asymmetric semantic priming by using a transfer paradigm. In this study, 24 L1 English–L2 Spanish unbalanced bilinguals performed a picture-naming task in L1 and L2 and then translated the named and new words both forward (L1 to L2) and backward (L2 to L1). Their translation of L1 words into L2 was completed significantly faster when the L1 words were named in the previous picture-naming task (i.e., they were previously activated concepts) either in L1 or L2 than when they were not (i.e., they were new concepts). On the other hand, the participants' backward translation was not affected by L2 words whether or not they were named previously. Given that picture–naming requires conceptual access, it is feasible that old concepts were pre–activated through naming, but new concepts were not. However, because forward translation but not backward translation was primed by picture naming, this finding confirms that L1 may be more sensitive to prior activation of semantic representation, whereas L2 is not.

In addition to the asymmetric priming effects, physiological evidence supports weak semantic activation in L2 word processing (Chee et al., 2001). In this study, more or less proficient L2 English speakers were compared on a "Pyramids and Palm Tree" task. As the participants judged which word was closest to a sample word in meaning among the stimulus triplets in this task (e.g., for *pillow* as a sample item, with *bed* and *chair* as comparatives), changes in their blood oxygen level dependent (BOLD) signal were measured through functional magnetic resonance imaging (fMRI) along with their response time. The results showed that participants who were lower in L2 proficiency registered greater BOLD signal changes and took a longer amount of time to respond. In other words, less proficient L2 speakers took more time to judge semantic closeness and showed more fluctuation in BOLD signal changes than more proficient L2 speakers. These compatible patterns with two measurements imply that with a less well–developed L2 semantic system, the less proficient L2 speakers may have more difficulty processing semantic information in their non-native language than the more proficient L2 speakers. The results showed that the less proficient L2 speakers were in their second language, the more time (longer reaction time) and effort (additional areas of blood activation) they required to process semantic information in L2.

Such weak semantic activation in L2 has appeared not only in word processing but also in sentence processing. Behavioral and neurocognitive evidence has shown less semantic activation in L2 sentence processing than in L1. For example, Hu and Jiang (2011) and Weber-Fox and Neville (1996) showed that L2 speakers processed semantic aspects during L2 sentential reading less accurately and less specifically than L1 speakers. More specifically, in Hu and Jiang's study (2011), 27 L1 Chinese–L2 English speakers performed a lexical decision task to complete an English sentence task along with L1 English speakers. In the task, an English sentence with a blank at the end was presented aurally followed by English letter strings. Participants had to decide whether the given word was an English word or not by pressing either a *yes* or a *no* buttons. The *yes* response target words belonged to one of the three semantic congruency conditions according to the sentence context and a target: congruent, neutral, and incongruent. For example, after listening to a sentence "*While her husband wants to have a son, Kate has always wanted to* *have a beautiful* _____", each participant saw one of the following words: daughter (congruent), girl (neutral), or war (incongruent).

The rationale here is that if the sentence context primed the target, respondents' lexical decision of a target word for a *yes* response would be affected by the semantic congruency conditions. Indeed, the participant pressed a *yes* response more quickly on congruent targets than on incongruent targets. Both groups responded as *yes* significantly faster when the target was congruent to the sentence context than when the target was not. Interestingly, only L1 English speakers showed different patterns between the neutral and incongruent conditions, whereas L2 English speakers did not. As shown in the sample sentence above, the neutral targets were not exactly predictable but were still semantically acceptable. Thus, the result that L2 speakers behaved similarly with neutral and incongruent conditions implies that they considered semantically acceptable targets to be incongruent when they were unexpected. In other words, when the target word is highly expected based on the context, L2 speakers can process semantic information as efficiently as do L1 speakers. When the semantic information was not highly expected, however, L2 speakers could not incorporate unexpected but semantically acceptable words into their mental representation of the sentence meaning. Given the difficulty L2 speakers have with semantic integration, which does not appear in L1 speakers, it can be implied that automatic semantic representation is activated in only a limited way during L2 processing. In other words, L2 participants tend to stick with their current mental representation too strongly to change it in an automatic manner.

Related to L2 speakers' difficulty with semantic incorporation, Weber–Fox and Neville's study (1996) exhibited how late L2 learners complete sentences inaccurately. In this study, 61 L1 Chinese–L2 English speakers, whose age of L2 exposure occurred between 3– to 18–years old, judged whether L2 sentences were accurate or not when a critical word was presented after the entire sentence was read. For example, after reading the sentence *"The scientist criticized Max's _____ of the theorem*," participants were given either the word *proof* (for a *yes* response) or *event* (for a *no* response). Only late L2 speakers, whose age of L2 exposure began at 16 years or older exhibited significantly lower accuracy in semantic judgment compared to L1 speakers. Such semantic misjudgment by late L2 speakers may be due to their difficulty in semantic integration and reflects a relatively low level of semantic activation in L2 sentential reading compared to earlier L2 speakers or L1 speakers.

In addition, this behavioral evidence of weak semantic activation during L2 sentential reading was supported by neuro–imaging evidence in the same study (Weber-Fox & Neville, 1996). In this study, event-related potentials (ERPs) were registered as the participants were presented with English sentences word–by–word as a self–paced reading task. Interestingly, the peak of N400, which is a good index of processing semantic anomalies in the ERP research (e.g., Kutas & Hillyard, 1980), was observed after the onset of a semantically violated word in late L2 speakers later than L1 speakers. Given the low semantic accuracy and later N400 peak in L2 processing, we can infer that L2 speakers, especially late L2 speakers (post–puberty L2 exposure), may not be able to activate semantic representation either as completely or as quickly as L1 speakers.

Ardal et al.'s study (1990) added neurocognitive evidence to support a deficit in semantic activation in L2 processing. In this study, N400 latency was compared in monolinguals' L1, bilinguals' L1, and bilinguals' L2 during English sentence reading with either a semantically congruent or incongruent word presented at the end of a sentence. The results showed that the N400 peak appeared significantly faster in monolinguals' L1 processing than bilinguals' L1 processing, with bilinguals' L2 processing being the slowest. The N400 delay in L2 processing as opposed to L1 suggests that less fluent language may activate semantic representation less automatically than can more fluent language. Therefore, the time needed to detect semantic incongruence in a less fluent language is longer than that required in a more fluent language.

In addition to N400 peak latency, the stimulus repetition effect on N400 provided more evidence regarding automatic semantic activation during L2 word processing. When a word appears a second time in context, it tends to alleviate semantic processing, resulting in reduced N400 amplitude (for a review, see Rugg, 1995). For example, Alvarez et al. (2003) observed repetition effects on N400 patterns produced by 28 L1 English–L2 Spanish unbalanced bilinguals in a mixed-language semantic categorization task. In this task, participants saw a word and then judged whether the target belonged to part of the human body. Their semantic category judgments were made for three conditions: when the target word was first presented (first presentation); when the word was repeated in the same language (within-language repetition); and when the word was repeated in the other language (between-language repetition). The repetition effect was measured based on the gap between the amplitude of N400 on the first presentation of the word and that of N400 on the word's repetition. This resulted in more repetition effects based on a higher N400 peak on the first presentation or based on a lower N400 peak on the second presentation.

Alvarez et al.'s (2003) results showed that Spanish (L2) words in within-language repetition produced the highest repetition effect; English (L1) words in within-language repetition produced the second highest repetition; English words in between-language repetition (L2–L1) showed the second lowest repetition; and Spanish words in between-language repetition (L1–L2) showed the least repetition effect. This reveals that overlaps in both form and meaning may contribute to the repetition effect, because within-language repetition produced more effects than between-language repetition regardless of language. Notably, however, the L2 word evoked N400 most when it appeared in an initial position.

More important for the present study, however, is that the N400 peak appeared earlier when L1 words followed L2 translations (L2–L1 repetition) than when the reverse (L1–L2 repetition) was the case. This finding can be understood in two ways. First, the L1 word is affected by prior semantic activation more than the L2 word, and second, the L1 word activates its meaning more automatically than the L2 word. In L2–L1 repetition, the immediate repetition facilitated L1 semantic activation, resulting in relatively speedy priming effects. In contrast, in L1–L2 repetition, although the L2 word's meaning just activated by the L1 word was available, the L2 word had to first activate the L1 lexical entry in addition to its L2 lexical entry to arrive at the meaning. In discussing their findings, Alvarez et al. (2003) noted, "This additional operation tended to shift the bulk of the priming effect to a later epoch. This view suggests that the later priming effect seen for the L1–L2 condition is actually a delayed N400 effect" (p. 301).

Considering the evidence revealed by the reaction time studies, as well as the ERPs and fMRI studies, the level of semantic activation in L2 seems to be inherently lower than that in L1. In the studies reviewed above, L2 speakers consistently showed incomparable semantic activation in two languages; that is, there was less sensitivity to semantic interference in L2 word processing (e.g., Sholl et al., 1995) or more difficulty in semantic integration during L2 sentence processing (e.g., Hu & Jiang, 2011), compared to L1 processing. In addition, L2 speakers needed more time and energy to judge semantic categorization of L2 words (e.g., Alvarez et al., 2003) and to detect semantic anomalies in reading L2 sentences (e.g., Ardal et al., 1990) than did L1 speakers. The findings reported above, therefore, suggest that L2 speakers' semantic activation may be deficient compared to L1 speakers', both in L2 word and sentence processing.

2.1.2. Comparable Activation of Semantic Representation in L1 and L2

Not all research, however, supports the hypothesis that automatic semantic activation is relatively weak in L2 processing, particularly at low L2 proficiency levels. Several conflicting findings exist as well.

Symmetric semantic priming effects in both language directions, which supports comparable semantic activation in two languages, were found during word processing even at an early stage of L2 development. For example, Duyck and De Houwer (2008) reported strong semantic priming effects in L2 processing on a more conservative task, which was assumed to trigger semantic information less than other tasks (e.g., a translation task). These authors tested 16 L1 Dutch–L2 English speakers with a letter–case judgment task in both L1 and L2. In this task, participants were asked to judge targets' letter–case using verbal labels. That is, they responded either “animal” or “occupation” to uppercase targets or lowercase targets, respectively, regardless of the target’s meaning. In *both* L1 and L2, participants’ responses were faster when the target’s meaning matched the semantic category of the response than when it did not (e.g., a response of “animal” was faster for LION than for LAWYER). In this study, the congruency effect can be considered automatic semantic activation; although the participants tried to disregard the word’s meaning, its semantic representation was activated so automatically and strongly that they could not suppress it completely. Notably, judging letter–case can be performed accurately without semantic processing. Therefore, the similar congruency effects between L1 and L2 by unbalanced, late bilinguals living in an L1 environment imply that L2 word forms can access their underlying semantic representation as automatically as do L1 words (Duyck & De Houwer, 2008).

Finkbeiner and Nicol (2003) also added contrary evidence to the weak activation of semantic representation during L2 word processing by studying semantic categorization effects on learning L2 words. In this study, 47 English monolinguals were asked to learn 32 novel words in either a semantically categorized or an unrelated order. They were then asked to translate those words in both directions. The participants performed the translation task more slowly on the categorized list than on the unrelated list with similar semantic interference effects in both translation directions. This finding suggests that both L1 and L2 words involve semantic activation to a compatible degree, thus contradicting asymmetric semantic activation between two languages.

In addition, some cross–language priming studies have shown a strong semantic priming effect not only in the direction of L1–L2 but also in the L2–L1 direction among distinctive populations. For example, Duñabeitia et al. (2010) examined highly balanced and simultaneous L1 Basque–L2 Spanish speakers in a lexical decision task in which primes were either cognates or non-cognates. In this study, non–cognates were translation equivalents with different spellings and sounds in the two languages (e.g., the Spanish word *mesa* and its English translation *table*). Cognates, in contrast, were translation equivalents with the same origin and usually a similar spelling or sound pattern (e.g., the Spanish word *rico* and its English translation *rich*) (Duñabeitia et al., 2010). Although cognates produced stronger semantic priming than non-cognates, semantic priming effects were generally similar in both the L1–L2 and L2–L1 directions, regardless of the cognate status of the prime words.

At first glance, this may seem to contradict the notion of asymmetric semantic activation between two languages. However, the study's participants were simultaneous and balanced bilinguals who were likely to reach native–like semantic development in the L2. Thus, the symmetric semantic priming effect shown in Duñabeitia et al.'s study (2010) cannot be considered contrary to the weak activation of semantic representation among late L2 speakers. Rather, Duñabeitia et al.'s study (2010) suggests that comparable degrees of semantic activation in L1 and L2 can be achieved, provided L2 proficiency is as high as L1 proficiency. Therefore, participants' L2 proficiency should be manipulated as an important factor in future research on semantic activation in L2 processing.

Two additional studies using the masked priming paradigm also provided evidence of symmetric semantic activation through similar degrees of semantic priming effects among early bilinguals with age of L2 acquisition earlier than age 6 (Basnight–Brown & Altarriba, 2007, experiment 2; Duyck & Warlop, 2009). In the masked priming paradigm, the prime usually follows a mask (e.g., #######) and is virtually invisible. With this characteristic, the masked priming is more likely to involve an automatic process rather than a strategic one (Lucas, 2000). Both balanced or early bilinguals in Basnight–Brown and Altarriba's study (2007) and Duyck and Warlop's study (2009) exhibited translation priming effects in both directions despite using a mask. As such, the authors highlighted that semantic interference occurred from L2 to L1 even when they assumed that a conscious strategy was blocked. These findings support the notion that automatic semantic activation in L2 may be as strong as it is in L1 when individuals have reached native–like proficiency in L2. (The importance of measuring and including participants' L2 proficiency as a potential factor to affect the level of semantic activation in L2 will be discussed in more detail in the following chapter.)

Finally, Duyck and Brysbaert (2004) showed equivalent magnitudes of semantic activation between L1–L2 and L2–L1 translations. They demonstrated semantic interference effects in backward and in forward translation through the number magnitude effect. The number magnitude effect occurs when it takes longer to translate number words indicating larger quantities (e.g., huit and acht [8]) than number words indicating smaller quantities (e.g., deux and twee [2]). The rationale is that if number words in a target language automatically activate their meanings, the number magnitude effect is likely to occur, independent of word frequency effect. Thus, the number magnitude effect is considered an outcome of semantic processing of number words. In Duyck and Brysbaert's study (2004), L1 Dutch–L2 French bilinguals were asked to name a number word that appeared in Arabic, their L1 (Dutch), and L2 (French), or a newly learned language for this task (Estonian). The results revealed significant semantic effects of number magnitude not only in forward translation (both L1–L2 and L1–just learned language) but also in backward translation (both L2–L1 and just learned language–L1) to comparable degrees. The similar size in number magnitude effects among unbalanced bilinguals with newly learned language for the task in both translation directions was considered evidence for the compatible semantic activation in L1 and L2 even at a low stage of L2 development. (The potential word type effect will be discussed in more detail in the following chapter.)

Comparable semantic activation in L1 and L2 was typically found in studies examining semantic activation at the lexical processing with one exception, which Hahne and Friederici (2001) revealed at the sentence processing level. They compared 12 L1 Japanese–L2 German speakers with L1 German speakers on ERP patterns when the participants listened to German sentences that were either semantically correct or incorrect (e.g., *Das Brot wurde gegessen*: The bread was eaten, or *Der Vulkan wurde gegessen*: The volcano was eaten). Both L1 and L2 groups produced the N400 peak in listening to semantically incorrect sentences, and no statistically significant difference was found in the amplitudes of N400 peaks between the two groups. This finding suggests that semantic processes do not necessarily slowdown in late L2 learners, provided they have at least "some" knowledge of the second language. (This obscure and inaccurate L2 proficiency assessment is noted as a limitation of the previous literature in the following chapter.)

2.1.3. Limitations of Previous Studies and Remaining Questions

The studies reviewed have reported divergent results regarding the relative levels of automatic semantic activation in L1 and L2 processing. To resolve discrepancies among the previous literature and to provide additional evidence regarding semantic activation in L2, the present study attempted to improve three methodological aspects: 1) assessing automatic semantic activation from new perspective which can disambiguate semantic processing from non–semantic processing, 2) controlling L2 speakers' levels of L2 proficiency and amount of L2 use, and 3) including both word and sentence processing.

First, the task of assessing semantic activation should encourage and guarantee *semantic* processing. The tasks employed in some of the previous literature, however, tend to entangle semantic processing with lexical processing, although they were interpreted as outcomes of purely semantic processing. The results prior authors reported are likely to be confounded with non–semantic processing such as lexical confusion and word type effect. This is the case because in previous studies, most tasks required participants to map given linguistic forms to other linguistic forms according to semantic intervention (e.g., a translation task with semantic primes). This roundabout way to examine semantic activation encompasses two potential problems. First, the outcome of language processing can be related more closely to other factors such as the degree of shared semantic information between a prime word and a target word than the degree of semantic activation of the word itself. For example, number words have a privileged position in producing semantic priming compared to other words, because their semantics are entirely shared across languages. Second, when word A in one language and word B in the other language are similar in form, this overlap can facilitate producing output (e.g., a response). This can create an impression that the word is semantically primed, even though the output comes mainly from lexical–level processing.

For example, the symmetric semantic activation shown in Duyck and Brysbaert's (2004) study seemed boosted by word type effect. Here, semantic activation in L2 may be as strong as in L1 for certain types of words (such as number words and color words) in that what they refer to is almost exactly the same across languages. According to the Distributed Feature Model (de Groot, 1992), the extent to which semantic representations are shared across languages can be determined by a word's lexical category. In this sense, we do not know whether the number magnitude effect produced by a number word in L2 resulted from pure semantic processing of it or from a combination of fast lexical processing from L2 to L1 and semantic processing of the number word in L1

Another example of semantic activation being entangled with other lexical factors in the previous literature is found in Finkbeiner and Nicol's study (2003). The symmetric semantic activation in L1 and L2 might possibly result from the design entangling semantic activation with non–semantic cues. Specifically, Finkbeiner and Nicol (2003) argued for symmetric semantic activation in L1 and L2, because words on a semantically categorized list during training were translated more slowly than words on a random list in both directions. It is doubtful, however, whether the new words that were created artificially for the study (e.g., dax, zek, ecrus, etc.) could trigger meanings as completely and as appropriately as did L1 words. Participants learned the new words in short training sessions that lasted only a few minutes across two days. Specifically, the procedure of a vocabulary training session was as follows: participants heard a new word, saw the word with its corresponding picture for 500 ms, heard the word again, and verbally repeated the word twice. After completing this procedure for each of 32 target items, they performed a recognition task in which they judged whether each of the 64 pairs of a picture and a new word was a correct pair or not. Considering the great number of new words and the short time to learn them, the semantic connections of these new artificial words were likely to be much weaker than those of L1 words.

How, then, could the new words display the semantic category effect in backward as well as forward translations (Finkbeiner & Nicol, 2003), which implied that their semantic connections were established strongly enough at the early stage of L2 development? The participants in Finkbeiner and Nicol's study (2003) could have linked an L2 form automatically to its competitors as seen together during the training session, not necessarily to its meaning as the researchers intended. In this case, the words on the same semantic category list could have more competitors than those on the random list. This is because the semantically categorized words were presented in a blocked group, whereas the non–semantically categorized words were not presented as such during training. Thus, semantically categorized words, which were always presented together, could have more competing words at a lexical level rather than semantic than randomized words in the mind, producing a greater delay in translation. Such lexical confusion likely produced pseudo–semantic priming effects when storing, recognizing, retrieving, and producing the new word. Indeed, Jiang and Forster (2001) showed that recently acquired L2 lexical items may be represented episodically not lexically. Given that data from about half of the participants (23 of 47) were discarded due to accuracy lower than 80%, it is possible that the new words were not learned correctly enough to produce semantic priming. Thus, the comparable semantic categorization effects in translation found by Finkbeiner and Nicol (2003) may not be due solely to semantic activation of the newly learned words. Thus, a new approach to disentangle semantic processing from possible lexical processing would be useful to examine semantic processing more directly than what has been done with multiple lexical routes.

In addition, to observe more dynamic aspects of semantic activation in L2, the present study tried to estimate and consider more carefully the participants' language backgrounds. Although L2 proficiency is suggested to influence the degree of automatic semantic activation during L2 word processing (Kroll & Stewart, 1994; Kroll & De Groot, 1997; Jiang, 2000), participants' L2 proficiency was neither measured carefully nor considered important in the previous literature. As briefly mentioned in the previous chapter, some studies have provided clues for the potentially crucial role of L2 proficiency in the level of semantic activation in L2 processing. In particular, studies with highly proficient L2 speakers (e.g., Duyck & Brysbaert, 2004; Duyck, & De Houwer, 2008) or balanced bilinguals (e.g., Basnight–Brown & Altarriba, 2007; Duñabeitia et al., 2010) have shown similar degrees of semantic activation in L1 and L2. For example, Duñabeitia et al. (2010) showed similar semantic priming effects in L1 and L2; the participants were simultaneous and presumably balanced L1 Basque–L2 Spanish bilinguals who were likely to fully develop the semantic structure as native speakers of both languages. Thus, considering the crucial role of L2 proficiency in L2 semantic system development, participants' L2 proficiency should be measured by a more reliable screening device than a self-rating and should be controlled for future studies to examine semantic activation in a more dynamic way.

With regard to participants' language backgrounds, L2 use should be considered as well. The extensive use of L2 can enhance different types of knowledge (e.g., explicit to implicit knowledge through practice; for details, see DeKeyser, 2003). That is, highly automatized use of explicit knowledge may lead to it becoming implicit (DeKeyser, 2001). For example, in Duyck and Warlop's (2009) study, which found comparable semantic activation in both languages, participants were L1 Dutch–L2 English speakers and described as less proficient L2 speakers. However, their L2 proficiency was not scrutinized carefully; a self–rating on a 7–point Likert scale was used to assess their L2 abilities for reading, writing, and speaking (4.2, 3.8, and 3.9 out of 7, respectively). More importantly, the participants reported frequent use of L2 on a daily basis (e.g., popular television dramas, etc.) and had learned the L2 formally since elementary school in Duyck and Warlop's (2009) study. Thus, although the participants' self–ratings indicated they were intermediate L2 speakers, their semantic activation could reach a native–like level with their extensive use of the L2.

Another example that suggests the importance of scrutinizing L2 use as well as L2 proficiency comes from Hahne and Friederici (2001). This study showed native–like semantic activation in L2 sentence comprehension by late L2 speakers. The L2 speakers described in the study, however, seemed to be quite advanced L2 users. They were studying at German universities and had lived in Germany for an average of 29 months. More importantly, they reported German as their most frequently used language in daily life (more than 65% on average). As such, the late L2 speakers in Hahne and Friederici's (2001) study with their favorable amount (as opposed to "some" as described by Hahne and Friederici) of L2 knowledge might have used the L2 extensively. This could have helped increase their semantic activation to a native–like level. Although the L2 speakers marked their L2 proficiency as 3.5 out of 6, they were more likely to be advanced L2 speakers. Thus, in future studies, L2 use and L2 proficiency should be objectively examined and included to understand the developmental aspects of semantic activation.

Third, although automatic semantic activation can vary according to processing levels, it has not been observed at diverse levels within a single study. That is, the issue of automatic semantic activation in L2 processing has been approached at different levels in previous studies, and these studies consequently have drawn conflicting conclusions. The weaker semantic activation in L2 than in L1 typically has been associated with studies that examined the issue in sentence processing (e.g., Hu & Jiang, 2011; Weber–Fox & Neville, 1996; Ardal et al., 1990; with the one exception of Hahne & Friederici, 2001), whereas symmetric results were obtained mostly from studies that examined lexical processing (Duñabeitia et al., 2010; Duyck & De Houwer, 2008; Finkbeiner & Nicol, 2003). Notably, semantic activation in word recognition occurs through isolated words and bottom–up processing, whereas semantic activation in sentential reading involves semantic integration and top–down processing. Both approaches are considered in the present study to understand its diverse aspects more thoroughly.

To summarize, previous studies possess some limitations to generalizing their findings on relative levels of semantic activation in L1 and L2. To improve such drawbacks, future study needs to disentangle semantic processing from other types of representation (e.g., lexical representation), more carefully control crucial variables for semantic system development (i.e., participants' L2 proficiency and L2 use), and approach this issue at both the word and sentence levels of processing.

Despite these improvements for future research, questions still remain: How can we assess automatic semantic activation in L2 more directly at both the lexical and sentential levels? How does it develop with increased L2 proficiency and greater L2 use? Regarding the first question, there has been a desperate need for a new approach to this research topic (e.g., see Duyck & De Houwer, 2008). Devising tasks to disambiguate semantic processing from other confounding factors will help suggest more exact semantic activation. One way to assess semantic activation more directly is to adopt a task that focuses on how participants link the linguistic forms to real-world referents. For these reasons, prior research has called for tasks that require individuals to access semantic information directly and not through multiple lexical routes (Altarriba & Basnight–Brown, 2008). Indeed, if a task requires L2 speakers to link a linguistic form to its real–world referent in each language, it may not be affected by the degree of form or meaning overlap between the two languages.

In the present study, two new approaches that tap automatic semantic activation were administered to disentangle semantic processing from non–semantic processing (lexical or episodic). The first approach is emotional involvement; the second is mental imagery. The following chapters introduce how these two approaches allow us to approach automaticity in semantic activation more directly than has been achieved in previous studies.

2.2. Examining Semantic Activation through Emotional Involvement

Given the need for an innovative approach that focuses on how participants link linguistic input to its real-world referent, instead of their lexical equivalent, examining emotional involvement during word recognition can provide more direct evidence regarding semantic activation. This is because emotional involvement can be observed when emotional words truly trigger the emotional state of mind. Emotional words directly refer to abstract concepts; these are particular affective states (e.g., *happy*, *angry*) or processes (*to worry*, *to rage*) and function to either describe (*she is sad*) or express (*I feel sad*) what we experience. Therefore, although emotional words convey abstract meaning, they can be learned through experience like concrete words by associating a certain emotion and a word meaning. This characteristic of emotional words will help us look into automatic semantic activation more closely. In the following chapter, it is discussed in more details.

2.2.1. The Unique Properties of Emotional Words

Emotional words “denote emotional states, moods, or feelings” (Vigliocco, Meteyard, Andrews, & Kousta, 2009, p. 222). Emotional words possess unique properties in terms of meaning representation. They indicate a specific state of mind charged with affection, mood, or feeling and thereby invoke the corresponding internal affective states (Pavlenko, 2008). This is because emotional words have additional components that differentiate them from other types of words such as abstract and concrete words, namely, valence and arousal. The emotional valence means pleasant or unpleasant feeling of an event, object, or situation (Barrett, 2006). For example, emotional words denoting negative feelings such as *angry* and *scary* have a negative valence while words referring to positive feelings such as *happy* and *excited* have a positive valence. In many studies, the emotional valence of words was determined by participants’ rating, for instance, on 5 points Likert scales from 1 (an extremely negative valence) to 5 (an extremely positive valence) where neutral words are usually rated around the middle, 2 to 3. (e.g., Altarriba, 2003; Dewaele, J.-M., 2004). On the other hand, arousal is the extent to which a physiological and psychological state is activated or deactivated to stimuli (Barrett, 2006). Thus, emotional arousal caused by emotionally charged words have been measured by skin conductance responses (SCRs) and recorded to evoke higher SCRs than neutral words in many studies (e.g., Harris, Ayçiçegi, & Gleason, 2003; Harris, 2004; Harris, Gleason, & Ayçiçegi, 2006). Due to the unique characteristics of emotional words, some authors even suggest categorizing emotion words separately from abstract and concrete words (Altarriba & Bauer, 2004; Pavlenko, 2008; Ayçiçegi -Dinn, & Caldwell-Harris, 2009).

Recent studies have supported the mutual interaction between the affective and cognitive systems (for review see Dolan, 2002). For example, Gaillard et al. (2006) and Naccache et al. (2005) showed that “processing an emotional valence in words engages the same region in a brain (the primarily subcortical system) as in processing emotion from non–verbal stimuli (e.g., faces)” (as cited in Vigliocco, Meteyard, Andrews, & Kousta, 2009). These findings dispute the traditional view in cognitive psychology and cognitive neuroscience where emotional information and semantic information are clearly divided. Traditionally, thus, non–linguistic expression of emotions (e.g., expression and recognition via face and voice) had been a prior interesting area to investigate emotion. Recently, however, language referring to emotions has attracted more and more attention.

There are accumulated studies showing that processing emotional words has distinctive characteristics as compared to processing neutral words. Negative words, especially, are better identified than positive words or neutral words (Gaillard et al. 2006) and trigger long lasting effects in the amygdala (Naccache et al. 2005). That is, since negative words automatically recruit more attention than neutral words (e.g., MacKay and Ahmetzanov, 2005; Jay, Caldwell-Harris and King, 2008), they tend to be represented, processed, and recalled better than neutral words in the mental lexicon (e.g., Dewaele, 2004). The strong tendency of negative emotional words to produce more interference than do positive emotional words appears across tasks (e.g., Wentura et al., 2000 on a lexical decision task; Algom et al., 2004 on a word naming task; Eilola et al., 2007 on an emotional Stroop task).

The underlying mechanisms are still under investigation; one plausible account for the emotional interference produced by negative words is that threatening stimuli prevent the disengagement of attention (see Fox, Russo, Bowles, & Dutton, 2001, as cited in Eilola et al., 2007). This is because the negative words capture attention resources in the manner of *an automatic vigilance hypothesis* (e.g., Ohman & Mineka, 2001). The automatic vigilance (i.e., sustained attention) hypothesis states that humans preferentially attend to negative stimuli, and negative words automatically trigger vigilance than do neutral or positive words. Such interference by negative emotion words is a well-established finding in L1 processing (Estes & Adelman, 2006). [2]

Given that emotional words automatically elicit corresponding valence and arousal when the word meaning is represented in L1, the comparison of emotional involvement between L1 and L2 can provide more direct clues regarding how automatically semantic representation is activated during L2 processing as compared to that in L1 processing than what the previous literature have done through multiple lexical routes. If L2 speakers show a similar degree of emotional involvement in L2 to that in L1, it means that the L2 semantic activation is as strong as in L1. If not, L2 semantic activation may be weaker than in L1. In the following section studies focusing on how the processing of emotional words differs in L1 and L2 will be reviewed.

2.2.2. The L1–L2 Difference in Processing Emotional Words

While much research about emotional word processing has been conducted in the monolingual domain (from patients for clinical purposes to a normal monolingual

[2] For this reason, the present study focused on processing negative emotion words in order to apply the consistent effects by negative words in L1 to L2 processing.

population), only a handful of studies have drawn attention about the bilingualism (for review see Harris et al., 2006 and Pavlenko, 2008). Among them, most qualitative research has shown that L1 has a more intense emotional effect than L2 does (Pavlenko, 2005; Harris et al., 2003; Dewaele, 2004; Jay, Caldwell-Harris and King, 2008). In these studies, L2 speakers reported to prefer using L1 to L2 when describing their feelings in detail, for instance, in writing an essay or in therapy because they felt stronger emotion in L1 than in L2. Likewise, the L2 speakers tend to maintain a certain emotional distance in L2 better than in L1 so that they did not hesitate to talk about an embarrassing moment in L2 compared to in L1 (Harris et al., 2003; Pavlenko, 2005; Wu & Thierry, 2012).

However, quantitative studies have administrated diverse tasks to compare the degrees of emotional involvement in L1 and L2 processing and revealed more complicating findings. One of the common methods is a surprise recall task. Participants performed other tasks on emotional and neutral words in both languages such as emotional-intensity rating (i.e., from zero=no feeling to seven= strong feeling); then they did a subsequent unexpected recall task. The rationale of this task is that if the participants feel emotional words more intensively in one language than in the other, emotional words in that language will be recalled better implying stronger semantic activation in that language. For instance, in Anooshian & Hertel's study (1994) using a recall task, L1 English-L2 Spanish and L1 Spanish-L2 English speakers who were fluent in both languages recalled emotional words better than neutral words, only in their L1s, but not in L2s.

Also, some studies attempted to examine automatic emotional intensity during language processing through detecting a change in the ability of the skin to conduct electricity. If L2 speakers respond to emotional words in L1 more than in L2, it would mean stronger semantic connections in L1 than in L2. For instance, Harris et al. (2003) measured skin conductance responses (SCRs) of L1 Turkish-L2 English speakers while listening to emotional words and phrases in L1 and L2. They found that emotional expressions heard in Ll elicited larger skin conductance amplitudes than comparable expressions in the L2 in this population. Considering that the participants in these two studies were proficient L2 speakers (balanced bilinguals), the findings underlie that explicitly knowing the emotional word meaning in L2 does not necessarily mean feeling it as intensively as in L1.

On the contrary, other quantitative studies using the same paradigm of emotion-memory effects also found the emotional word effect in L2 similar to or more than in L1. For example, regarding the emotion-memory effects, Ayçiçegi-Dinn, A., & Caldwell-Harris (2009) examined highly proficient L1 Turkish-L2 English speakers on a recall task and reported their superior memory of emotional words to neutral words both in L1 and L2 with a similar magnitude of differences, implying a comparable establishment of semantic structure in the two languages among advanced L2 speakers. Considering the participants' language profile, it is possible that such a high development of the semantic system in L2 could be assisted by their five years of residency in L2-speaking countries.

On top of that, Ayçiçegi & Harris (2004) observed stronger emotion-memory effects in L2 than in L1 by L1 Turkish-late L2 English bilinguals (the mean age of acquisition was later than age 12). In this study, participants performed a word rating task and then were asked to write down and to check words that appeared in the previous rating task as many as they remembered, respectively. As a result, they remembered and recognized emotional words in L2 better and more frequently than in L1. However, unlike previous studies showing a memory advantage of emotional words more in L1 than L2, this study employed unblocked presentation of stimuli. As the authors also pointed out, in a mixed-language condition, "the unexpectedness and novelty of the English (L2) items could have facilitated elaborative processing of these items" (p. 11) and consequently better memorized than familiar and less surprising L1 items.

Finally, on one hand some studies have shown inconsistent findings regarding the L1-L2 differences in emotional word processing across tasks; on the other hand, other studies have done so within the same task mainly because they did not control the confounding factors. Specifically, among a few of studies using emotional Stroop task for L1–L2 difference in emotional word processing, Eilola et al. (2007) and Eilola and Havelka (2010) found no L1-L2 difference in emotional interference effects while Sutton et al. (2007) and Winskel (2013) found more emotional interference effect in their dominant languages than in their weaker languages.

One main reason for the conflicting findings among these studies may be that participants' language profile such as L2 proficiency and L2 use are not comparable. For example, slower responses elicited by negative words were significant in L1 but not in L2 in Winskel's study (2013) while such patterns appeared in a reverse way (significant interference in L2 not in L1) in the Sutton et al.' (2007) study. However, in the two studies showing regarding L1-L2 differences, participants' dominant languages differed. Given that participants in Sutton et al. (2007) were early L1 Spanish- L2 English speakers whose age of L2 acquisition was prior to 7 and who had lived in L2-speaking country, their dominant language was L2. Therefore, their findings (more emotional effects in L2 than in L1) are actually consistent to those of Winskel's (2013) that showed more emotional involvement in their dominant language than weaker language.

On the other hand, in Eilola and her colleagues' studies (2007), the emotion word effects were comparable between two languages. The L1 Finnish-L2 English speakers and L1 Greek-L2 English speakers in their studies, although their L2 proficiency self-rating was not as high as their L1 proficiency, can be considered highly proficient in L2 English given that they were immersed in L2-speaking environment or using L2 in everyday life as frequently as L1: more than half of the L1 Finnish-L2 English speakers reported to listen to music (88.2%) and watch TV programs (64.7%) in L2 everyday; the L1 Greek-L2 English speakers were studying in UK. Therefore, it is plausible that the L2 speakers in these studies, regardless of whether they were exposed to L2 early or late in life, might have used L2 extensively so that they were able to automatically access to and activate emotions through emotional words even when they were presented in L2.

2.2.3. Implications of the Emotion Word Processing for Examination of Semantic Activation

Although most qualitative studies reported that L2 speakers tend to feel emotional words in L1 more strongly than in L2 (e.g., Harris et al., 2003; Dewaele, 2004), some quantitative studies showed contrary findings (e.g., a higher emotionality rating on L2 words than L1 words). One of the reasons for showing discrepancy between qualitative and quantitative studies may be the task that each qualitative study employed. The depth of processing in which a task is performed can determine the degree of emotional word effect. That is, if a task allows deeper processing in one language than the other, more emotional effects will be produced in the more deeply processed language than the less deeply processed language. For instance, Ayçiçegi–Dinn, & Caldwell-Harris (2009) tested L1 Turkish–L2 English speakers living in Turkey on a surprise recall task following four types of tasks. These tasks were an emotional-intensity rating task, a letter-counting task, a translation task, and a word association task, varying in terms of the depth of the demanded mental processing. Among them, the emotional-intensity rating task showed stronger effects for emotional words in L1 than in L2 while the translation task, which required additional depth of processing that accompanies translation, revealed stronger effects in L2 than in L1. That is, the L2 advantage over L1 occurred as a byproduct of a deeper processing in the L2 induced by task demands. Regarding the task–specific effects on the L1-L2 differences in emotional word processing, the authors mentioned that "because L2 words are more novel or amusing than L1 words, sometimes processing in L2 leads to deeper processing and thereby to enhanced recall more than in L1." Thus, the task itself should require equivalent depth of processing in both languages in order to compare the automatic activation of emotional connotation (Ayçiçegi-Dinn, & Caldwell-Harris, 2009, p. 293).

Moreover, for the present study, it is important to make sure that a given task leads participants to automatic processing rather than conscious control of stimuli. If a task makes participants use their explicit knowledge such as comparing and rating the meaning of emotional words without a time constraint (e.g., a rating task which asks people to judge emotional word valence 1 to 7), the result cannot address automatic activation of semantic representation for the present study. There have been many tasks used to examine bilinguals' processing of emotional words in their two languages: a questionnaire (Dewaele, 2006), a recall task (Anooshian & Hertel, 1994), a memory task (Ferre et al., 2010), a rating task (Ayçiçegi & Harris, 2004), a word association task (Ayçiçegi-Dinn, & Caldwell-Harris, 2009), SCRs (Harris et al., 2006), an implicit affect association task (Segalowitz et al., 2008), etc. However, these tasks are more likely to measure explicit knowledge or different processing rather than implicit knowledge or automatic processing: whether L2 speakers have knowledge of emotional words (a word association task), how they feel them consciously (a rating task), and how strongly they become pulsated (SCRs).

In this sense, an emotional Stroop task, which requires equivalent depth of automatic processing in both languages (Eilola et al., 2007; Jiang, 2012; Sutton et al., 2007), was used in the present study. In the emotional Stroop task, each of neutral and emotionally charged word is presented to the participants, and then they are asked to identify the print color of each word, regardless of its meaning, as fast and accurately as possible. In this experiment participants should ignore the meanings of the words and focus on responding to the print color. Thus, "any effect of the words on the response times in the emotional Stroop paradigm is likely to reflect automatic and early lexical processing" (Eilola et al., 2007, p. 1072).

Besides, the emotional Stroop effect is a well-established finding in L1 research. Previous studies using the Stroop paradigm have consistently shown that L1 speakers have the capacity to detect differences between words that imply threat (negatively emotional words) and those that do not (neutral words). Larsen et al. (2008) explained this mechanism in terms of tasks: “The underlying mechanism for the emotion Stroop is thought to be a generic interrupt system that acts early and in an automatic fashion when threatening information is detected in the perceptual stream (Algom et al., 2004)” (p. 445).

In sum, the present study examined automatic activation of the emotional involvement in L2 in order to investigate more directly automatic semantic activation in L2 processing. This paradigm is based on the consistent findings that emotional involvement occurs when rich connotations of emotional words are established in a given language. For the emotional interference effect, an emotional Stroop task was administered in that such effect can be observed at the equivalent depth of processing in both languages as well as in the automatic processing. Also, other variables such as L2 proficiency and L2 use were manipulated to investigate how the level of automatic semantic activation in L2 changes according to these factors.

Given that emotional Stroop task zooms in on semantic activation during word processing, another approach to tap automatic semantic activation during sentence processing will be needed. Besides, it is important to note that physical attributes as well as affective properties of a word are central aspects when learning semantic representation (Andrew et al., 2009). That being said, since emotional involvement tends to focus on semantic activation based on affective aspects of words (i.e., linguistic form–to–emotional mental state mapping), a different approach focusing on physical properties in sentence processing will provide another piece of the puzzle for examining semantic activation from the other angle. In the following chapter, it will be discussed how important the investigation of mental imagery during L2 sentence reading is for examining semantic activation more directly.

2.3. Examining Semantic Activation through Mental Imagery

Given the need for an innovative approach that focuses on how participants link linguistic input to its real-world referent, instead of their lexical equivalent, examining emotional involvement during word recognition can provide more direct evidence regarding semantic activation.

With regard to the need to look into automatic semantic activation at sentence processing from the new perspective, the second approach focuses on how participants generate mental imagery according to sentence meaning automatically. This is because mental imagery is produced during language processing when linguistic inputs are automatically linked to real–world referents (depicted objects).

In order to assess mental imagery, a sentence–based picture recognition task can be a useful tool. In this task participants are asked to judge whether a picture has been mentioned in the preceding sentence. The picture will show an object in exactly the same way or in a slightly different way (e.g., different orientation or shape) than depicted in the previous sentence. If the sentence meaning primes the picture recognition, a *yes* response will be faster when an image is congruent to what is implied in a sentence than when it is incongruent. In this task, the sentence–picture congruency effect in the form of faster responses to congruent items than to incongruent can occur when sentential meaning is activated automatically enough to trigger an imagery of the depicted object.

This approach was inspired by a set of studies reported by Zwaan and colleagues (Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002). These studies were intended to explore and test the theory that meaning is represented as a form of mental imagery rather than propositions. The following section will explain the theoretical rationale for this approach.

2.3.1. Theoretical Issues on How Meaning is Represented during Language Comprehension: Embodied Cognition

There have been two major approaches in an on-going debate of how meaning is represented, depending on whether meaning representation is based on linguistic input only or on interaction with cognitive systems (e.g., sensorimotor system). The traditional view or *the propositional theory* was established based on the former; on the other hand, the new embodied cognition approach was originated from the latter.

The mental imagery in the embodied cognition approach takes into account that sensorimotor information may play a crucial role in semantic representation so that meaning is represented in a form of imagery. Sensorimotor information indicates what we experience in the real world and what is in the sensory system and the motor system. This information is automatically recruited during language comprehension, resulting in mental imagery. Thus, when people cannot generate mental imagery relevant for the reference in a sentence (e.g., a sentence conveying completely new or highly abstract concepts), their comprehension of the sentence is relatively weak compared to that of a sentence that easily activates mental imagery (Stanfield & Zwaan, 2001). For example,

> "Suppose that one is talking about a *double lutz* to another who has no idea what a double lutz is and cannot remember ever hearing of it. With the verbal explanation and related background knowledge, the listener can figure out that it is a jump that an ice-skater performs. However, the listener cannot "comprehend it in the same capacity that the speaker can" (Taylor & Zwaan, 2009, p. 52).

Therefore, complete semantic activation may depend on the availability of sensorimotor information during language processing according to the embodied cognition.

In the meantime, some argue that mental imagery may not be required in semantic activation all the time (e.g., Rogers et al., 2004) since not all concepts have specific images, and most knowledge that we learn from reading books is not learned from experience. For these reasons, it is often questioned how all linguistic input, including abstract concepts, can be represented in the perceptual symbol system. Regarding this issue, even among the embodied theorists, there exist various versions of embodied theories about the role of sensorimotor involvement from weak to strong on the continuum. Since the purpose of this study is not to argue for embodied cognition, a brief background of embodied cognition on which the mental imagery is based will be discussed below.

Figure 1. Schematic of Theories' Position along the Continuum from Amodal to Modal [3]

	Traditional view	Weak embodiment		Strong embodiment
	←			→
Relationship to sensory-motor systems	Complete independence (e.g., Collins & Loftus, 1975)	Independent but associated (e.g., Rogers et al., 2004)	Partial dependence (e.g., Vigilocco et al., 2004)	Complete dependence (e.g., Barsalou, 1999; Zwaan, 2004)

As shown in Figure 1, theories regarding the relationship between the sensorimotor system and semantic representation are divided into four broad groups. The leftmost are amodal theories which are based on the traditional view, arguing semantic representation completely independent from sensory–motor systems; weak embodiment theories insist that meaning representation during language processing is mediated or partially modulated by sensory-motor systems; and strong embodiment theories claim that semantic representation is completely dependent on and directly interacts with sensory-motor systems (Meteyard & Vigliocco, 2008, p. 296).

To sum up the consensus among diverse embodied cognition, all versions of embodied cognition admit sensorimotor's involvement in language comprehension at least to some extent and assume that perceptual symbols which automatically come to our mind during language comprehension are modal. For instance, after reading a sentence saying that *John put a pencil in a cup*, people recognized the picture of a vertical pencil faster and more accurately than a horizontal pencil, while after reading a sentence saying that *John put a pencil in a drawer*, they recognized a horizontal pencil faster and more accurately than a vertical pencil (Stanfield & Zwaan, 2001). According to the embodied cognition

[3] Adapted from Meteyard & Vigliocco (2008)

theory, this is because people automatically recruited sensory-motor information during sentential reading, and the mental imagery produced by the sensory-motor system can be concrete in terms of orientation.

Consequently, this view challenges the classic perspective regarding how meaning is represented in the mind. Traditionally, it has been believed that linguistic input was converted to propositional representations (e.g., Kintsch, 1998, as cited in Zwaan et al. 2002). Based on the relationship between agent, object, and action, each of the sample sentences *John put a pencil in a cup* and *John put a pencil in a drawer* would be represented as [[PUT[JOHN, PENCIL]], [IN[PENCIL,CUP]]] and [[PUT[JOHN,PENCIL]], [IN[PENCIL,DRAWER]]], respectively. According to the traditional perspective, the connection between propositions, but not sensory modality, was important for sentence processing, and the image activated in the mind was expected to be amodal. In short, the embodied image of the pencils should be the same for each sentence. Therefore, sentence-picture congruency effects cannot be explained by the traditional perspective, whereas the embodied cognition can account for these (Zwaan et al., 2002).

To summarize, although it is still an on-going debate whether mental representation resulting from sensory-motor activation is always essential in semantic processing, there are consistent findings showing the sensorymotor involvement in language comprehension, especially when action verbs or concrete nouns constitute the core meaning of the linguistic input. The following chapters will introduce the findings to support this view and explain how this approach can help us examine automaticity in semantic representation more directly than previous methodologies.

2.3.2. Empirical Evidence of the Embodiment Theory

Recently, more and more research has reported empirical evidence based on the embodiment theory. First, Zwaan and colleagues' studies (Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002), which inspired the present study, were intended to test the theory that meaning is represented in mental imagery rather than propositions. In Stanfield and Zwaan (2001), 40 L1 English speakers were tested on a picture-recognition task. The participants responded to a picture following a sentence. If an object in a picture had been mentioned in the previous sentence, they responded *yes*; if not, they responded *no*. The 24 experimental items requiring *yes* responses were all mentioned in the previous sentences; however, half of the items differed in terms of the orientation (either vertical or horizontal) implied in the sentence context. This manipulation resulted in two conditions: match and mismatch. For a match item, the sentence stated that *John put a pencil in a cup* and the picture showed a vertical pencil. For a mismatch item, the sentence stated that *John put a pencil in a drawer* but the picture showed the vertical pencil. This was a mismatch because the orientation of the pencil differed from what was expected in the sentence. Performance was significantly faster in the match condition than the mismatch. These results suggest that when L1 speakers comprehend a sentence in their native languages, they generate mental imagery corresponding to the context of a sentence in their mind.

Numerous following studies suggest that such mental imagery may be automatically activated during language comprehension (e.g., Zwaan et al., 2002; Pecher et al., 2009). Zwaan et al. (2002) added a picture-naming task which is less likely to evoke a comparison between a sentence and a picture than a picture recognition task. The participants' picture naming was significantly faster and more exact when the picture appeared in the same way as implied in the previous sentence than when the picture did not. Moreover, in Pecher et al.'s study (2009), the pictures were presented 45 minutes after the sentences were read. They still found sentence-picture congruency effects even when the tasks did not require an overt comparison between a sentence and a picture. This sentence-picture congruency effect suggests that L1 speakers are able to automatically activate the imagery corresponding to the meaning of input during sentential processing.

Such sentence-picture congruency effects have been considered strong evidence for the embodiment theory arguing that meaning is represented as an image in our mind during language comprehension (for more detail, refer to the Perceptual Symbol Theory in Barsalou, 1999). Even when leveling it down to the weak version of the embodiment theory, many studies with behavioral (e.g., Taylor & Zwaan, 2009; Pecher et al., 1998; Zwaan & Yaxley, 2003) as well as neuroscientific evidence (e.g., Coppens, Gootjes, and Zwaan, 2012) have supported the theory that sensory-motor information may be automatically involved in the semantic representation of, at least, linguistic input referring to action verbs and concrete objects, and thereby, mental imagery routinely occurs during comprehension of the depicted motions or objects. For instance, Zwaan and Taylor (2006) examined L1 English speakers' reading time during sentential reading implying manual rotation (e.g., *Before the big race the driver took out his key and started the car*) with turning a knob either clockwise or counterclockwise to proceed through the sentences in a self-paced reading task. Their reading time at the verb region was faster (i.e., their turning a knob was smoother) when the direction implied in the sentence and the direction of their manual rotation were congruent than when they were not. With four extended experiments, this study suggests that “motor system assists in or is required for the comprehension of language about action (Taylor & Zwaan, 2009, p. 49).”

In addition, many brain-imaging studies consistently showed that the cortical areas that are involved with body actions are automatically activated for processing language referring to actions in a normal population (e.g.,Vigliocco et al., 2006). In the same vein, Coppens, Gootjes, and Zwaan (2012) supported the mental representation during language comprehension through ERPs. In this study, the influence of prior visual experience on subsequent reading was assessed. Participants saw a picture of an object and read a text about the object implying the same or a different shape as shown in the previous picture. In spite of a 15-minute gap between the picture and the sentence, when the shapes in the picture and the sentence meaning mismatched, ERPs during reading showed larger N400 amplitude than when matched.

These results discussed above strengthen the case for the interaction between language and visual experience during language comprehension, suggesting that complete semantic activation produces mental imagery.

2.3.3. Implications of the Mental Imagery for Examination of Automatic Semantic Activation

Putting aside a question of whether the amodal approach or the embodiment approach is more persuasive, it is worth paying attention to the consistent empirical evidence which shows that meaning representation automatically involves sensory-motor information during language processing. The reliable findings indicate that meaning representation can trigger certain "embodied (visual or motor) information" (Taylor & Zwaan, 2009), and that L1 processing is highly efficient, since certain images are activated in the same way as described in the sentences. Given that mental imagery is automatically generated as a consequence of semantic activation during sentential reading in L1, the investigation of mental imagery during L2 processing can provide more direct clues regarding the automatic activation of semantic representation in L2 processing.

Considering the need for a methodology that disambiguates semantic processing from lexical processing, the sentence-based picture recognition task assessing the sentence-picture congruency effect may be a promising alternative. If L2 speakers are tested with the sentence-based picture-recognition task, their L2 semantic processing can be compared to that of L1 speakers'. The results will provide more direct clues regarding to what extent meaning representation can be automatically activated in L2 processing.

To this end, this present study aimed to apply the Stanfield & Zwaan's (2001) task to L2 speakers along with L1 speakers. If L2 speakers show the same patterns as L1 speakers, then one might conclude that L2 processing activates semantic representation to an extent similar to L1 processing. If not, one might conclude that automatic semantic activation in L2 speakers is not strong enough to generate mental imagery during L2. Therefore, a comparison of L1 and L2 speaker performance on the sentence-based picture-recognition task is expected to provide more direct evidence on the relative levels of semantic activation in L1 and L2 processing.

2.4. Motivation of the Present Study

The extent to which semantic representation is activated automatically in L2 processing is a fundamental issue in the efficiency of L2 comprehension, but emphasis on this aspect has been scant. The few previous studies on this issue have limited ability to answer this question for two reasons. First, after they entangled semantic and lexical representation, the authors tried to infer how strong the semantic activation would be from the results of the mixed–level processing (e.g., Finkbeiner & Nicol, 2003). Second, most of the previous studies did not pay enough attention to crucial factors for developing semantic activation; that is, L2 proficiency and L2 use. Although L2 proficiency and use greatly influence L2 semantic structure development (e.g., Kroll & de Groot, 1997) and automaticity in L2 processing (DeKeyser, 2001), respectively, previous literature neither scrutinized nor manipulated these factors. As a result, they have produced divergent findings.

Furthermore, whereas some studies found weaker semantic activation in L2 than in L1 mostly in sentence processing (e.g., Hu & Jiang, 2011), others found symmetric semantic activation between two languages mainly in word processing (e.g., Duyck & DeHouwer, 2008). Such contrary results render unclear the issue of whether automatic activation of semantic representation in L2 can develop to a native-like level as L2 proficiency and L2 use increase at lexical and sentential levels. These limitations call for a more direct comparison of semantic activation between two languages at both levels of processing.

Given the need for a new approach to allow a more direct investigation of semantic activation, the present study administrated two innovative paradigms for two different processing levels. First, in order to examine automatic semantic activation more directly at the level of word processing, the present study examined emotional involvement which appears in L2 speakers during L2 word reading and compared to that in L1 speakers. This approach is based on the consistent findings that emotional words automatically activate an emotionally charged state of mind (e.g., Sutton et al., 2007). Previous studies that have used the emotional Stroop tasks revealed that participants' judging or naming a color of a word tend to be slower when the words are emotional words, especially negative words (e.g., *panic*, *depressed*), than neutral words (e.g., *floor*, *driver*) in their stronger language more than in their weaker language. If advanced L2 Korean speakers in the present study do not show emotional involvement in L2, this would mean that their emotional connotations of Korean words in L2 speakers' mind are not activated as automatically as in L1 speakers'.

Second, as another approach to more directly examine semantic activation at the level of sentence processing, mental imagery which appears in L2 speakers during L2 sentence reading was examined and compared to that in L1 speakers in the present study. This is based on the consistent findings that semantic representation can take the form of mental imagery. Previous research on mental imagery has shown that L1 speakers can involve the sensory-motor system automatically and generate mental imagery according to the content of a sentence during a sentential reading in L1. For example, in Zwaan and his colleagues' successive studies (e.g., Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002; Zwaan & Yaxley, 2003; Taylor & Zwaan, 2009; Coppens, Gootjes, & Zwaan, 2012). L1 speakers recognized an object significantly faster when an object orientation or shape matched what the previous sentence implied than when they did not match. The sentence–picture congruency effect, in the form of a faster response time to congruent items than incongruent items, is considered evidence of automatic semantic activation.

Given that mental imagery can be observed only when semantic information is processed, examining mental imagery allows the present study to disentangle semantic processing from non–semantic processing (e.g., lexical representation and episodic representation). Moreover, given that previous research has consistently found that L1 speakers routinely generate mental imagery, comparing mental imagery between L1 and L2 processing can show more directly whether L2 processing activates semantic representation as automatically as does L1 processing. In the present study, the advanced L2 Korean speakers were compared with L1 Korean speakers. If participants do not show sentence-picture congruency effects during sentential reading in L2, this means that the meaning representation activation in L2 is not as automatic as in L1.

Furthermore, as discussed, in previous studies on semantic activation, L2 proficiency and L2 use were confounded with the degree of automatic semantic activation (e.g., emotional involvement). Therefore, in the present study, potentially confounding factors were either controlled or manipulated. That is, L2 proficiency was controlled by limiting L2 speakers' L2 proficiency to *advanced* and then manipulated according to their L2 proficiency scores within the category of advanced L2 speakers. Also, participants' amount of L2 use was calculated by a thorough language background questionnaire. These two variables were included as factors in the analyses. This manipulation allowed a teasing out of L2 use effects from the general L2 proficiency effects.[4] In this way, the present study is expected to shed more light on how automatically L2 word and sentence can activate semantic representation and how it changes as L2 proficiency and the amount of L2 use increase among advanced L2 speakers.

2.4.1. Research Questions

The present study pursued to address the following research questions:

1. Do advanced L2 Korean speakers exhibit emotional involvement during word processing in L2 comparably with L1 Korean speakers?
2. How does emotional involvement during L2 word processing change according to L2 proficiency and amount of use?
3. Do advanced L2 Korean speakers generate mental imagery during a sentential reading in L2 as automatically as L1 speakers do?
4. How does the ability to automatically generate mental imagery during a sentential reading in L2 change according to L2 proficiency and amount of use?

To address these research questions, an emotional Stroop task and a sentence-based picture recognition task were administered to measure emotional involvement and mental imagery, respectively. Advanced L2 Korean speakers performed these two tasks in their L2 (Korean) along with L1 Korean speakers. In the following chapters, participants, methodologies, and results of each task are described in detail.

---

[4] Considering that the focus of the present study is *automatic* L2 processing, the factor that may contribute to automaticity should be included. In the sense that "L2 use" in the present study indicates how many hours L2 speakers might have been using L2 in an automatic manner (e.g., personal interaction in L2, media use in L2, and so on), this variable is expected to explain some parts of the L2 speakers' automatic semantic activation beyond what L2 proficiency does. L2 proficiency will be measured through an off-line task and thereby may allow the use of strategic rather than automatic knowledge in the present study.

Chapter 3. Pilot Studies

3.1. Overview

Two pilot studies were conducted for two separate reasons. The purpose of the first pilot study was to check L1–L2 differences in emotional involvement and mental imagery before applying these methods directly to L2 Korean speakers. This study was undertaken because few studies have used these methods in L2 processing (e.g., only a few published studies have used emotional Stroop effects in L2 processing, and no known research has used mental imagery in L2 processing). For convenient data collection, English was established as a target language so that L1 English speakers and L2 English speakers could participate in this pilot study.

Also, the aim of the second pilot study is to confirm that the Korean stimuli created for the present study produce condition effects (i.e., emotional Stroop effect and sentence-picture congruency effect) as intended among L1 speakers as a control group.

3.2. Methods

In the emotional involvement experiment in English (Ahn, 2013b), a total of 57 L2 English speakers at two different L2 proficiency levels (advanced and intermediate; based on self–ratings) and 44 L1 English speakers participated at the University of Connecticut (Storrs) from October 15, 2013 to November 1, 2013.

The experimental task and procedure were basically the same as described in the *Methods* section of the present study. Briefly, respondents participated in an emotional

Stroop task in which they were asked to judge the color of a presented word by pressing a button (either the right shift key for blue or the left shift key for green). A set of 20 negative words (e.g., lonely, war) and a set of 20 neutral words in English (e.g., table, driving) were presented after a fixation mark (*) once in each of two colors in a separate block. The order of the blocks was counterbalanced across participants so that half of the participants saw negative words first then neutral words, whereas the other half saw the words in the reverse order. The word frequency and word length were equivalent across word conditions (ps > 0.10). It took less than 10 minutes to complete an emotional Stroop Task (details regarding this method are described in section 4.3 in the present paper).

Furthermore, in a mental imagery experiment presented in English (Ahn, Jiang, & Osthus, 2012), a total of 38 L2 English speakers with two different L2 proficiency levels (advanced and intermediate) and 18 L1 English speakers participated in a study conducted at the University of Michigan (Ann Arbor) from January 9, 2012 to February, 20, 2012. The experimental task and design were the same as described in the *Methods* section of the present study. Briefly, in a sentence–based picture recognition task used for the pilot study, participants read a sentence, saw a picture, and then judged whether the object in a picture had been mentioned in the previous sentence, by pressing *yes* (right shift key) or *no* (left shift key). All critical pictures required *yes* responses by showing objects as mentioned in the previous sentences; however, only half matched the sentence in terms of orientation and shape, whereas the other half did not. For example, a picture of a bird with folded wings followed either a sentence saying that *the child saw the bird on a wire* (match condition) or a sentence stating that *the child saw the bird in the sky* (mismatch condition). For the critical items, a total of 64 sentences were paired with 32 corresponding pictures, producing 32 pairs of each of match and mismatch condition. This combination resulted in two lists, each of which consisted of different versions of 16 pairs for each condition. The list order was counterbalanced across participants. The materials were pretested with 10 native English speakers to confirm that each picture was matched with only one of the two sentences, but not likely vice versa. It took about 30 minutes to finish the experiment (more information regarding this method is described in section 5.3 in the present paper).

3.3. Results and Discussion

First, the response latency in the emotional condition was significant in the advanced L2 group, similar to the L1 group, but not significant in the intermediate L2 group (Table 1). The two L2 groups' different patterns in the emotional involvement effects in the form of delayed response in the emotional condition suggest that L2 speakers can activate emotional word meanings in L2 as automatically as L1 speakers based on their L2 proficiency levels: the more proficient L2 speakers are in L2, the more likely they are to activate semantic representation automatically during L2 word processing. This finding of the pilot study with L2 English speakers on emotional involvement is consistent to the prediction of the present study with L2 Korean speakers.

Second, the results of the mental imagery experiment showed that the difference in the response latency was significant in the advanced L2 group, similar to the L1 group, but not significant in the intermediate L2 group (Table 2). The two L2 groups' different patterns in the sentence–picture congruency effects, in the form of delayed response in the mismatch condition, suggest that L2 speakers can produce mental imagery during L2 sentential reading as automatically as L1 speakers based on their L2 proficiency levels: the more proficient L2 speakers are in L2, the more likely they are to activate semantic representation automatically during L2 sentential reading. This finding of the pilot study with L2 English speakers on mental imagery is consistent to the prediction of the present study L2 Korean speakers.

Table 1. Results of Piloting Emotional Involvement in L2 (English) Processing

		RT
L1	Emotional	493.73 (126.95)
(n=44)	Neutral	473.49 (85.71)
	Difference	20.24*
Advanced L2	Emotional	511.25 (71.05)
(n=24)	Neutral	473.03 (58.25)
	Difference	38.22*
Intermediate L2	Emotional	515.99 (152.40)
(n=33)	Neutral	519.92 (123.77)
	Difference	-3.93

(*SD* in parentheses, $p < .05$)

Table 2. Results of Piloting Mental Imagery in L2 (English) Processing

		RT
L1	Match	678.07 (147.56)
(n=18)	Mismatch	785.58 (193.39)
	Difference	107.51*
Advanced L2	Match	689.12 (196.07)
(n=17)	Mismatch	855.45 (267.89)
	Difference	166.33*
Intermediate L2	Match	807.04 (225.27)
(n=21)	Mismatch	861.99 (192.49)
	Difference	54.95

(*SD* in parentheses, $p < .05$)

Given that the pilot studies revealed L1–L2 differences in emotional involvement and mental imagery during L2 processing according to L2 proficiency, a second pilot study was conducted to confirm that such emotional involvement and mental imagery are produced with Korean stimuli as well. To this end, the major modifications from the first pilot study are the language (Korean instead of English, see section 5.3.2 for the procedure to select and confirm Korean stimuli) and participants (L1 Korean speakers instead of L1 English speakers as a control group).

Therefore, a total of 34 L1 Korean speakers (19 males and 15 females) were recruited in the Connecticut area through flyers and personal contacts from October 15, 2013 to January 20, 2014. They participated in emotional involvement and mental imagery experiments in Korean and showed similar patterns to those of L1 English speakers when using English stimuli. The results showed significant delays in responding to the emotional condition and the mismatched condition in the emotional involvement experiment and the mental imagery experiment, respectively (ps<.05) as shown in Table 3. The two experiments produced high reliabilities (rs>.78) by L1 Korean speakers.

Table 3. Results of Piloting with Korean Stimuli

Tasks	Conditions	RT
Emotional Stroop task	Emotional	555 (201)
(n=34)	Neutral	531 (199)
	Difference	24*
Sentence-based picture recognition task	Match	745 (280)
(n=34)	Mismatch	792 (291)
	Difference	47*

(*SD* in parentheses, $p < .05$)

Chapter 4. Emotional Involvement in L2 Processing (Experiment 1)

4.1. Study Overview

To examine semantic activation during L2 processing more directly, two experiments were conducted in the present study. In this chapter, we discuss the first experiment. The paradigm of Experiment 1 is to look into emotional involvement[5] during L2 word recognition. The rational for this paradigm is that emotional words, especially negative words (e.g., depressed), automatically command attention more than neutral words (e.g., baggage). This sustained attention interferes with subsequent cognitive activity such as button pressing (Larsen et al., 2006). An emotional Stroop task (EST) was used to assess emotional involvement during L2 reading. In the EST, participants were instructed to press a blue–colored button (the right shift key on the keyboard) with their right index finger if the word appeared in blue and the green–colored button (the left shift key) with their left index finger if the word appeared in green. They were told to respond as quickly and accurately as possible.

Color judgment of L1 speakers becomes more delayed when emotional words appeared than when neutral words appeared, and this tendency consistently persists in previous studies (e.g., McKenna & Sharma, 2004). The emotional Stroop effect in the form of a slower response to emotional words than to neutral words can be considered a result of automatic and complete meaning activation in recognizing words. Thus, emotional Stroop

[5] This emotional word effect is often called *emotional intrusion*, *emotional interference*, or *emotional Stroop* (McKenna & Sharma, 2004), because such an effect is not intended and disrupts processing relevant stimuli. In the present paper, the term *emotional involvement* is used to eliminate the negative nuance in the context of semantic activation.

effects (or *emotional involvement*) are observed only when participants automatically process the emotional word's meaning. If the word's meaning remains inactivated or incomplete when L2 speakers respond to the printed color of the word, their color judgment would be similar regardless of the emotional or neutral condition. By comparing the L2 respondents' ability to generate emotional involvement to the same ability of the L1 speakers, the present study sought to address the issue of how semantic representation can be activated automatically in L2 word processing.

Among various versions of emotional Stoop tasks, depending on the ways to respond to colors such as a color naming task (e.g., Strauss et al., 2005), a word naming task (e.g., Algom et al., 2004), and a lexical decision task (e.g., Wentura et al., 2000), a button–pressing or color decision task (e.g., Eilola et al., 2007) was selected for two practical reasons. First, it is used to exclude potentially confounding factors such as participants' production skills (in the case of naming tasks) or explicit knowledge of L2 lexicons (in the case of a lexical decision task). Second, it is used to reduce additional time and potential errors in analyzing the data (i.e., listening to all recordings of participants' vocal responses and manually coding each of them as correct or incorrect).

During the experiment, the emotional and neutral words were presented in separate blocks, because emotional involvement (or *emotional interference*) consists of fast and slow components (McKenna & Sharma, 2004). The fast component refers to the interference within a trial, whereas the slow component implies a lingering effect on the words after the emotionally charged word is presented. To maximize emotional involvement, this study presented stimuli in blocked conditions. This manipulation of separate blocks is also typical in the bilingual domain to prevent the confounding lingering effects caused by emotional words on neutral words in a mixed-block design (e.g., Sutton et al., 2007). The order of word blocks was counterbalanced across participants. Thus, half of the participants viewed the emotional word block first then the neutral word block, whereas half did the task in the reverse order. Within a block, each word appeared once in blue and once in green on a white background in random order. Including two colors in the present study rather than many (e.g., Eilola & Havelka, 2010 used four colors), was an attempt to minimize participants' burdens to memorize which button is assigned to which color; otherwise, possible confusion in matching colors and buttons may affect the task execution.[6] To create and run the experiment, DMDX[7] was used.

Furthermore, L2 speakers' emotional involvement during L2 word processing was compared along with their L2 proficiency and the extent to which the participants use their L2 in daily life other than in a classroom setting. The results shed more light on the developmental aspects of automatic semantic processing; that is, how automatic semantic activation during L2 processing changes according to L2 proficiency and L2 use. The details of the Experiment 1 are described below.

4.2. Participants

For the experimental group, a total of 94 L2 Korean (non–heritage) speakers, who were allegedly assessed as advanced L2 Korean learners by their institutions or self–reports,

[6] Considering the L2 speakers' response in Eilola et al. (2007) was much slower (mean 717 ms) than those in the present study (mean 500 ms), including more colors was probably done because there were four possible manual responses as pointed out by Duyck and De Houwer (2008).
[7] Tutorial at http://www2.gsu.edu/~eslnxj/dmdx/usedmdx.html

were initially recruited in the US and Korea from June 18, 2014 through August 22, 2014. After being recruited, their Korean proficiency was measured by a Korean C-Test adapted from Lee-Ellis (2009) to confirm that their Korean proficiency levels were advanced enough for the purpose of the present study. In this study, being *advanced* in L2 (Korean) required the ability "to read with almost complete comprehension a variety of authentic prose material on unfamiliar subjects" (the Interagency Language Roundtable (ILR), 2014) in general which corresponds to levels 2+ to 3 of the ILR, or levels from 2+ to 3 of the Defense Language Proficiency Test 5 System (DLPT5).[8]

The Korean C–test (Lee-Ellis, 2009) was created based on the ILR scale by a Defense Language Institute–certified Korean ILR passage level rating expert in order to measure Korean proficiency levels from 1 to 3 in a more practical and economically convenient way. The adapted C–test consisted of four passages containing 25 blanks in each and the second half of every second word deleted. Examinees were asked to fill in the blanks. The raw test scores were converted into percentage accuracy scores resulting in 0% to 100%. Based on the criterion of the DLPT5, if a participant answered correctly equal to or more than 60% of the blanks, s/he was determined as advanced in Korean and was asked to participate further in the entire experiment. This criterion can be considered conservative given that even some of Korean heritage speakers who had near–native proficiency in Korean obtained 64.5% on average in this Korean C–test (e.g., Ahn, 2013a).

Finally, a total of 60 (male = 13, female = 47) L2 Korean speakers who met the criterion (60% accuracy on the Korean proficiency test) participated in Experiment 1 as

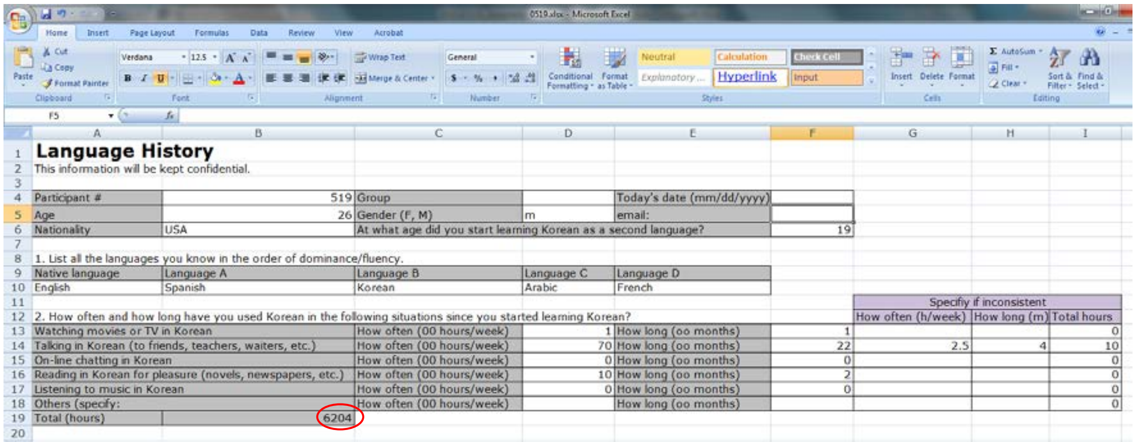
[8] For more information on the level description, see http://www.dliflc.edu/file.ashx?path=archive/documents/KPDLPT5FamGuideMC.pdf.

advanced L2 learners. Others with scores lower than 60% were not asked to participate in any further experiments. The native languages of the participants with advanced proficiency in Korean were Chinese (*n* = 32), English (*n* = 23), French (*n* = 1), German (*n* = 1), Japanese (*n* = 1), Mongolian (*n* = 1), and Thai (*n* = 1).

All L2 Korean speaking participants were late L2 learners who began learning Korean as their second language at age 18 or older. Their accuracy on the Korean C–test ranged from 60% to 94% with an average of 79%. All of them had at least some experience staying in Korea, ranging from one month to six years. Each was also taking advanced–level Korean courses for academic purposes (e.g., Korean literature, Media Korean, and Teaching Korean as a second language) at their institutions at the time of testing.

The accumulated amount of using L2 Korean (hereafter L2 use) was calculated based on the language background questionnaire adapted from Lee-Ellis (2012). Specifically, to examine the potential effect of their amount of L2 use on the automatic semantic activation separately from the amount of L2 instruction, L2 use was measured for the situations in which they were likely to use L2 in a spontaneous manner (e.g., online chatting in L2, face–to–face interaction in L2, watching television or movies). The language background questionnaire was in the format of an Excel file, and after the participants typed the required information for how many hours a week they had used L2 Korean in each of the given situations and for how many months, then the total amount was automatically calculated as shown in Figure 2.

Figure 2. An Example of L2 Use Data in Language Background Questionnaire



	A	B	C	D	E	F	G	H	I
1	**Language History**								
2	This information will be kept confidential.								
3									
4	Participant #	519	Group		Today's date (mm/dd/yyyy)				
5	Age	26	Gender (F, M)	m	email:				
6	Nationality	USA	At what age did you start learning Korean as a second language?			19			
7									
8	1. List all the languages you know in the order of dominance/fluency.								
9	Native language	Language A	Language B	Language C	Language D				
10	English	Spanish	Korean	Arabic	French				
11							Specify if inconsistent		
12	2. How often and how long have you used Korean in the following situations since you started learning Korean?						How often (h/week)	How long (m)	Total hours
13	Watching movies or TV in Korean		How often (00 hours/week)	1	How long (oo months)	1			0
14	Talking in Korean (to friends, teachers, waiters, etc.)		How often (00 hours/week)	70	How long (oo months)	22	2.5	4	10
15	On-line chatting in Korean		How often (00 hours/week)	0	How long (oo months)	0			0
16	Reading in Korean for pleasure (novels, newspapers, etc.)		How often (00 hours/week)	10	How long (oo months)	2			0
17	Listening to music in Korean		How often (00 hours/week)	0	How long (oo months)	0			0
18	Others (specify:		How often (00 hours/week)		How long (oo months)				0
19	Total (hours)	6204							
20									

As a result, participants' L2 use in the present study varied from 116 to 24,848 hours with a mean of 4,168 hours (*SD* 5,043.8) from the time of beginning studying Korean to the time of testing. The hours and duration of L2 use also varied across participants, ranging from about four hours a week for less than one year up to about 60 hours a week for about seven years.[9] The language background questionnaire took about 20 minutes to complete. Table 4 summarizes the advanced participants' language backgrounds.

Table 4. Advanced L2 Korean Speakers' Language Backgrounds

	Age at Test	Age of Acquisition of L2	Korean C-Test accuracy (%)	Length of Residence (m)	Use of L2 (h)
Mean	24.9	19.6	79	22.3	4168
SD	2.3	3.6	15.4	16.4	5043.8
Minimum	19	18	60	1	116
Maximum	33	30	94	72	24848

[9] The detailed calculation may still have limitations in that it is a rough approximation solely based on a participant's memory rather than verified data.

For the control group, 36 (male=20, female=16) L1 Korean speakers whose native and dominant language is Korean either living in the U.S. or in Korea attended in this study. Their age at testing was 30.2 on average. Their performance was compared with L2 Korean speakers' one as a norm in Korean (L2) processing.

All participants were paid $10 USD per hour, with a maximum of $30 for a three-hour experiment as L2 Korean speakers and a minimum of $10 for an hour experiment as L1 Korean speakers.

4.3. Methods

4.3.1. Materials

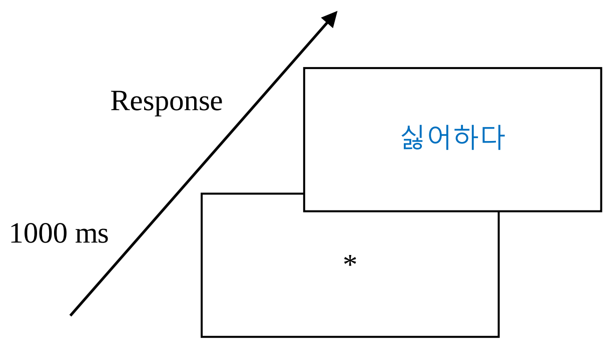
The EST included a total of 30 Korean negative emotional words (e.g., 싫어하다 *dislike*, 걱정 *worry*) and 30 Korean neutral words (e.g., 운전하다 *drive*, 가족 *family*) in each of the two colors. First, 50 emotional words in Korean and 50 neutral words in Korean were chosen mainly from five Korean text books (e.g., Cho et al., 2000a; 2000b) that are widely used in the U.S. The neutral words were chosen from various semantic categories (e.g., transport, food, furniture, etc.) because the emotional Stroop effect better appears especially either when each neutral word is from various categories or when all neutral words from one category (Sutton et al., 2007). Next, Korean stimuli were reviewed by five Korean instructors for familiarity and difficulty for L2 Korean speakers using a Likert scale anchored by1 = *for beginners* to 5 = *for natives*. Based on the reviewers' ratings, 40 emotional words and 40 neutral words rated as 1 or 2 (beginning to advanced beginning) on the Likert scale by all reviewers were selected. After piloting 34 L1 Korean speakers, the 30 emotional and 30 neutral words that most reliably produced the emotional involvement effect were chosen as the final stimuli (see Appendix 1). All words are equivalent in terms of word frequency and word length (all ps>.10) and the number of part of speech (10 words for each of noun, adjective, and verb in each) across word conditions. The split-half analyses showed high reliabilities of the task using these materials (rs > .9).

4.3.2. Procedures

All participants were tested individually. Instructions appeared in Korean on the computer screen and were repeated verbally by the experimenter. A given experimental trial proceeded as follows: a fixation mark (*) appeared for 1000 ms in the centre of the screen at the beginning of each trial, followed by a word (Figure 3). The word remained on the screen until a participant responded; if not, it remained for a maximum of 2000 ms. Then it was replaced immediately by the fixation mark as typically done in the previous studies (e.g., Sutton et al., 2007).

Participants first completed 10 trials in the practice session, each consisting of a list of symbols (e.g., ##### or $$$$$ in a specific color) instead of a word. If they pressed a wrong button, they were told "틀림" (meaning *wrong*) on the screen; otherwise, they were told "맞음" (meaning *correct*) while no feedback was provided during the main test session. Once participants reached 80% accuracy in the practice session, they moved to the main test session. Otherwise, they practiced the same set of 10 trials again. All participants except one could pass the practice session after trying it once.

Figure 3. The Stimulus Sequence within a Trial of the Emotional Stroop Task



After Experiment 1, all participants were asked to complete a word check list which included the words that were used in the EST. The purpose of administering a word check list was to make sure that participants knew all words so that the participants' different performance may be not due to their vocabulary capacity but due to their automatic processing. Participants were asked to put a check (√) in a column next to the word the meaning of which they do not know. Unknown words, if any, were removed from analysis later.

The EST was conducted before the sentence-based picture recognition task, which took about 15 minutes for L2 Korean speakers to complete. Once participants signed consent forms, they were asked to complete a Korean language proficiency test (Korean C-test), performed two main tasks (for Experiments 1 and 2), and filled out the word checklist and the language background questionnaire.

4.4. Data Preparation and Analyses

A total of 96 participants who were 36 L1 and 60 advanced L2 Korean speakers responded to 120 words in the EST. This yielded a total of 11520 data points for the EST. The data were inspected in terms of error rates, errors, and outliers. First, error rates were examined for each participant and item. If a participant had an error rate greater than 20%, he or she was excluded from analyses as conventionally done in reaction time research (e.g., Jiang, 2007). None of the participants was removed in this study; L1 Korean speakers' and L2 Korean speakers' error rates ranged from 0% to 4% and from 0% to 6%, respectively. Also, if an item had an error rate greater than 40%, it was excluded from analyses since it was close to the results of random responses. None of the items were rejected in this study based on this criterion: error rates of items in the EST task ranged from 0% to 25%. However, given that advanced L2 Korean speakers checked one word in each condition (i.e., 지치다 by 8 participants, 설거지 by 5 participants) as unknown in the vocabulary list given after the experiment, these two words in each of two colors, and thereby 4 items, were deleted in further analyses. The item deletion accounts for the loss of 384 data points (3%).

Furthermore, 136 (1%) reaction times were deleted due to incorrect decision responses, and 691 (6%) reaction times fell outside the range of acceptable latencies (mean +/- 2*SD* in each case). This threshold to remove outliers has been traditionally taken in reaction time research (e.g., Jiang, 2012). The whole procedure of data trimming accounted for total data loss of 10% in EST. Finally remaining data points were 10309 (90% of the original data) in EST. Table 5 displays the means and standard deviations (*SD*s) of the remaining reaction times in each condition and group for EST.

Table 5. Summary of Remaining Data for Experiment 1

Group	L1 Speakers		L2 Speakers	
Condition	Neutral	Emotional	Neutral	Emotional
Mean (*SD*)	490 (115)	506 (127)	498 (126)	502 (127)
Difference	16		4	

Remaining reaction times were log-transformed (base 10) to approximate normality (Baayen, 2008). All data analyses were conducted with R, an open-source statistical package R (R development core team, 2012), which is available at http://cran.r-project.org. For running mixed-effects models, the lme4 package (Bates, 2005) was used.

4.5. Results

4.5.1. Emotional Involvement in L1 Word Recognition

The mixed effects model was used to analyze L1 speakers' reaction times (RTs) on the EST.[10] In this model, the independent variable of our primary interest was condition which was set up as fixed effect. Random effects were imposed on both participants and items to control for the variation across them. This is because participants and items were

[10] Unlike the previous studies on the emotional Stroop effects where repeated measures analyses of variance (ANOVAs) were typically adopted, the present study ran mixed-effects models. This is because mixed-effects models (MEMs) have more benefits over ANOVAs for this kind of data set where participants are tested on a series of items and the same items are tested on a series of participants. MEMs can include participants and items as crossed random effects simultaneously in a single design and does not require averaging over participants or items, which allows deeper understanding of a dataset as it is. Besides, MEMs are robust against missing data which many SLA research confronts in practice, and so too does the present study. For more information regarding the benefits of MEMs on SLA research, see Cunnings (2012).

randomly selected from a larger population, and the random sampling can presumably affect the results.

Based on the descriptive statistics in Table 5, L1 speakers' RTs were 16 ms slower in the emotional condition (506 ms, *SD* 127) than in the neutral condition (490 ms, *SD* 115). The lmer( ) function in R was used to construct MEMs to test the significance of this difference. The command used in R to run the model was as follows:

```
> Model 1 = lmer (logged RTs ~ condition + (1 | subject) + (1 | item),
                  data = EST_NS)
```

We used log transformed RTs (i.e., logged RTs in the formula) to better follow the statistical assumption of the model that the error term follows the normal distribution. The independent variable was condition which was set as fixed effect. The formula (1 | subject) + (1 | item) specifies crossed random effects for participants and items. The results of Model 1 revealed significant condition effect in L1 word recognition (Table 6).

Table 6. Summary Statistics of Model 1 on L1 Speakers' Emotional Involvement

AIC	BIC	logLik	deviance
-2077.101	-2045.816	1043.550	-2087.101

Random effects:

Groups	Name	Variance	Std.Dev.
subject	(Intercept)	0.0239903	0.15489
item	(Intercept)	0.0006633	0.02575
Residual		0.0321945	0.17943

Number of obs: 3854, groups: subjects, 36; item, 116

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.174513	0.026353	234.30
condition emotional	**0.028025**	**0.007507**	**3.73**

To be specific, when the condition was emotional, L1 speakers responded more slowly (0.028025) than the neutral condition, and the difference was statistically significant ($|t| \geq 2.0$ at α=.05).[11]

In addition, we ran another model in which by-subject random slope for condition was included. Given that participants were repeatedly measured on emotional and neutral words, participants may vary in terms of their sensitivity toward the manipulation (e.g.,

[11] The significance in the mixed-effects model using R can be determined based on a *t* value instead of *p*; if the absolute *t* value is equal to or larger than 2.0, the null hypothesis is rejected and the conclusion is that the parameter is significantly different from zero.

The reason of not using *p* value for the mixed-effects model is that "the calculation of exact *p* values is not straight forward for mixed-effects models where, for example, it is not yet understood how the degrees of freedom should appropriately be calculated (see Baayen et al., 2008, p. 396; Bates, 2006). Although there are ways in which *p* values can be estimated (e.g., Baayen, 2008, p. 248), the estimated *p* value can be anticonservative (i.e. has an increased risk of Type I error) for small datasets. There is another way of estimating *p* values using the pvals.fnc() function from the LanguageR package, however, using this function results in an error, as currently the pvals.fnc() function is not yet implemented for the types of model discussed in the present study that contain random slopes" (Cunnings, 2012, pp. 377-378).

some participants may respond to the emotional words much slower than to the neutral words while other participants' response across two conditions show minimal difference). Note that since items were not crossed for emotional and neutral conditions (i.e., a word in emotional condition cannot be measured in another neutral condition, and vice versa), by-item random slope for condition cannot be included. Thus, the second model was run using the following formula:

> Model 2 = lmer (logged RTs ~ condition + (1 + condition| subject) + (1 | item),
data = EST_NS)

Table 7. Summary Statistics of Model 2 on L1 Speakers' Emotional Involvement

AIC	BIC	logLik	deviance
-2155.673	-2111.875	1084.837	-2169.673

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
subject	(Intercept)	0.0229488	0.15149	
item	(Intercept)	0.0007057	0.02656	
	condition emotional	0.0045486	0.06744	0.01
Residual		0.0310113	0.17610	

Number of obs: 3854, groups: subject, 36; item, 116

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.17478	0.02580	239.31
condition emotional	**0.02793**	**0.01353**	**2.06**

The results of the second model showed the *significant* condition effect as well. That is, even when allowing different participants to vary in terms of the sensitivity to the manipulation (condition), L1 speakers responded significantly more slowly in emotional condition than in neutral condition (0.02793), and the difference was statistically significant ($|t| \geq 2.0$ at α=.05) as shown in Table 7.

For the model comparison between the models with and without the by-subject random slope, AIC scores were considered. AIC scores measure how much variance is remained unexplained in the model. The AIC score for Model 2 (-2155.673) was smaller than that of Model 1 (-2077.101), indicating that the second model is explaining more of the variance in the data. That being said, the lower *t* value for the fixed effect of condition in the second model (2.06) than in the first (3.73) suggests that the first model containing random intercepts only produced an overconfident estimate of this effect.

In addition, in order to test whether a particular model provides a significantly improved fit to the data over another, likelihood ratio tests were conducted with the anova ( ) function in R as follows:

```
> anova (Model 1, Model 2)
```

Table 8. Summary Statistics of Model Comparison for L1 Speakers' Emotional Involvement

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
Model 1	5	-2077.1	-2045.8	1043.5	-2087.1			
Model 2	7	-2155.7	-2111.9	1084.8	-2169.7	82.573	2	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

As shown in Table 8, the result, expressed as a chi-squared statistic, displayed that Model 2 provides a significantly better fit to the data than Model 1 ($x^2(2)$=82.573, $p < .001$), indicating that the by-subject random slopes have improved model fit, and thus need to be included in the model for this data set.

In conclusion, the AIC scores and likelihood ratio test revealed that the difference between these two models was statistically *significant*, and the best-fit model confirmed that L1 speakers' color judgment of words became significantly slower when they were emotional words than when they were neutral words.[12] This result was even congruent to that of conventional ANOVAs (F=4.734, p < .05)[13] as shown in Table 9.

Table 9. Summary Statistics of ANOVA on L1 Speakers' Emotional Involvement

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Error: subject					
condition	1	0.24	0.2421	0.095	0.76
Residuals	34	86.72	2.5507		
Error: subject:condition					
condtion	**1**	**0.743**	**0.7428**	**4.734**	**0.0364 ***
Residuals	35	5.491	0.1569		
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

4.5.2. Emotional Involvement in L2 Word Recognition

The previous section confirmed that L1 speakers as a control group show a significant condition effect in the EST. In this section, L2 speakers' RTs on the EST were

[12] Its effect size was 0.5 based on the formula of Borenstein et al. (2009).
[13] The results of ANOVAs were provided here to see if the results of MEMs newly reported in the present study remain the same or different with those of the traditionally used statistic techniques for emotional Stroop research.

analyzed through the mixed-effects model and were compared with L1 speakers in regard to the emotional involvement in L2 word recognition. As same in the analysis of L1 speakers' emotional involvement, condition was our independent variable of main interest and set as fixed effect. Participants and items effects were controlled as random effects.

Based on the descriptive statistics in Table 5, L2 speakers' RTs were 4 ms slower in the emotional condition (502 ms, *SD* 127) than in the neutral condition (498 ms, *SD* 126). These RTs were log-transformed for analyses. The lmer( ) function in R was used to construct MEMs to test the significance of this difference. The command used in R to create a model was as follows:

> Model 1 = lmer (logged RTs ~ condition + (1 | subject) + (1 | item),
>
> data = EST_NNS)

Logged RTs was analyzed in terms of the independent variable, the fixed effect condition. In this model, (1 | subject) + (1 | item) specifies crossed random effects for participants and items. The results of Model 1 revealed significant condition effect in L2 word recognition (Table 10).

To be specific, when condition was emotional, L2 speakers responded significantly slower (0.010952) than neutral condition, and the difference was statistically significant ($|t| \geq 2.0$ at α=.05).

Table 10. Summary Statistics of Model 1 on L2 Speakers' Emotional Involvement

AIC	BIC	logLik	deviance
-2416.164	-2382.301	1213.082	-2426.164

Random effects:

Groups	Name	Variance	Std.Dev.
subject	(Intercept)	0.019211	0.13860
item	(Intercept)	0.000171	0.01308
Residual		0.038588	0.19644

Number of obs: 6455, groups: subject, 60; itemN, 116

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.183808	0.018305	337.8
condition emotional	**0.010952**	**0.005463**	**2.0**

In addition, another model in which by-subject random slope for condition was included was tested. As explained in the analyses of L1 speakers' RTs, this was included to control for heterogeneity in participants' responsiveness to the condition. Thus, the second model was ran using the following formula:

```
> Model 2 = lmer (logged RTs ~ condition + (1 + condition| subject) + (1 | item),
                        data = EST_NNS)
```

The results of the second model showed the *non-significant* condition effect unlike those of the first model. That is, when allowing different participants to vary in terms of the sensitivity to the manipulation (condition), L2 speakers' slower responses in the emotional condition than in the neutral condition (0.011139) became statistically non-significant ($|t| < 2.0$ at α=.05) as shown in Table 11.

Table 11. Summary Statistics of Model 2 on L2 Speakers' Emotional Involvement

AIC	BIC	logLik	deviance
-2477.994	-2430.586	1245.997	-2491.994

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
item	(Intercept)	0.0002005	0.01416	
subject	(Intercept)	0.0198455	0.14087	
	condition emotional	0.0032799	0.05727	-0.17
Residual		0.0377389	0.19426	

Number of obs: 6455, groups: subject, 60; item, 116

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.183826	0.018599	332.5
condition emotional	**0.011139**	**0.009226**	**1.2**

As the result of comparing these two models tested above, the AIC score for Model 2 (-2477.994) was smaller than for Model 1 (-2416.164), indicating that the second model is explaining more of the variance in the data. That being said, the lower *t* value for the fixed effect of condition in the second model (1.2) than in the first (2.0) suggests that the first model containing random intercepts but not random slopes provided an overconfident estimate of this effect.

In addition, the results of the likelihood ratio test showed that the inclusion of by-subject random slope for condition significantly improved the model fit ($x^2(2)=65.83$, $p < .001$) as shown in Table 12.

Table 12. Summary Statistics of Model Comparison for L2 Speakers' Emotional Involvement

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)	
Model 1	5	-2416.2	-2382.3	1213.1	-2426.2				
Model 2	7	-2478.0	-2430.6	1246.0	-2492.0	65.83	2	5.072e-15	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Thus, the difference between these two models was statistically *significant*, and the best-fit model canceled the significance of condition effect in L2 word recognition, suggesting that L2 speakers' color judgment of words may not significantly differ according to the word condition when admitting individuals' different sensitivity to condition.[14] This result was congruent to that of conventional ANOVAs (F=1.534, $p > .05$) as shown in Table 13.

[14] Its effect size (of Model 2) was 0.2 based on the formula of Borenstein et al. (2009).

Table 13. Summary Statistics of ANOVAs on L2 Speakers' Emotional Involvement

Error: subject

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
condition	1	11.05	11.053	5.58	0.0215	*
Residuals	58	114.89	1.981			

Error: subject:condition

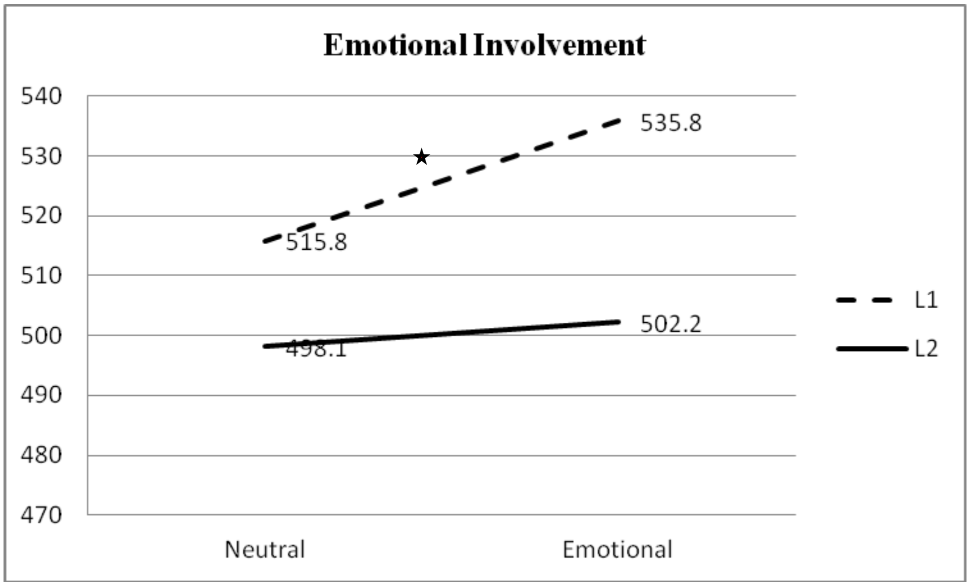
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
condtion	**1**	**0.195**	**0.1953**	**1.534**	**0.22**
Residuals	59	7.512	0.1273		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.5.3. Differences in Emotional Involvement between L1 and L2 Word Recognition

Given that L2 speakers did not show a significant condition effect unlike L1 speakers, no further analysis was needed to see the group difference in the condition effect. Figure 4 displays the significant emotional involvement in L1 word recognition but not in L2.

Figure 4. The Results of Experiment 1



(The asterisk mark (*) indicates significance at α=.05)

4.5.4. L2 Proficiency and L2 Use in Emotional Involvement in L2 Word Recognition

It was examined whether L2 speakers' levels of L2 proficiency and amounts of L2 use in daily lives contribute to the condition effect on word recognition in L2. For the analysis, each of these two predictors was centered around its mean by subtracting the predictor's overall mean from each individual value of the predictor. This method is known to help reduce collinearity within the model when including continuous predictors in a mixed-effect model (Cunnings, 2012). Thus, the two predictors (L2 proficiency and L2 use) were specified in the model as c_score and c_L2use, respectively. To see two predictors' contribution to the condition effect, their interaction effect with the condition were analyzed. Thus, interaction between condition and c_score and interaction between condition and c_L2use were included as independent variables as follows:

```
> Model = lmer (logged RTs ~ condition + condition*c_score +
                condition*c_L2use + (1 + condition| subject) + (1 | item),
                data = EST_NNS)
```

As the results, the interaction between condition effect and L2 proficiency was not significant (estimate = -1.298e-04, SE = 5.436e-04, *t*(6133) = -0.2) while that between condition effect and L2 use was positively *significant* (estimate = 1.156e-03, SE = 5.231E-04, *t*(6133) = 2.2 at α=.05) as shown in Table 14.

Table 14. Summary Statistics of Interaction between Contributing Factors and Emotional Involvement in L2

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
item	(Intercept)	0.000182	0.01349	
subject	(Intercept)	0.018967	0.13772	
	condition emotional	0.002248	0.04741	-0.11
Residual		0.037849	0.19455	

Number of obs: 6133, groups: item, 116; subject, 57

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.177e+00	1.869e-02	330.4
condition emotional	1.204e-02	8.414e-03	1.4
c_score	1.707e-03	1.260e-03	1.4
c_L2use	-7.684e-05	1.212e-03	-0.1
condition emotional:c_score	**-1.298e-04**	**5.436e-04**	**-0.2**
condition emotional:c_L2use	**1.156e-03**	**5.231e-04**	**2.2**

The significant interaction between condition effect and L2 use means that the more L2 speakers used L2 (Korean) in their daily lives, the slower responses they had in the emotional condition (and thereby, the bigger condition effects they produced). Therefore, we can conclude that the advanced L2 Korean speakers' ability to automatically activate emotional word meanings in L2 word recognition tended to be boosted with the increase in L2 use but not in L2 proficiency.

4.6. Discussion

Experiment 1 was conducted to address two research questions as follows:

- Do advanced L2 Korean speakers exhibit emotional involvement during word processing in L2 comparably with L1 Korean speakers?
- How does emotional involvement during L2 word processing change according to L2 proficiency and amount of L2 use?

Sixty advanced L2 Korean speakers who participated in this study showed less emotional involvement in L2 word recognition compared with L1 Korean speakers. That is, condition effect in the EST (or the emotional Stroop effect) was non-significant in the L2 group but significant in the L1 group. Also, the degree of emotional involvement of L2 speakers changed as L2 use, but not L2 proficiency, increased in the sense that they displayed significant interaction between L2 use and emotional Stroop effect, but non-significant interaction between L2 proficiency and emotional Stroop effect.

These two findings of Experiment 1, taken together, suggest that L2 speakers' emotional involvement in L2 word recognition can be assisted by the degree of their L2 use. Given that the interaction between condition effect and L2 use was positively significant even when allowing for different levels of sensitivity to word conditions among participants, late L2 speakers' ability to automatically activate word meaning may advance as the amount of L2 use increases. This hypothesis seems plausible when comparing L2 speakers' emotional Stroop effects across studies. Eilola et al. (2007) and Eilola and Havelka (2010) examined L1 Finish–L2 English and L1 Greek–L2 English speakers, respectively, and both found that L2 English speakers displayed significant emotional Stroop effects in their L2 as well as L1s. On the other hand, Winskel (2013) found that L1 Thai-L2 English speakers exhibited significant emotional Stroop effect in their L1 but not in L2 similar to the results of the present study. The major difference among the participants of these studies (Eilola et al., 2007; Eilola & Havelka, 2010; Winskel, 2013) seems the intensity or amount of their L2 use. The participants showing similar degree of emotional Stroop effects in their L1s and L2 (Eilola et al., 2007; Eilola & Havelka, 2010) reported to have used their L2 even more than their L1s in daily lives while those showing significant emotional Stroop effect only in their L1s but not in L2s in Winskel's study (2013) and the present study did not. Thus, considering the significant interaction between emotional Stroop effect and L2 use in the present study and such inconsistent findings according to the degrees of L2 use across the previous studies, how much L2 speakers use their second language in daily life should play a significant role in the automatic semantic activation during L2 word recognition.

In addition to amount of L2 use, level of proficiency in L2 may play a crucial role in automatic semantic activation in L2 word recognition. However, as noted previously, earlier studies on L2 speakers' emotional Stroop effects neither scrutinized nor considered the potentially contributing factors to developing automatic semantic activation. The present study attempts to address the potential roles of these factors by manipulating the L2 Korean proficiency and use of participants in analyses. Unlike the amount of L2 use, L2 proficiency appeared not to affect L2 speakers' performance in Experiment 1 in that interaction between emotional Stroop effect and L2 proficiency was not significant. The non-significant interaction between L2 proficiency and emotional Stroop effect, however, might be due to little variance in L2 proficiency among the L2 speakers recruited for this study rather than to no significance of L2 proficiency in nature. The 60 L2 speakers selected for this study were advanced Korean speakers chosen from a pool of 96 original participants based on their accuracy scores in percentage, and therefore, the range of their L2 proficiency (60%~94%) was not large enough to show its significant effect in this study.

On the other hand, amount of L2 use by participants, which appeared to be significant in condition effect, ranged from 116 to 24848 hours by the time of testing. Frequency and duration of L2 use since beginning study ranged from about four hours a week for less than one year up to around 60 hours a week for about seven years. With such a wide range (24732 h) in amount of L2 use, the advanced L2 speakers showed significant interaction between L2 use and condition effect in this study.

To sum up the major finding of this section, the present study suggests that while emotional involvement in L2 word recognition may not be as strong initially as with L1, but with increased use may develop to comparable levels in the long run.

Chapter 5. Mental Imagery in L2 Processing (Experiment 2)

5.1. Study Overview

In this chapter, we discuss Experiment 2. The purpose of the second experiment is to probe mental imagery during L2 sentence reading. The rationale of this paradigm is that sentence meaning is automatically represented as a form of imagery in readers' minds (Barsalou, 1999). To assess participants' ability to generate mental imagery during L2 sentence reading, a sentence–based picture recognition task (SPT) was employed. In SPT, participants were instructed to press the *yes* button (right shift key on the keyboard) if the object was mentioned in the previous sentence and the *no* button (left shift key) if it was not. They were told to respond as quickly and accurately as possible.

Pictures are recognized more quickly and accurately when they are congruent to what the previous sentences implied, such as orientation or shape when it concerns L1 speakers (e.g., Stanfield & Zwaan, 2001). The sentence–picture congruency effect in the form of faster response to matching pictures than to mismatching pictures can be considered a result of complete sentence comprehension. Otherwise, there would be no difference in recognizing the same object immediately following the sentence either in congruent or incongruent shape. By comparing the participants' ability to generate mental imagery to the same extent as L1 speakers, Experiment 2 attempted to address the critical issue of how meaning representation can be activated automatically in L2 sentence reading. If the sentence meaning remain inactivated or incomplete when reading a sentence, the L2 speakers' picture recognition performance would be similar, regardless of congruent or incongruent conditions. On the other hand, if the participants show sentence–picture congruency effects similar to L1 speakers', this would mean that the L2 speakers comprehend the sentence as automatically as do L1 speakers.

In addition, Experiment 2 further compared the L2 speakers' ability to generate mental imagery along with their L2 proficiency and the extent of L2 use. The result provides important evidence regarding how the level of semantic activation in L2 sentence reading may change based on L2 proficiency and L2 use. The methods and participants of the Experiment 2 are described in more detail below.

5.2. Participants

The same population as in Experiment 1 participated in Experiment 2 as well.

5.3. Methods

5.3.1. Materials

In the sentence-based picture recognition task, stimuli consisted of 48 critical items and 144 filler items (with 48 requiring *yes* responses and 96 requiring *no* responses). For the critical items, two versions of the sentences were manipulated in terms of the object shape or orientation per image, resulting in a total of 96 sentences.[15] In all critical items, although both sentences were supposed to require *yes* responses by mentioning the same

[15] This is different from the previous studies (e.g., Stanfield & Zwaan, 2001) where two pictures were used for two sentences. The way to manipulate stimuli in this study may be limited in dealing with potential idiosyncratic characteristics of the picture or the sentence. However, this limitation is not likely to affect the main purpose of this study which is investigating relative efficiency of semantic processing between L1 and L2 in that the same pair of a picture and a sentence was applied to both L1 and L2 speakers.

object as shown in a picture, one sentence of the two versions was matched and the other was mismatched with an object in a picture in terms of shape or orientation. For instance, to accompany the image of an opened book upside down (Figure 5), two sentences were used: 교수님이 복사기 위에 책을 놓아요 (*The professor placed the book on the copy machine*) for a match condition, and 교수님이 가방 안에 책을 넣어요 (*The professor put the book into the backpack*) for a mismatch condition. Therefore, the match sentences were expected to trigger *yes* responses faster than the mismatch sentences.

Figure 5. An Example of an Image

The sentences were created by the researcher including five items from the Zwaan's studies (e.g., Stanfield & Zwaan, 2001). The sentence stimuli were matched in terms of length and position of a target object noun in a sentence across conditions. A professional illustrator drew images as black-and-white line drawings based on many sources for typicality, complexity, and familiarity reported in previous studies (Snodgrass & Vanderwart, 1980; Zwaan et al., 2002, and so on). Each picture was scaled to fit into a square of about 3 inches.

In order to ensure that Korean sentences were easy enough for the participants of the present study to immediately understand Korean stimuli were reviewed by five L1

Korean speakers who were Korean linguistics as well as Korean instructors with a teaching experience of 10 years on average. Then, they participated in a pretest. The purpose of the pretest was to ensure that only one of the two versions of sentences matches the corresponding picture (i.e., a vertical image should not be matched with a horizontal sentence, and *vice versa*). The 120 critical sentences (60 congruent + 60 incongruent sentences) were presented with the 60 corresponding images. The reviewers were asked to choose one of three options: 1) only one version of the critical sentences best matches the picture; 2) two versions of the critical sentences equally match the picture; 3) neither sentence matches the picture. The sets of two sentences and an image rated as an option 1) by all five reviewers were chosen as the 48 final critical items (Appendix 2).

In addition, a total of 144 filler items were created to direct participants' attention away from the critical items: 48 *yes*-response filler items, in order to have the equal number of critical items, and 96 *no*-response fillers, in order to have the equal number of 'yes' items. As an example of *yes*-response fillers, an image of a watch was presented after a sentence saying 아빠는 30 년 동안 이 시계를 사용하세요 (*My father has been using this watch for 30 years*) since no particular shape is implied in this filler. As a *no*-response filler item, an image of a car was presented after a sentence saying 태수가 기차역에 도착했을 때 기차가 모두 떠나고 없어요 (*All trains were gone when Paul arrived at the station)* where no car was mentioned in the sentence at all. In this way, a total of 144 sentence-image pairs functioned as filler items.

There were two counterbalanced stimulus lists. Each list had one of two versions of 48 critical items counterbalanced (i.e., among 96 matched and mismatched pairs of an images and a sentence, having the first half of match pairs (24) and the second half of mismatch pairs (24) or vice versa) and 144 identical filler items. As a result, each participant saw one of the lists containing 48 critical items (24 items for match condition and 24 items for mismatch condition) counterbalanced and requiring *yes* responses, 48 *yes*-response filler items, and 96 *no*-response filler items in random order. The finalized lists were pretested with another 34 L1 Korean speakers to ensure the items produced the intended sentence-picture congruency effects. As a result, the Korean stimuli showed significant sentence-picture congruency effects (p <.05). The items neither produced error rate 40% or higher (one item) nor elicited the reverse patterns (longer response latency in match condition than in mismatch condition) among the native Korean speakers. The split-half analyses showed high reliabilities of the task using these materials (rs > .7).

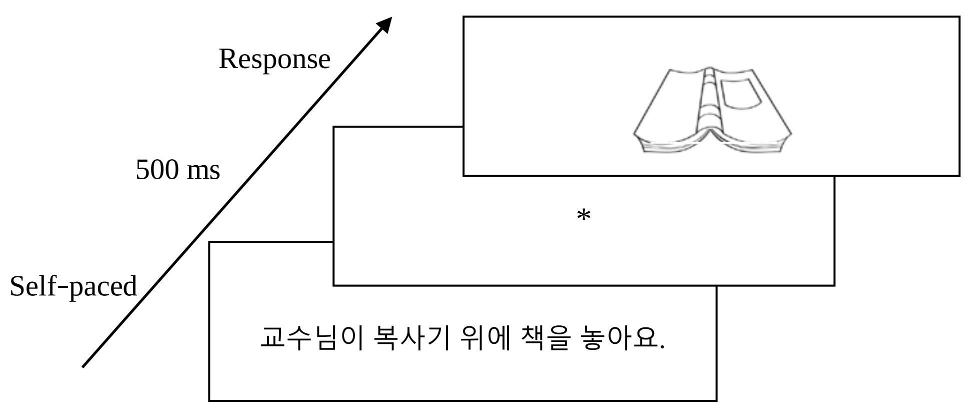
5.3.2. Procedures

In the SPT, participants were asked to decide whether the object presented in the picture was mentioned in the sentence. They were instructed to press *yes* if the object was mentioned, and they were instructed to press *no* if not. The task presented a sentence in the center of the computer screen first. When the participants understood the sentence, they were supposed to press the right shift key on a QWERTY keyboard. A fixation mark (*) then appeared for 500 ms in the center of the screen followed by the image of the object. As soon as participants pressed the *yes* (right-shift key) or *no* (left-shift key) buttons, the next trial (*sentence* → * → *image*) automatically began (Figure 6). Reaction time was measured from the time an image appeared to the time the participants pressed the response key.

The task was developed and conducted with DMDX. It consisted of two phases: the practice phase and the test phase. In the practice phase, participants were familiarized with the experimental task by practicing 12 trials with feedback telling them whether their response was correct or not on the screen after each trial. Any feedback other than this (e.g., feedback regarding orientation) was not provided in order to prevent participants from noticing the purpose of this study. If a participant's accuracy was below 80%, they repeated the practice items until accuracy exceeded 80%. Once they reached the 80% accuracy level in the practice phase, they moved to the test phase by pressing the space bar on a keyboard.

Figure 6. The Stimulus Sequence within a Trial of

the Sentence-based Picture Recognition Task



In the test phase, each participant saw a total of 192 trials and 48 comprehension questions in random order in each language. There was no feedback on their responses in the test phase. For an example of the comprehension questions, a sentence asking 그가

연필을 갖고 있어요? (*Did he have a pencil?*) followed participants' responses for a sentence stating 훈이는 연필을 컵에 넣어요 (*Hoon puts the pencil in the cup*). The purpose of including the comprehension questions was to ensure that the participants were paying attention to the semantics of the sentences. A pilot study indicated that some participants became more strategic after a few trials, so that they only read the noun in a sentence without fully processing the whole sentence. However, understanding the meaning of a sentence is essential to successfully examine the sentence-picture congruency effects (Jiang, 2012). Including comprehension questions was, in fact, an effective way to prevent participants from only reading nouns to recognize a picture quickly and to encourage them to try to fully understand the meaning of the entire sentence.

The experiment took about 50 minutes for L2 Korean speakers to complete. The order of the stimulus list was counterbalanced across participants.

5.4. Data Preparation and Analyses

A total of 96 participants, who participated in Experiment 1 (36 L1 and 60 advanced L2 Korean speakers), responded to 48 critical items in the SPT. This yielded a total of 4608 data points for SPT. This data set was inspected in terms of error rates, errors, and outliers. First, error rates were examined for each participant and item. If a participant had an error rate greater than 20%, he or she was excluded from analyses as done for the experiment 1. None of the participants was removed in this study: L1 Korean speakers' error rates ranged from 0% to 9% in SPT; L2 Korean speakers' error rates fell between 0% and 17% in SPT. Also, if an item had an error rate greater than 40%, it was excluded from analyses. None of the items were rejected in this study based on this criterion: error rates of items in the SPT fell between 0% and 31%. However, given that advanced L2 Korean speakers checked two items as difficult to understand in the vocabulary checklist given after the experiment (i.e., item #14 and #44 by 5 and 8 participants, respectively), these two items were deleted in further analyses. The item deletion accounts for the loss of 192 data points (4%).

Furthermore, 286 (6%) reaction times were deleted due to incorrect decision responses in SPT, and 346 (7%) reaction times fell outside the range of acceptable latencies (mean +/- 2*SD* in each case) in SPT. The whole procedures of data trimming counted for the total data loss of 16% in SPT. Finally remaining data points were 3899 (84% of the original data) in SPT. Table 15 summaries the means and *SD*s of the remaining reaction times in each condition and group for SPT.

Table 15. Summary of Remaining Data for Experiment 2

Group	L1 Speakers		L2 Speakers	
Condition	match	mismatch	match	mismatch
Mean (*SD*)	724 (278)	753 (290)	837 (397)	881 (441)
Difference	29		44	

Remaining reaction times were log-transformed (base 10) to approximate normality (Baayen, 2008). All data analyses were conducted with R. For the mixed-effects modeling, the lme4 package (Bates, 2005) was used.

5.5. Results

5.5.1. Mental Imagery in L1 Sentence Reading

The mixed effects model was used to analyze L1 speakers' reaction times (RTs) on the SPT.[16] In this model, the independent variable of our primary interest was condition which was set up as fixed effect. Random effects were imposed on both participants and items to control for the variation across them. This is because participants and items were randomly selected from a larger population, and the random sampling can presumably affect the results.

Based on the descriptive statistics in Table 15, L1 speakers' RTs were 29 ms slower in the mismatch condition (753 ms, *SD* 290) than in the match condition (724 ms, *SD* 278). The lmer( ) function in R was used to construct MEMs to test the significance of this difference. The command used in R to run the model was as follows:

```
> Model 1 = lmer (logged RTs ~ condition + (1 | subject) + (1 | item),
                  data = SPT_NS)
```

We used log transformed RTs (i.e., logged RTs in the formula) to better follow the statistical assumption of the model that the error term follows the normal distribution. The independent variable was condition which was set as fixed effect. The formula (1 | subject) + (1 | item) specifies crossed random effects for participants and items. The results of Model 1 revealed *significant* condition effect in L1 sentential reading (Table 16).

[16] Likewise, running MEMs has more advantages over ANOVAs to better understand this type of data where participants are measured on a series of items, and vice versa. See Cunnings (2012) for more detailed explanations.

Table 16. Summary Statistics of Model 1 on L1 Speakers' Mental Imagery

AIC	BIC	logLik	deviance
661.4665	688.0626	-325.7333	651.4665

Random effects:

Groups	Name	Variance	Std.Dev.
item	(Intercept)	0.01339	0.1157
subject	(Intercept)	0.04762	0.2182
Residual		0.07886	0.2808

Number of obs: 1509, groups: item, 46; subject, 36

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.53834	0.04147	157.65
condition mismatch	**0.04183**	**0.01451**	**2.88**

To be specific, when condition was mismatch, L1 speakers responded more slowly (0.04183) than match condition, and the difference was statistically significant ($|t| \geq 2.0$ at α=.05).

In addition, we ran another model in which by-subject random slope for condition was included. Given that participants were repeatedly measured on match and mismatch items, participants may vary in terms of their sensitivity toward the manipulation (e.g., some participants may respond to the mismatch items much slower than to the match items while other participants' response across two conditions show minimal difference). Thus, the second model was run using the following formula:

> Model 2 = lmer (logged RTs ~ condition + (1 + condition| subject) + (1 | item)),

data =SPT_NS)

The results of the second model showed the significant condition effect as well. That is, even when allowing different participants to vary in terms of the sensitivity to the manipulation (condition), L1 speakers responded significantly more slowly in mismatch condition than in match condition (0.04151), and the difference was statistically significant ($|t| \geq 2.0$ at α=.05) as shown in Table 17.

Table 17. Summary Statistics of Model 2 on L1 Speakers' Mental Imagery

AIC	BIC	logLik	deviance
665.3995	702.6339	-325.6997	651.3995

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Item	(Intercept)	1.338e-02	0.115668	
subject	(Intercept)	4.852e-02	0.220275	
	cond mismatch	1.741e-05	0.004173	-1.00
Residual		7.885e-02	0.280810	

Number of obs: 1509, groups: item, 46; subject, 36

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.53847	0.04177	156.53
condition mismatch	**0.04151**	**0.01453**	**2.86**

On top of that, we ran another model in which by-item, in addition to by-subject, random slopes for condition was included. Given that items were repeatedly measured on match and mismatch conditions within minimal pairs of sentences (i.e., a sentence in match condition can be measured in mismatch condition with a minimal pair changed), items may vary in terms of their sensitivity toward the manipulation (e.g., some item pairs may trigger much more delayed responses in mismatch condition than in match condition while other item pairs across two conditions show minimal difference). Thus, the third model was run using the following formula:

> Model 3 = lmer (logged RTs ~ condition + (1+condition| subject)

(1 + condition | item), data = SPT_NS)

The result of the third model showed the significant condition effect as well. That is, even when allowing different items, in addition to different participants, to vary in terms of the sensitivity to the manipulation (condition), L1 speakers responded more slowly in a mismatch condition than in a match condition (0.04346), and the difference was statistically significant ($|t| \geq 2.0$ at α=.05) as shown in Table 18.

Table 18. Summary Statistics of Model 3 on L1 Speakers' Mental Imagery

AIC	BIC	logLik	deviance
665.5459	713.4188	-323.7730	647.5459

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
item	(Intercept)	0.0116958	0.108147	
	cond mismatch	0.0038071	0.061702	0.10
subject	(Intercept)	0.0481583	0.219450	
	cond mismatch	0.0000144	0.003795	-0.83
Residual		0.0779110	0.279125	

Number of obs: 1509, groups: item, 46; subject, 36

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.53737	0.04120	158.69
cond mismatch	**0.04346**	**0.01708**	**2.54**

For the model comparison among the models with and without the by-subject and/or by-item random slopes, AIC scores were considered. The AIC score for Model 1 (661.47) was smaller than the others (the AIC scores of Model 2 and Model 3 were 665.40 and 665.55, respectively), indicating that the first model is explaining more of the variance in the data.

In addition, in order to test whether a particular model provides a significantly improved fit to the data over another, likelihood ratio tests were conducted with the anova ( ) function in R as follows:

> anova (Model 1, Model 2, Model 3)

Table 19. Summary Statistics of Model Comparison for L1 Speakers' Mental Imagery

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
Model 1	5	661.47	688.06	-325.73	651.47			
Model 2	7	665.40	702.63	-325.70	651.40	0.0671	2	0.9670
Model 3	9	665.55	713.42	-323.77	647.55	3.8535	2	0.1456

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

As shown in Table 19, the result, expressed as a chi-squared statistic, displayed that none of the three models provides a significantly better fit to the data than the others (*p*s > .05), indicating that the neither by-subject nor by-item random slopes have improved model fit, and thus do not need to be included in the model for this data set.

In conclusion, the AIC scores and likelihood ratio test revealed that the difference between these three models were not statistically significant. All of the three models confirmed that L1 speakers' picture recognition after reading a sentence became significantly slower when the picture showed the object in a direction or shape mismatched to what was implied in the previous sentence than matched, no matter of admitting different sensitivity of items and subjects to condition.[17] This result was even congruent to that of conventional ANOVAs (F=8.419, $p < .01$) as shown in Table 20.

[17] Its effect size (based on Model 1) was 0.6 based on the formula of Borenstein et al. (2009).

Table 20. Summary Statistics of ANOVAs on L1 Speakers' Mental Imagery

Error: subject

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
condition	1	3.02	3.023	1.693	0.202
Residuals	34	60.71	1.786		

Error: subject:condition

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
condtion	**1**	**0.6547**	**0.6547**	**8.419**	**0.00638 ****
Residuals	35	2.7218	0.0778		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5.5.2. Mental Imagery in L2 Sentence Reading

The previous section confirmed that L1 speakers as a control group showed significant condition effect in the sentence-based picture recognition task (SPT). In this section, L2 speakers' RTs on the SPT were analyzed through the mixed-effects model and were compared with L1 speakers in regard to the mental imagery in L2 sentential reading. As same in the analysis of L1 speakers' mental imagery, condition was our independent variable of main interest and set as fixed effect. Participants and items effects were controlled as random effects.

Based on the descriptive statistics in Table 15, L2 speakers' RTs were slower in the mismatch condition (881 ms, *SD* 441) than in the match condition (837 ms, *SD* 397). These RTs were log-transformed for analyses. The lmer( ) function in R was used to construct MEMs to test the significance of this difference. The command used in R to create a model was as follows:

> Model 1 = lmer (logged RTs ~ condition + (1 | subject) + (1 | item),

data = SPT_NNS)

Logged RTs were analyzed in terms of the independent variable, the fixed effect condition. In this model, (1 | subject) + (1 | item) specifies crossed random effects for participants and items. The results of Model 1 revealed significant condition effect in L2 sentence reading (Table 21).

To be specific, when condition was mismatch, L2 speakers responded more slowly (0.05464) than match condition, and the difference was statistically significant ($|t| \geq 2.0$ at α=.05).

Table 21. Summary Statistics of Model 1 on L2 Speakers' Mental Imagery

AIC	BIC	logLik	deviance
2356.898	2385.794	-1173.449	2346.898

Random effects:

Groups	Name	Variance	Std.Dev.
subject	(Intercept)	0.02826	0.1681
item	(Intercept)	0.02810	0.1676
Residual		0.14132	0.3759

Number of obs: 2390, groups: subject, 60; item, 46

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.65105	0.03462	192.13
condition mismatch	**0.05464**	**0.01548**	**3.53**

In addition, another model in which by-subject random slope for condition was included was tested. As explained in the analyses of L1 speakers' RTs, this was included to control for heterogeneity in participants' responsiveness to the condition. Thus, the second model was run using the following formula:

> Model 2 = lmer (logged RTs ~ condition + (1 + condition| subject) + (1 | item),

data = SPT_NNS)

The results of the second model showed the significant condition effect as well. That is, even when allowing different participants to vary in terms of the sensitivity to the manipulation (condition), L2 speakers' slower responses in mismatch condition than in match condition (0.05436), and the difference was statistically significant ($|t| \geq 2.0$ at α=.05) as shown in Table 22.

Table 22. Summary Statistics of Model 2 on L2 Speakers' Mental Imagery

AIC	BIC	logLik	deviance
2360.817	2401.270	-1173.408	2346.817

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
subject	(Intercept)	2.903e-02	0.170372	
cond	mismatch	2.151e-05	0.004638	-1.00
item	(Intercept)	2.810e-02	0.167629	
Residual		1.413e-01	0.375921	

Number of obs: 2390, groups: subject, 60; item, 46

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.65117	0.03480	191.12
condition mismatch	**0.05436**	**0.01549**	**3.51**

On top of that, we ran another model in which by-item, in addition to by-subject, random slopes for condition were included. Given that items were repeatedly measured on match and mismatch conditions within minimal pairs of sentences, items may vary in terms of their sensitivity toward the manipulation (e.g., some item pairs may trigger much more delayed responses in mismatch condition than in match condition while other item pairs across two conditions show minimal difference). Thus, the third model was run using the following formula:

> Model 3 = lmer (logged RTs ~ condition + (1 + condition | subject)
+ (1 + condition | item), data = SPT_NNS)

The result of the third model showed the significant condition effect as well. That is, even when allowing different items, in addition to different participants, to vary in terms of the sensitivity to the manipulation (condition), L2 speakers responded more slowly in mismatch condition than in match condition (0.05398), and the difference was statistically *significant* ($|t| \geq 2.0$ at α=.05) as shown in Table 23.

Table 23. Summary Statistics of Model 3 on L2 Speakers' Mental Imagery

AIC	BIC	logLik	deviance
2348.809	2400.821	-1165.405	2330.809

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
subject	(Intercept)	2.953e-02	0.171841	
	cond mismatch	3.122e-05	0.005587	-1.00
item	(Intercept)	3.234e-02	0.179835	
	cond mismatch	1.199e-02	0.109511	-0.34
Residual		1.382e-01	0.371758	

Number of obs: 2390, groups: subject 60; item, 46

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.65242	0.03619	183.82
cond mismatch	**0.05398**	**0.02231**	**2.42**

For the model comparison among the models with and without the by-subject and/or by-item random slopes, AIC scores were considered. The AIC score for Model 3 (2348.8) was smaller than the others (the AIC scores of Model 1 and Model 2 were 2356.9 and 2360.8, respectively), indicating that the third model is explaining more of the variance in the data. That being said, the lower *t* value for the fixed effect of condition in the third model (2.42) than in the others (3.53 and 3.51 in Model 1 and Model 2, respectively) suggests that the others (Model 1 and Model 2) without by-item random slope for condition provided an overconfident estimate of the condition effect.

In addition, the results of the likelihood ratio test showed that the inclusion of by-item random slope for condition in addition to subject random slope for condition significantly improved the model fit ($x^2(2)=16.01$, $p < .001$) as shown in Table 24.

Table 24. Summary Statistics of Model Comparison for L2 Speakers' Mental Imagery

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)	
Model 1	5	2356.9	2385.8	-1173.5	2346.9				
Model 2	7	2360.8	2401.3	-1173.4	2346.8	0.0818	2	0.9599487	
Model 3	**9**	**2348.8**	**2400.8**	**-1165.4**	**2330.8**	**16.0071**	**2**	**0.0003343**	*******

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Taken together, the AIC scores and likelihood ratio test revealed that the difference between these three models were statistically significant, and the best-fit model confirmed the significance of condition effect in L2 sentence reading shown in the other models. It means that L2 speakers recognized a picture after reading a sentence significantly slower when an object of the picture was mismatched to what was implied in the previous sentence than matched, no matter of admitting different sensitivity of items and subjects to condition, similarly to L1 speakers.[18] This result was congruent to that of conventional ANOVAs (F=9.523, $p < .01$) as displayed in Table 25.

[18] Its effect size (of Model 3) was 0.4 based on the formula of Borenstein et al. (2009).

Table 25. Summary Statistics of ANOVAs on L2 Speakers' Mental Imagery

Error: subject					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
condition	1	1.46	1.459	1.248	0.268
Residuals	58	67.79	1.169		
Error: subject:condition					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
condtion	**1**	**1.125**	**1.1245**	**9.523**	**0.00309 ****
Residuals	59	6.967	0.1181		
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

5.5.3. Differences in Mental Imagery between L1 and L2 Sentence Reading

The previous section confirmed a significant condition effect in sentence reading for both L1 and L2 speakers. In this section, interaction effect between the condition variable and the group variable was examined to see if there was any difference in the degree of condition effect between groups. Namely, we investigated whether there was any meaningful difference in the degree of automatic semantic activation between two groups.

In this analysis, the primary interest, the interaction between the condition variable and the group variable was set up as a fixed effect. Also, intercept and condition effect of every subject and item were set as random effects as below:

> Model = lmer (logged RTs ~ condition*group + (1 + condition| subject)

+ (1 + condition | item), data = SPT_ALL)[19]

[19] To see group difference in condition effect, the model having both by-subject and by-item random slopes for condition was considered given that its fit was the best for L2 group data. Two alternative models (i.e., one having random effects' intercepts only and the other having intercepts and by-subject random slope for

The result showed that the difference in condition effect on sentence reading (i.e., sentence-picture congruency effect) between L1 and L2 groups was *not significant* (Table 26).

Table 26. Summary Statistics of Interaction between Group and Mental Imagery

AIC	BIC	logLik	deviance
3179.477	3248.430	-1578.739	3157.477

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
subject	(Intercept)	3.595e-02	0.189605	
	desc mismatch	1.059e-05	0.003254	-1.00
item	(Intercept)	1.980e-02	0.140697	
	desc mismatch	7.215e-03	0.084944	-0.23
Residual		1.186e-01	0.344435	

Number of obs: 3899, groups: subject, 96; item, 46

Fixed effects:

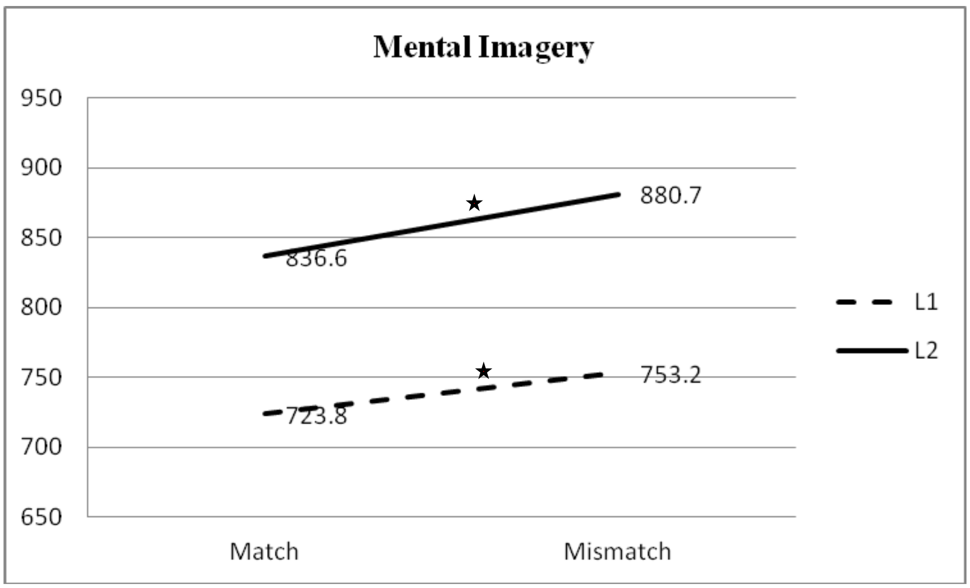
	Estimate	Std. Error	t value
(Intercept)	6.53723	0.03985	164.05
cond mismatch	0.04371	0.02176	2.01
group_L2	0.11186	0.04306	2.60
cond mismatch:group_L2	**0.01001**	**0.02276**	**0.44**

To be specific, both L1 and L2 speakers responded more slowly (0.04371) when condition was mismatch than when it was match, and the condition effect was significant (t = 2.01). Also, L2 speakers (group_L2) responded more slowly than L1 speakers overall

condition) were tested as well. However, the likelihood ratio test revealed the current model fits L1-and-L2 combined data better than the others ($x^2(2)$= 20.2329, p < .001).

(0.11186), and the difference was significant (t = 2.60). However, the condition effect in L2 speakers (0.01001) was *not significantly* different from that in L1 speakers ($|t|<2.0$ at α=.05), indicating that the level of automatic semantic activation in L2 speakers was similar to that in L1 speakers during L2 sentence reading. Figure 7 displays the significant condition effects both in L1 and L2 sentence reading.

Figure 7. The Results of Experiment 2



(The asterisk mark (*) indicates significance at α=.05)

5.5.4. L2 Proficiency and L2 Use in Mental Imagery during L2 Sentence Reading

Given that L2 speakers showed a significant condition effect similar to L2 speakers, it was further examined whether L2 speakers' levels of L2 proficiency and amounts of L2 use in daily lives contribute to such comparable degrees of condition effect on sentence reading between L1 and L2 speakers.

For the analysis, each of these two predictors was centered around its mean by subtracting the predictor's overall mean from each individual value of the predictor. This method is known to help reduce collinearity within the model when including continuous predictors in a mixed-effect model (Cunnings, 2012). Thus, the two predictors (L2 proficiency and L2 use) were specified in the model as c_score and c_L2use, respectively. To see two predictors' contribution to the condition effect, their interaction effect with the condition were analyzed. Thus, interaction between condition and c_score and interaction between condition and c_L2use were included as independent variables as follows:

```
> Model = lmer (logged RTs ~ condition + condition*c_score +
                condition*c_L2use + (1 + condition| subject) +
                (1+condition | item), data = SPT_NNS_LGB)
```
[20]

As the results, the interaction between a condition effect and L2 proficiency was *not significant* (estimate = -9.602e-04, SE = 8.679e-04, $t(2293) = -1.11$) while that between condition effect and L2 use was positively *significant* (estimate = 7.601e-06, SE = 2.709e-06, $t(2293) = 2.81$ at α=.05) as shown in Table 27.

The significant interaction between a condition effect and L2 use means that the more L2 speakers used L2 (Korean) in their daily lives, the slower responses they had in

[20] The model having both by-subject and by-item random slopes for condition was considered first given that its fit was the best for L2 group's language background data among the three. The other two models (one having random effects' intercepts only and the other having intercepts and by-subject random slope for condition) were tested as well. However, the likelihood ratio test revealed the current model fits L2 group's language background data better than the others ($x^2(2)$= 13.9172, $p < .001$).

the mismatch condition (and thereby, the bigger condition effects they produced). Therefore, we can conclude that L2 use positively affected the sentence-picture congruency effect among advanced L2 speakers while L2 proficiency did not.

Table 27. Summary Statistics of Interaction between Contributing Factors and Mental Imagery in L2

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
subject	(Intercept)	3.100e-02	0.176057	
	cond mismatch	8.524e-05	0.009233	-1.00
item	(Intercept)	3.275e-02	0.180979	
	cond mismatch	1.191e-02	0.109119	-0.33
Residual		1.381e-01	0.371643	

Number of obs: 2293, groups: subject, 57; item, 46

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.644e+00	3.816e-02	174.12
cond mismatch	3.972e-02	2.326e-02	1.71
c_score	4.986e-04	1.414e-03	0.35
c_L2use	-3.013e-07	4.528e-06	-0.07
cond mismatch:c_score	**-9.602e-04**	**8.679e-04**	**-1.11**
cond mismatch:c_L2use	**7.601e-06**	**2.709e-06**	**2.81**

5.6. Discussion

Experiment 2 was conducted to address the following two research questions:

- Do advanced L2 Korean speakers generate mental imagery during a sentential reading in L2 as automatically as L1 speakers do?
- How does the ability to automatically generate mental imagery during a sentential reading in L2 change according to L2 proficiency and amount of use?

Results showed that the 60 advanced L2 Korean speakers who participated in this study generated mental imagery in L2 sentential reading at a comparable level of automaticity with L1 Korean speakers in that both L1 and L2 speakers showed significant condition effect in the sentence-based picture recognition task (or sentence-picture congruency effect) without significant difference between the L1 and L2 groups. Also, the L2 speakers' ability to generate mental imagery changed as L2 use, but not L2 proficiency, increased. That is, they displayed significant interaction between L2 use and sentence-picture congruency effect but not between L2 proficiency and sentence-picture congruency effect.

Taken together, the first and second findings of Experiment 2 suggest that advanced L2 speakers, despite being 18 years of age or older when beginning to learn their second language, are likely to eventually reach a level of automatic semantic activation in L2 sentence reading comparable with native speakers. First, sentence-picture congruency effects among L2 speakers appeared to be reliably significant through all of the three statistic models, regardless of whether by-subject and/or by-item random slopes for condition were added, as with L1 speakers. This implies that most of the advanced L2 speakers who participated in this study have generated mental imagery according to sentence meaning even when admitting different levels of sensitivity of participants and items to manipulations (match versus mismatch conditions) in the model.

Furthermore, L2 speakers generated mental imagery during sentential reading in L2 not only reliably, but also comparably with L1 speakers in the sense that there was no significant group difference in terms of sentence-picture congruency effect. The size of sentence-picture congruency effect was bigger in L2 speakers (44 ms) than in L1 speakers (29 ms), but the 15 ms difference between the groups was not significant, indicating that it could just be an accidental occurrence. The ability of a late L2 learner to achieve levels of automatic semantic activation during L2 sentence processing comparable with L1 speakers may be assisted by extensive L2 use. This finding suggests that extensive L2 use can help even late L2 learners to activate sentence meaning during L2 sentence reading as quickly as L1 speakers.

On the other hand, the effect of L2 proficiency appeared non-significant on sentence-picture congruency effect in the present study. As discussed in the previous section, lack of significant impact of L2 proficiency on a sentence-picture congruency effect could result from little variance in L2 proficiency among participants in the present study. Again, their scores –measured on the L2 (Korean) proficiency test– only varied between 60% and 94% accuracy. With such insufficient variances, levels of L2 proficiency might be too homogenous to reveal significance of its effect on a sentence–picture congruency effect. Thus, non-significance of L2 proficiency effect in mental imagery in the present study cannot be interpreted to mean that this variable does not play a role in developing semantic activation in L2 sentence processing.

Indeed, there is a study reporting a noticeable impact of L2 proficiency on generating mental imagery.[21] Ahn, Jiang, and Osthus (2012) examined more or less fluent L2 English speakers using the same format of sentence–based picture recognition task. Their results showed different degrees of sentence-picture congruency effect according to participants' L2 proficiency levels. Specifically, the findings showed that more fluent L2 English speakers exhibited significant sentence–picture congruency effect, while less fluent L2 English speakers did not. These findings suggest that L2 proficiency may appear as a contributing factor in developing automatic semantic activation in L2 processing as long as participants are sufficiently varied in terms of L2 proficiency.

To conclude this section, the present study found that late L2 speakers could generate mental imagery corresponding to sentence meaning in L2 with similar automaticity to L1 speakers. It also reveals that extensive L2 use can contribute to late L2 learners attaining levels of semantic activation comparable with L1 speakers.

[21] To date, the study by Ahn, Jiang, and Osthus (2012) is the only known research to probe mental imagery in the minds of L2 speakers during L2 processing.

Chapter 6. General Discussion

The present study examined whether late L2 learners show semantic activation in L2 processing as automatically as do L1 speakers. It also investigated whether L2 proficiency and amount of L2 use contribute to increasing the level of automatic semantic activation in L2 processing. To address these questions, the present study probed late L2 Korean learners' emotional involvement and mental imagery generation in L2 processing to examine semantic activation at the word– and sentence– levels of processing, respectively. This chapter discusses the main findings of the present study and their significance along with the implications for future research and L2 learning.

6.1. Automatic Semantic Activation in Late L2 Speakers

The findings of the present study reveal that the degrees of automaticity in semantic activation are asymmetric during word recognition but symmetric during sentential reading between L1 and L2 processing (Table 28). The same participants took part in the study's two experiments, and the L2 group showed different degrees of automatic semantic activation according to the levels of processing when compared to the L1 group. Specifically, late L2 Korean learners did not appear to process word meaning as automatically as did L1 speakers in Experiment 1 (EST). However, Experiment 2 (SPT) showed automatic activation of sentential meaning to be comparable between the L1 and L2 groups.

Table 28. The Findings of the Present Study

Experiments	Emotional involvement in L2 word processing	Mental imagery in L2 sentence processing
Results	L2 speakers ≠ L1 speakers	L2 speakers = L1 speakers

6.1.1. Automatic Semantic Activation in L2 Word Recognition

Advanced L2 speakers' limited access to L2 word meaning revealed in the present study is consistent with previous findings of weak semantic representation in L2 word processing (e.g., Sholl et al., 1995; Zhao et al., 2011). In Sholl et al.'s study (1995), L2 Spanish speakers, who rated their L2 proficiency as high (about 8 out of 10 on a 10-point scale, with 10 being highly fluent), translated L1 words to L2 words more quickly when the L1 words were mentioned in a previous picture naming task than when they were not. On the other hand, their translation of L2 words to L1 words was not affected by previous naming. This indicates L1 words are more sensitive to prior semantic activation than L2 words, even among highly fluent L2 speakers. Also, Zhao et al.'s study (2011) displayed similar results that proficient L2 English speakers produced more semantic priming effects in forward translation than in backward. This implies that L2 words are not conceptually mediated as much as L1 words are when they are translated. Taken together, semantic representation of L2 words may be inherently less automatic and weaker than L1 words. This serves as evidence for the asymmetric strengths of conceptual links from L1 and L2 words as suggested by the revised hierarchical model (RHM; Kroll & Stewart, 1994).

From this perspective, although the L2 speakers were judged as proficient by an objective test, emotional connotations of L2 words could not be accessed in their minds as automatically as in L1 speakers' via weaker semantic connections from L2 words than from L1 words in the present study. This is similar to findings in the studies of Sholl et al. (1995) and Zhao et al. (2011).

At the same time, however, inconsistent findings also exist in the previous literature. For example, Duyck and De Houwer (2008) reported similar degrees of semantic interference effects on a letter–case judgment task between L1 and L2. They thereby suggested this as strong evidence against the RHM (especially because such symmetry occurred among late L2 speakers in a conservative task). However, two major differences between the study of Duyck and De Houwer (2008) and the present study cast doubts to such a claim against the RHM. First, L2 English speakers in their study could have been even more proficient in their L2 than those of the present study, even though both groups were described as unbalanced and late L2 speakers. The authors described their L2 English speakers' language backgrounds as having used their L2 extensively through mass media since the teenage years. Unlike the L2 Korean speakers in the present study who began studying Korean later than age 18 as the earliest, the L2 English speakers in Duyck and De Houwer's study (2008) may have had near–native proficiency in their L2. As such, they cannot be put under the same umbrella of late and unbalanced L2 speakers with those in the present study.

Second, the letter–case judgment task seems even less conservative in terms of allowing semantic access than EST adopted in the present study. In Duyck and De

Houwer's study (2008) L1 Dutch– L2 English speakers were asked to say either "animal" when a target word was presented in uppercase or "occupation" when it was in lowercase, regardless of its meaning. Thus, whereas both the letter–case judgment task and the EST can be conducted successfully without semantic processing at all (by focusing on letter case and color, respectively), the former still required verbal responses, allowing semantic access more than the latter. Consequently, participants could be more easily distracted by semantic information in the Duyck and Houwer's study (2008) than those in the present study, resulting in different degrees of semantic activation. In this sense, the discrepancy between the present study and Duyck and Houwer's study (2008) is likely attributed mainly to the different stages in L2 semantic development of more or less proficient L2 speakers and presumably to the extent to which a task involved semantic information in each study, not necessarily as compelling evidence against each other.

How, then, do we interpret the inconsistent findings within the same paradigm of emotional Stroop, which is the case where diverse groups of L2 speakers were examined on a truly conservative task in terms of semantic access? Among the four previous studies using an EST for L2 processing research, only one, Winskel (2013), showed findings consistent with those of the present study. The other studies (Eilola & Havelka, 2010; Eilola et al., 2007; Sutton et al., 2007) displayed conflicting results. In the case of Sutton et al.'s study (2007), the evidence is not exactly conflicting in that L2 speakers who were balanced and simultaneous bilinguals produced emotional involvement (or emotional Stroop effect) to the comparable degrees between L1 and L2.

On the other hand, the other studies all examined unbalanced L2 speakers and still exhibited contrary findings: L2 speakers showed difficulty in automatically representing emotional connotations of L2 words in Winskel's study (2013) similar to those of the present study but displayed no difficulty in the other two studies (Eilola & Havelka, 2010; Eilola et al., 2007). The inconsistent findings of semantic activation in the same paradigm of emotional Stroop could be attributed to two reasons. Fundamentally, the L2 speakers showing non–significant emotional Stroop effect in L2 in Winskel's study (2013) and the present study could be less proficient in their L2s than those showing significant effect in L2 in the studies of Eilola and Havelka (2010) and of Eilola et al. (2007). Indeed, L2 speakers in the present study and Winskel's study (2013) started learning their L2s at age 19.6 and age 7, respectively, in a formal setting (i.e., at school in their L1–speaking countries). In contrast, those in the studies of Eilola and Havelka (2010) and of Eilola et al. (2007) began learning at age 8 and age 9, either completely immersed in or partly exposed to L2 environments, respectively. The distinctive ages of L2 acquisition and L2 learning contexts among the L2 speaker groups imply that each group could be at different stages of L2 semantic development, consequently resulting in different degrees of semantic activation (Harris et al., 2006).

Alternatively, the different degrees of emotional involvement in L2 processing between the L2 speaker groups across the studies mentioned may be attributed to the orthographic differences between their L1s and L2s. Participants who appeared to activate emotional word meaning in L2 as automatically as L1 speakers had orthographically similar L1s and L2 (i.e., L1 Greek–L2 English in Eilola & Havelka, 2010 and L1 Finnish–

L2 English in Eilola et al., 2007), but those who did not had dissimilar L1s and L2s (i.e., L1 Thai–L2 English in Winskel, 2013 and L1 English–L2 Korean, for instance, in the present study). Different scripts may be much easier for L2 speakers to ignore their meaning activation when desirable, such as in an EST, and thereby for them to suppress its meaning activation better than that of similar scripts. In the same vein, Duyck and De Houwer (2008) pointed out the possibility that the extent to which semantic activation appears during L2 word recognition may depend on the similarity of the two writing systems (i.e., L1 German–L2 English speakers can show Simon effects better than those with different writing systems cannot). A future study can test the alternative with two different groups of bilinguals who have comparable proficiency in their L2s but different orthographic relations between their L1 and L2s on the EST.

6.1.2. Automatic Semantic Activation in L2 Sentence Reading

Contrary to the results in L2 word recognition, the present study showed clear evidence for comparable degrees of automaticity in semantic activation between L1 and late L2 speakers during sentential reading. This finding coincides with that of Hahn and Friederici's study (2001), in which late L1 Japanese–L2 German speakers were compared with L1 German speakers on N400 latency as they listened to semantically correct and incorrect sentences in German. The patterns of N400 produced by semantic anomalies were not significantly different between the L1 and L2 speaker groups in Hahn and Friederici's study (2001), suggesting that even late L2 speakers who started learning L2 around age 19 can activate and integrate semantic information during L2 sentential reading as automatically as L1 speakers.

Although some studies contrast with such comparable semantic activation during sentence reading between L1 and late L2 speakers (e.g., Hu & Jiang, 2011, Weber–Fox & Neville, 1996), the studies may have some points that are comparable with the present study even if other parts are in contrast to its findings. For example, Hu and Jiang (2011) also compared late L2 speakers with L1 speakers on a sentence-based lexical decision task. In this task, participants listened to a sentence that was missing the last word and saw a target word that was semantically predictable (congruent condition), not predictable but possible (neutral condition), or not possible (incongruent condition) to the sentence. Although the late L2 speakers' lexical decision patterns did not exactly resemble those of L1 speakers between the neutral and incongruent conditions, both groups' responses were significantly faster in the congruent condition than in the neutral condition (Hu & Jiang, 2011). L2 speakers could facilitate their lexical decision through the meaning of a sentence they just heard, similar to L1 speakers. This indicates that late L2 speakers may represent the sentence meaning while short listening as automatically as L1 speakers do, as shown when the target word was congruent with their current mental representation. Such a facilitative effect in Hu and Jiang's study (2011) adds partial supporting evidence to the comparable degrees of semantic activation during sentence processing between L1 and late L2 speakers as shown in the present study.

Similarly, in Weber–Fox and Neville's study (1996) late L2 speakers (age 16 as their age of L2 acquisition on average) displayed less accuracy and a later N400 peak in L2 sentence completion judgment (i.e., deciding whether a target word was semantically correct or incorrect to complete a sentence) than did L1 speakers, due to failure to reject semantic violations. That is, the L2 speakers were as accurate in deciding semantically correct words but not as accurate in deciding semantically incorrect words as L1 speakers. Taken together, late L2 speakers may activate L2 sentence meaning as automatically as L1 speakers, at least when it comes to semantically congruent sentences.

Moreover, in Weber–Fox and Neville's study (1996), late L2 speakers displaying less accuracy and later N400 peak in L2 sentence processing than L1 speakers were not advanced in their L2 proficiency; indeed, their self-rated comfort and proficiency in L2 were distinctively lower than those of other L2 groups ($ps < .05$) within the study. Thus, it is dubious whether such a low level of semantic activation in L2 sentence reading among the late L2 speakers would remain the same, even when their L2 proficiency become as high as the early L2 speakers displaying similar levels of automatic activation with L1 speakers in the study (Weber–Fox & Neville, 1996).

To summarize, the present study found positive evidence that late L2 speakers may activate L2 sentence meaning as automatically as L1 speakers. This finding is supported by previous studies examining late L2 speakers' semantic activation in L2 sentence processing with overall consistent (e.g., Hahn & Friederici, 2001) or partially consistent findings (e.g., Hu & Jiang, 2011). Moreover, late L2 speakers' deficit semantic activation in L2 sentence processing seems confounded with low levels of L2 proficiency, not necessarily only to the late exposure to L2 (e.g., Weber-Fox & Neville, 1996). The issue of what contributes to L2 speakers' different patterns between facilitation (which occurs in semantically correct contexts) and inhibition (which appears in semantically incorrect contexts) compared to L1 speakers as shown in Hu and Jiang (2011) and Weber-Fox and Neville (1996) calls for further research.

6.1.3. Different Automaticity in Semantic Representation between L2 Word and Sentence Processing

Given that late L2 speakers' semantic activation appeared to be comparable with L1 speakers in the second experiment (mental imagery generation in L2 sentence reading) but not in the first experiment (emotional involvement in L2 word recognition), how can we reconcile the inconsistent findings across the two experiments?

Apparently, these two experiments focused on different levels of L2 processing. Given that the processing level was "word" in Experiment 1 and "sentence" in Experiment 2, the two experiments focused on fundamentally different types of processing in which participants were engaged. Thus, the inconsistency across experiments in the present study can be contributed to primarily different mechanisms of the two processing levels. As mentioned, semantic activation in word recognition only needs bottom–up processing or decoding, whereas sentential reading may require additional processing such as semantic and syntactic integration and top–down processing. Therefore, we may presume that comprehending a L2 word is less demanding and more automatic than that of an L2 sentence. Indeed, previous studies showed this tendency according to the assumption that comparable semantic activation between L1 and L2 was generally seen when examining L2 lexical processing (Duñabeitia et al., 2010; Duyck & De Houwer, 2008; Finkbeiner & Nicol, 2003) but with some exceptions (e.g., Hahne & Friederici, 2001). The weaker semantic activation in L2 compared with L1 has typically been observed in studies looking at L2 sentence processing (e.g., Weber-Fox & Neville, 1996; Ardal et al., 1990).

The present study, however, challenges both the assumption and the tendency. The major difference between the present and previous studies is the different aspects used to examine an outcome of automatic semantic activation during L2 processing. Previous studies took multiple lexical routes to look at the degree of automaticity in semantic representation (e.g., semantic priming effects in translation), whereas the present study examined L2 speakers' mental representation according to word or sentence meaning. For L2 speakers to activate mental states corresponding to word meaning and sentence meaning, the more sources they could obtain from stimuli and the better they could show semantic activation in the present study. In this sense, additional processing load in sentence reading (e.g., syntactic information and semantic integration), which was absent in word processing, could be useful rather than taxing for L2 speakers to represent semantic information completely.

In the case of sentence processing, Kaup and colleagues (2012) suggested that while mental simulation can occur both based on word and sentence, the mental simulation automatically produced during sentence reading reflects a mixture of word– and sentence–based processes. In their study, L1 speakers generated mental imagery (i.e., match and mismatch effects) in seeing a mixture of words and nonwords in sequence (e.g., *paint brush/lorfing/finding/tempe/water mug/karumpe* versus *paint brush/lorfing/finding/tempe/paint box/karumpe*) as well as in reading two versions of sentences with the same structure but one critical word different (e.g., *Mary finds the paint* *brush in the studio* <u>*in the water mug*</u> versus *Mary finds the paint brush in the studio* <u>*in the paint box*</u>). To produce a match and mismatch effect after reading the sentence, the participants had to process sentence structure and integrate meanings from individual words. On the other hand, generating mental imagery in seeing word sequence necessitated sole dependency on a decoding process of an individual word. Interestingly, the match and mismatch effects that participants produced in Kaup et al.'s study (2011) were larger in sentence reading (76 ms) than in word sequence reading (74 ms), implying that sentence–based simulation may be equally or more automatic compared to word–based simulation. [22] If the mechanism of sentence–based simulation is applicable to L2 processing as well, the inconsistent finding across processing levels in the present study can be explained by the greater feasibility of sentence–based simulation than that of word–based simulation.

Another possible explanation for different levels of semantic activation in the present study is cross-task variation. The tasks adopted in Experiments 1 and 2 differed in terms of the extent that semantic activation of late L2 learners was constrained or promoted. Under the first experimental paradigm, participants were asked to focus on the color in which a word was presented. Thus, word meaning, which was task–irrelevant was likely to be constrained rather than promoted.

Unlike the EST, however, the SPT used in the second experiment may provide a more optimal environment for late L2 speakers to maximally process L2 sentence meaning. Because participants were asked to judge a picture based on a previous sentence, semantic processing was essential to successfully conduct the SPT. Moreover, the SPT encouraged

[22] The significance of 2 ms difference was not indicated in Kaup et al.'s study (2011).

respondents to fully understand the meaning of the whole sentence through self–paced reading and were even asked random questions to check their comprehension.

Taken together, an EST can be considered more conservative than an SPT in terms of the extent to which participants should process semantic information. In this sense, L2 speakers' different degrees of semantic activation across tasks in the present study may be related to the dissimilar extents of which the two tasks promoted to involve semantic processing. Indeed, the degrees of semantic activation in L2 speakers tend to depend on tasks in which their semantic processing was observed in previous studies. For example, Grainger and Frenck–Mestre (1998) showed stronger translation priming effects in the semantic categorization task than in the lexical decision task using the same stimuli. This is because semantic categorization evokes intensive semantic processing, whereas lexical decision can be made at the early stage of visual word processing without deeply involving semantic information.

If this is the case, why did only L2 speakers show different patterns according to the tasks, unlike L1 speakers in the present study? In other words, what caused L2 speakers to disregard word meaning more effectively than L1 speakers in an EST? First, it might be simply that L2 speakers had limited access to word meaning with lower proficiency in Korean than L1 speakers. In this regard, less proficient language tends to be suppressed more effectively than more proficient language; indeed, the level of semantic activation in the former is lower than that in the latter. As such, relatively less effort is needed to suppress it (Green's Inhibitory Control Model, 1998). Thus, given that L2 Korean speakers' proficiency in Korean was much lower than L1 Korean speakers', it is possible that the L2 speakers could suppress Korean word meaning more easily than L1 speakers in the present study.

On the other hand, it is also possible that semantic activation in L2 speakers was as strong as in L1 speakers in the EST (as shown in the SPT), but L2 speakers were better able to suppress word meaning than L1 speakers' when it was desirable, such as in the EST. Considering the bilingual advantage in cognitive control in general (e.g., Bialystok et al., 2008) and no data about L1 Korean speakers' bilingual status in the present study, we cannot completely rule out this alternative explanation for the absence of L2 speakers' automatic semantic activation in L2 word recognition compared to L1 speakers. It will be interesting to see if a task requiring word meaning to be processed rather than ignored produces different results in a future study. Also, future research should test the alternative with manipulating the bilingual status of participants in order to exclude or adopt the alternative of the bilingual advantage.

In conclusion, the present study showed that late L2 speakers can activate semantic representation in L2 processing as automatically as do L1 speakers, provided an optimal environment is given with regards to processing level, task demand, and stimulus word type. Thus, the question becomes which individual factors affect the degrees of automaticity in semantic representation in L2 processing? This is discussed in the next section.

6.2. Factors Contributing to Automatic Semantic Activation in L2 Processing

The second purpose of the present study was to identify factors that contribute to developing automatic semantic activation in late L2 speakers. Among many potential predictors, the two chosen for the present study are L2 proficiency level and accumulated amount of L2 use. These are not only crucial variables for developing a semantic system and automaticity (e.g., see Jiang, 2000 on L2 proficiency and De Keyser, 2001 on L2 use) but are also the confounding factors most commonly found in the previous literature.

In regard to L2 proficiency, the present study found no statistically significant effects either in word or in sentence processing. This contrasts with the findings of previous studies. For example, Zhao et al. (2011) divided L2 speakers into two groups based on their L2 proficiency and found a significant difference in semantic priming effect between low– and high–proficiency L2 groups during L2 word recognition. This result is relevant to the Developmental Hypothesis (de Groot, 1995), which posits that initially weak semantic–conceptual links of L2 words strengthen with increased L2 proficiency. Indeed, in Duñabeitia et al.'s study (2010) almost balanced L1 Basque–L2 Spanish speakers who were likely to reach native-like semantic development in L2 reported comparable levels of a non-cognate priming effect in both the L1–L2 and L2–L1 directions.

Likewise, in Weber–Fox and Neville's study (1996) more or less proficient bilinguals (although confounded by their ages of L2 acquisition) were compared on semantic accuracy and N400 latency during L2 sentence reading and exhibited different degrees of semantic activation according to their L2 proficiency. That is, the bilinguals who rated themselves as high (above 3 out of 4) showed similar patterns in accuracy and N400 pattern with L1 speakers', whereas those with self-ratings below 3 displayed significantly distinctive patterns compared to L1 speakers and more proficient L2 groups. To the best of our knowledge, no known model explains the relationship between semantic activation in

L2 sentence processing and L2 proficiency. However, the findings of the previous studies (e.g., Weber–Fox & Neville, 1996) clearly suggest that L2 proficiency may mediate semantic activation level in L2 sentence processing as well.

Why, then, did the present study fail to reveal the potentially important role of L2 proficiency in semantic activation during L2 processing? It is less likely to be a matter of the current proficiency measurement adopted in the present study. In fact, the Korean C–test allowed the present study to overcome the limitation of self–reporting (e.g., subjective and less reliable), which is frequently used in the previous studies. [23]

Rather, it seems related to little variance in this regard among the current sample. As discussed in the previous chapter, the L2 proficiency levels of participants in the present study were relatively homogenous, because they were already defined as advanced L2 learners.[24] With such relatively small variances within the L2 proficiency variable, its effect is not likely to appear significant. If the present study included low–proficiency L2 Korean speakers, the L2 proficiency effect could be significant as in previous studies. For example, Ahn and colleagues (2012) reflecting greater variance did show a significant effect of L2 proficiency on semantic activation. Specifically, advanced L2 English speakers displayed a significant sentence–picture congruency effect in L2 sentence reading similar to L1 speakers, whereas intermediate L2 speakers did not. This indicates that L2 proficiency may play a critical role in developing automatic semantic activation in L2 processing. Thus,

[23] If the late L2 Korean speakers' scores were kept as by items, the reliability could have been checked to see whether it can provide a further alternative explanation for the lack of the effect of the L2 proficiency in the present study. For the future research, it is recommended to calculate the reliability of the test to see if the test reliably measured the given participants' L2 proficiency.

[24] Note that this narrower range of L2 use resulted from cutting 34 L2 speakers who were not defined as advanced due to their low scores (<60%) from the original pool (*N*=94).

despite the non–significance of the L2 proficiency variable in the present study, it cannot be argued that L2 proficiency does not contribute to the level of automatic semantic activation in L2 processing.

Unlike the case of L2 proficiency, the present study found L2 use played a significant role in semantic activation both in L2 word and sentence processing. L2 use could appear to be significant more so than L2 proficiency in the present study for two reasons. First, given that the present study focused on how L2 speakers map linguistic input to its real–world referent (i.e., word meaning to emotional involvement and sentence meaning to generating mental imagery) compared to L1 speakers, how extensively they have used L2 in their daily life (other than in a classroom) may be more related to the results than L2 proficiency. Participants were asked about their amount of L2 use in daily life, and these responses were calculated in Excel files as detailed as possible. Their amount of L2 use was calculated based on daily use in an automatic manner such as online chatting, face-to-face interaction, watching television or movies, and so on. The calculation was based on a combination of the age at which they begin studying the L2, hours per week L2 was used, and how many months were spent on each situation mentioned above. With more direct and detailed measurement of L2 use, the accumulated amount of L2 use could influence automatic semantic activation more directly than the other variable.

On the other hand, variance in the amount of L2 use was sufficient enough in the present study to reveal its significant role in semantic activation in L2 speakers. To reiterate, the accumulated amount of using L2 Korean varied from 116 to 24,848 hours with a mean of 4,168 hours (*SD* 5043.8). The results indicated that the accumulated amount of L2 use significantly affected the degree of emotional involvement, as well as of mental imagery among L2 speakers in the present study. Based on the findings, the present study suggests that more extensive L2 use in daily life is associated with a greater likelihood that L2 speakers will comprehend L2 words and sentences as automatically as do L1 speakers.

Few studies directly examine how L2 use affects semantic activation in L2, except one. Degner et al.'s study (2012) investigated L1 German–L2 French and L1 French–L2 German speakers' intensity of L1 and L2 use and its role in affective priming effects both in their L1s and L2s. The two groups matched each other in terms of their L2 proficiency, age of L2 acquisition, and length of residence in the L2 speaking–countries, only differing in the frequency of L1 and L2 use. That is, L2 French speakers were using their L2 significantly less than their L1, whereas L2 German speakers were using their L2 equally as frequently with their L1 in daily life. As a result, affective priming effects appeared significant in all cases, except in L2 French. Thus, the evidence from Degner et al.'s study (2012) combined with that from the present study suggests that the degree of automaticity in semantic representation during L2 processing may reflect the extent to which L2 speakers have used their L2 in daily life.

In conclusion, L2 speakers' automatic semantic activation in L2 is likely to be mediated by how proficient they are in their L2 and how much they have used the language in their daily lives. The present study and previous studies show that late L2 speakers' semantic activation may strengthen to a comparable level with L1 speakers given extensive

L2 use and high proficiency levels, even if the individual began learning their second language after puberty. [25]

[25] Note that the similar patterns between L2 and L1 groups in the present study should be interpreted with two limitations in mind: first, findings are based on the null hypothesis testing; second, two tasks showing the patterns could be less or more conservative than others. Therefore, the L2 group's similar performance to L1 speakers' on either of the two experiments described in the present study does not necessarily reflect a native–like or near–native performance among the L2 group in general.

Chapter 7. Conclusion

The present study replicated the finding that emotional involvement and mental imagery automatically occur during L1 reading, with L2 reading for comparing automatic levels of semantic activation in L1 and L2 processing. By examining emotional involvement and mental imagery in L2 speakers, which appear when they can map linguistic input to its corresponding mental state or real–world referent automatically (as do L1 speakers), the present study addressed to what extent semantic representation is activated automatically during L2 processing from the new angles. Indeed, most of the previous literature has examined semantic activation through multiple lexical routes, where semantic representation is confounded with non–semantic (e.g., lexical) representation.

In addition, the present study observed emotional involvement and mental imagery in L2 speakers according to their proficiency level in L2 and accumulated amount of L2 use. This observation revealed how automatic semantic activation develops among late L2 speakers and whether they can reach a comparable level of automaticity in semantic processing as do L1 speakers when assisted with high L2 proficiency and extensive L2 use.

Considering the significant effect that onset age of L2 exposure plays in an L2 speaker's ultimate attainment, as well as its presumably reduced effects on an L2's semantic aspects, the present study found positive clues that it may be possible for late L2 speakers to obtain native–like automatic semantic activation. Given that the target language (Korean) of the present study is one of the less commonly taught languages in the US (i.e., individuals are less likely to be exposed to Korean in the daily life of an L1 environment), the present study could control the accidental L2 exposure effect and examine the relationship between semantic processing and L2 use more clearly. Given that late L2 Korean speakers showed automatic semantic activation similarly to L1 speakers (at least in sentence processing), even when possible L2 exposure was minimal, this finding can be generalized (as a baseline) to other cases of late L2 learners who are likely to be exposed to their L2 as much as or more than those in the present study.

Furthermore, in the present study, both lexical and sentential levels were considered to examine semantic activation. Most previous studies have approached the issue only at one of the processing levels, making it challenging to draw any conclusions across studies. The present study, however, incorporated both approaches through emotional involvement in word processing and mental imagery in sentence processing within a single study. As such, it was possible to investigate semantic activation in L2 processing from diverse aspects.

In addition, to date, the present study is the first to apply a mental imagery paradigm to L2 processing. So far, accumulated research based on embodied cognition has been conducted exclusively in L1 processing and has gathered extensive evidence that meaning representation may automatically evoke mental simulation. The findings of the present study widen the scope of applying such an argument. Given that late L2 speakers also showed mental imagery similar to L1 speakers' in the present study, generating mental simulation along with semantic representation may not be limited to L1 processing. Rather, the present study's findings may provide some clues regarding the nature of semantic representation in the human mind. Ultimately, the present study answered the question of whether late L2 speakers can immediately understand, visualize, and feel what a word or a sentence denotes during L2 reading as well as do L1 speakers.

Lastly, the present study offers two major and one minor suggestions for future research and one pedagogical emphasis for second language learning related to late L2 speakers' automatic semantic activation.

First, further research is needed to examine emotional involvement in less conservative tasks than the EST for semantic processing. As pointed out in Chapter 6, the absence of L2 speakers' emotional involvement might be due to the characteristic of the EST, which can be performed more effectively without semantic processing. Therefore, the current result is confounded with L2 speakers' better inhibitory control in general, not necessarily only due to their less semantic activation in L2. By showing or not showing L2 speakers' emotional involvement in a task that requires semantic processing to complete, the result would provide clearer evidence of whether late L2 speakers can activate emotional connotation in recognizing L2 words as automatically as do L1 speakers.

In addition, a follow-up study of the Experiment 2 may test late-advanced L2 speakers in a more automatic manner, such as timed reading instead of self–paced, to determine to what level of automaticity respondents exhibit semantic activation during L2 processing comparable with L1 speakers. In Experiment 2 of the present study, late L2 speakers had enough time to comprehend a sentence before obtaining a picture and displayed similar degrees of automaticity in semantic representation to L1 speakers. Thus, the current finding on a self–paced reading condition may have limitations to address regarding to what extent late L2 speakers can activate semantic representation during L2 sentence reading compared to L1 speakers. In a future study, however, if L2 speakers show the sentence–picture congruency effects in speed reading (or speed listening if using auditory stimuli) to similar degrees with L1 speakers, this finding would indicate that late L2 speakers' automatic level of semantic activation may be higher than that revealed in the present study. Otherwise, this finding would suggest that late L2 speakers with advanced L2 proficiency may comprehend L2 sentences as completely as L1 speakers, but only when enough time is given. As such, a future study that manipulates time for comprehension will reveal valuable insights into the question of whether late L2 speakers can reach native–like automatic levels of semantic activation in L2 processing.

Furthermore, as a minor methodological revision, a future study should consider using two pictures with two sentences for the SPT. In the present study, critical items used for this task were created using a combination of two versions of the sentences per picture. Both sentences were intended to trigger *yes* responses by mentioning the same object as shown in a picture. One sentence of the two versions, however, was matched and the other was mismatched with an object in a picture in terms of shape. For example, prior to the image of an opened book placed upside down (see Figure 5), either one of the two sentences were presented: 교수님이 책상 위에 책을 놓아요 (*The professor placed the book on his desk*) for a match condition, or 교수님이 가방 안에 책을 넣어요 (*The professor put the book into his bag*) for a mismatch condition. The matched sentences were expected to trigger *yes* responses more quickly than the mismatched sentences. The way to manipulate stimuli in the present study, however, may be limited in terms of potential idiosyncratic characteristics of the picture or the sentence.[26]

Alternatively, a future study can manipulate stimuli through the technique used by Stanfield and Zwaan (2002) in which two pictures were used for two sentences.[27] Employing this method to create stimulus, a future study may add a new image and pair it with two original sentences, resulting in four pairs (two versions of matched pairs and two versions of mismatched pairs) for each item with two versions of the sentence and two versions of the pictures crossing each other.

Finally, the present study reveals an important finding for second language learning: more exclusive use of L2 may increase the likelihood that L2 speakers will show emotional involvement and mental imagery at a level of automatic activation comparable to L1 speakers. The similar degrees of automatic semantic activation between L1 and L2 processing imply that comprehension in L2 may be as efficient as in L1. In other words, even individuals who began learning L2 after puberty may be able to reach native–like efficient comprehension in L2 with extensive and regular use. Some ERP studies reported consistent findings with that of the present study. For example, in Hahne and Friederici's (2001) study, late L2 German speakers showed similar patterns of N400 with L1 German speakers when reading semantically incorrect sentences in German. This implies that

[26] This limitation, however, is not likely to affect the results of the present study, which examines relative levels of automatic semantic processing between L1 and L2 speakers (the same sentence–picture pair was applied to the two groups). Moreover, the original study using two pictures for two sentences found no significant interaction between condition effects and lists (Zwaan et al., 2002), implying that the way to pair stimuli in the present study may not significantly differ in terms of producing match and mismatch effects.

[27] Note that the purpose of Stanfield and Zwaan's study (2002) was to examine whether sentence meaning was represented as an image in L1 speakers. Therefore, the idiosyncratic issue of stimuli could be more important in their study.

semantic aspects may be the least vulnerable to the age at which an individual begins to acquire L2. Taken together, proficiency in a second language coupled with extensive use may enable L2 speakers to immediately understand, visualize, and perceive the meaning of a word or sentence they have read as automatically as do L1 speakers, regardless of the initial age of L2 exposure. Clearly, therefore, when learning a second language, it is important to use it extensively in daily life.

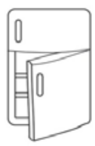
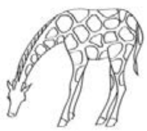
Appendices

Appendix 1. Stimuli for the Emotional Involvement (Experiment 1)

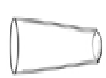
Emotional Words	Neutral Words
시끄럽다	운전하다
실수	빠르다
싫어하다	요리
심심하다	노래하다
어렵다	운동하다
외롭다	음식점
울다	마시다
교통사고	전화하다
잃어버리다	가족
죄송하다	은행
걱정	공부하다
틀리다	조용하다
헤어지다	날씨
힘들다	같다
불편	다르다
아프다	작다
비싸다	여행하다
실망	주말
피로	이야기
지치다*	아침
떨어지다	가볍다
나쁘다	가깝다
모르다	비슷하다
때리다	깊다
무서움	건물
미워하다	안녕하다
바쁘다	걷다
불행	빨래하다
슬픔	기다리다
거짓말하다	설거지*

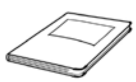
Appendix 2. Critical Items for the Mental Imagery (Experiment 2)

	Image	Congruent sentences	Incongruent sentences
1		아이가 나무에 있는 새를 봐요.	아이가 하늘에 있는 새를 봐요.
2		훈이는 필통에 연필을 넣어요.	훈이는 컵에 연필을 넣어요.
3		아저씨가 열쇠로 문을 열어요.	아저씨가 벽에 열쇠를 걸어요.
4		친구가 하늘을 나는 큰 새를 봐요.	친구가 물 위에 있는 큰 새를 봐요.
5		미영이가 접시에 있는 계란을 봐요.	미영이가 상자 안에 있는 계란을 봐요.
6		철수가 가게에서 양파를 사요.	철수가 음식에 양파를 조금 넣어요.
7		아줌마가 아이에게 바나나를 먹여요.	아줌마가 손님에게 바나나를 팔아요.
8		손님이 가게에서 새 안경을 써요.	손님이 가방에 안경을 넣어요.
9		교수님이 복사기 위에 책을 놓아요.	교수님이 가방 안에 책을 넣어요.

10		그 남자가 옷걸이에 모자를 걸어요.	그 남자가 식탁 위에 모자를 놓아요.
11		그 여자가 벽에 있는 거미를 보고 놀랐어요.	그 여자가 바닥에 있는 거미를 보고 놀랐어요.
12		친구가 냉장고에서 음료수를 찾아요.	친구가 냉장고 위에 사진을 붙여요.
13		기린이 땅 위의 풀을 먹어요.	기린이 나무 위의 잎을 먹어요.
14*		물 속에서 물고기가 놀고 있어요.	막대기에 물고기가 매달려 있어요.
15		소녀가 우산을 쓰고 있어요.	소녀가우산을가방에넣어요.
16		운전사가 버스에서 사람들을 기다려요.	운전사가 버스를 빨리 운전해요.
17		동생이 케이크를 잘라요.	동생이 케이크를 사요.
18		진이가 이사하려고 박스에 물건을 넣어요.	진이가 이사하기 위해 박스를 쌓아요.
19		빨래통에 셔츠가 있어요.	옷걸이에 셔츠가 있어요.

20		아저씨가 드릴로 바닥을 고쳐요.	아저씨가 드릴로 문을 고쳐요.
21		그 손님은 소금통으로 음식에 소금을 넣어요.	그 손님은 소금통을 음식 옆에 놓아요.
22		그가 휴지통에 음료수 캔을 버려요.	그가 가게에서 음료수 캔을 사요.
23		그 웨이터는 와인잔을 천장에 걸어요	그 웨이터는 와인잔에 와인을 부어요.
24		아영이가 칫솔에 치약을 발라요.	아영이가 여행가방에 치약을 넣어요.
25		그 소녀가 숟가락으로 국을 먹어요.	그 소녀가 숟가락을 식탁 위에 놓아요.
26		나는 선생님에게 편지를 보내요.	선생님에게 편지를 쓰고 있어요.
27		친구가 노트북으로 숙제를 해요.	친구가 노트북을 가방에 넣어요.
28		그녀가 새 펜으로 일기를 써요.	그녀가 새 펜을 필통에 넣어요.
29		그가 화가 나서 의자를 던져요.	그가 피곤해서 의자에 앉아요.

30		그가 커피잔을 벽에 걸어요.	그가 커피잔을 식탁 위에 놓아요.
31		농부가 사과를 맛있게 먹어요.	농부가 사과를 팔아요.
32		거북이가 개를 보고 무서워해요.	거북이가 바다를 보고 바다로 가요.
33		원숭이가 나무에서 놀고 있어요.	원숭이가 바닥에서 먹고 있어요.
34		개구리가 계단을 올라가요.	개구리가 계단 위에서 쉬어요.
35		아기 곰이 나무 위로 올라가요.	아기 곰이 사람들에게 가요.
36		나는 큰 초를 창고에 넣어요.	나는 밝은 초를 식탁 위에 놓아요.
37		나는 컵을 바닥에 떨어뜨렸어요.	나는 컵을 식탁에 놓아요.
38		친구가 새 책가방을 메고 있어요.	친구가 책을 새 책가방에 넣고 있어요.
39		그 개가 나를 보고 와요.	그 개가 나와 같이 자요.
40		아빠가 수박을 먹어요.	아빠가 수박을 사요.

41		동생이 자전거를 타고 학교에 가요.	동생이 자전거에서 넘어졌어요.
42		그는 새 바지를 햇빛에 말려요.	그는 새 바지를 가방에 넣어요.
43		날씨가 너무 더워 꽃도 나무도 물이 필요해요.	날씨가 좋아 꽃도 나무도 잘 자라요.
44*		소녀가 부츠를 벗고 방에 들어가요.	소녀가 부츠를 신고 거리를 걸어요.
45		그 학생은 공책을 선생님에게 줘요.	그 학생은 공책에 열심히 쓰고 있어요.
46		고양이가 나와 같이 텔레비전을 봐요.	고양이가 나를 따라와요.
47		주전자를 불 위에 올려놓아요.	주전자 안에 물을 부어요.
48		바람이 조금 불어 배가 천천히 가요.	바람이 많이 불어 배가 쓰러져요.

Note: Two deleted items with the asterisk mark (*) were deleted from analyses given that the L2 speakers checked them as unknown in the post-vocabulary check list.

References

Ahn, S., Jiang, N, & Osthus, Peter. (2012). *Mental imagery in second language processing*. Paper presented at the Second Language Research Forum 2012, Pittsburgh, PA.

Ahn, S. (2013a). *Conceptual attrition: Emotional intensity in Korean heritage speakers*. Poster presented at the National Heritage Language Institute 2013, Chicago, IL.

Ahn, S. (2013b). *Can you feel in L1 as automatically as in L2? Automatic activation of emotional word meaning*. Paper presented at the Second Language Research Forum 2013, Provo, UT.

Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional Stroop phenomenon: A generic slowdown, not a Stroop effect. *Journal of Experimental Psychology: General*, *133*(3), 323–338.

Altarriba, J., & Basnight-Brown, D. M. (2008). Ch. 4 An overview of semantic processing in bilinguals: Methods and findings. In *The Bilingual Mental Lexicon: Interdisciplinary Approaches* (pp. 79–98). Bristol • Buffalo • Toronto: Multilingual Matters.

Altarriba, J. & Bauer, L. M. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. *American Journal of Psychology*, *117*, 389-410.

Altarriba, J. (2003) Does carino equal 'liking'? A theoretical approach to conceptual nonequivalence between languages. *International Journal of Bilingualism 7*, 305-322.

Alvarez, R. P., Holcomb, P. J., & Grainger, J. (2003). Accessing word meaning in two languages: An event-related brain potential study of beginning bilinguals. *Brain and Language*, *87*(2), 290–304.

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463–498.

Anooshian, L. J., & Hertel, P. T. (1994). Emotionality in free recall: Language specificity in bilingual memory. *Cognition & Emotion*, *8*(6), 503–514.

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., & Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and Language*, *39*(2), 187–205.

Ayçiçegi, A., & Harris, C. (2004). Bilinguals' recall and recognition of emotion words. *Cognition and Emotion*, *18*(7), 977–987.

Ayçiçegi-Dinn, A., & Caldwell-Harris, C. L. (2009). Emotion-memory effects in bilingual speakers: A levels-of-processing approach. *Bilingualism: Language and Cognition*, *12*(03), 291–303.

Baayen, R. H. (2008). Analyzing *linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.

Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–660.

Barrett, L. F. (2006). Valence is a basic building block of emotional life. *Journal of Research in Personality 40*, 35–55.
Basnight–Brown, D. M., & Altarriba, J. (2007). Differences in semantic and translation priming across languages: The role of language direction and language dominance. *Memory & Cognition*, *35*(5), 953–965.
Bates, D. M. (2005). Fitting linear mixed models in R. *R News*, *5*, 27–30.
Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 859–873.
Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, UK: Wiley.
Chee, M. W. L., Hon, N., Lee, H. L., & Soon, C. S. (2001). Relative language proficiency modulates BOLD signal change when bilinguals perform semantic judgments. *NeuroImage*, *13*(6). 1155–1163.
Cho, Y-M, Lee, H. S., Schulz, C., Shon, H-M, & Sohn, S. (2000a). *Integrated Korean: Beginning 1*. Honolulu, University of Hawai'I Press.
Cho, Y-M, Lee, H. S., Schulz, C., Shon, H-M, & Sohn, S. (2000b). *Integrated Korean: Beginning 2*. Honolulu, University of Hawai'I Press.
Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic priming. *Psychological Review, 82*, 407-428.
Coppens, L. C., Gootjes, L., & Zwaan, R. A. (2012). Incidental picture exposure affects later reading: evidence from the N400. *Brain and Language*, *122*(1), 64–69.
Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research, 28*, 369–382.
Degner, J., Doycheva, C., & Wentura, D. (2012). It matters how much you talk: On the automaticity of affective connotations of first and second language words. *Bilingualism: Language and Cognition 15*(1), 181–189.
De Groot, A. M. (1995). Determinants of bilingual lexicosemantic organization. *Computer Assisted Language Learning, 8*, 151–180.
De Groot, A. M. B. (1992). Bilingual lexical representation: A closer look at conceptual representations. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 389–412). Amsterdam: Elsevier Science Publishers.
De Groot, A.M.B., & Nas, G.L. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language, 30*, 90–123.
DeKeyser, R. M. (2012). Age effects in second language learning. In S. Gass & A. Mackey (Eds.), *Handbook of Second Language Acquisition* (pp. 442–460). London: Routledge.
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). Oxford: Blackwell.
Dekeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.) *Cognition and Second Language Instruction* (pp. 125–151). NY: Cambridge University Press.

Dewaele, J.–M. (2004). Perceived language dominance and language preference for emotional speech: The implications for attrition research. In M.S. Schmid, B. Köpke, M. Kejser & L. Weilemar (Eds.) *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*. Amsterdam: John Benjamins, pp. 81–104.

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science*, *298*(5596), 1191–1194.

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2010). Masked Translation Priming Effects With Highly Proficient Simultaneous Bilinguals. *Experimental Psychology (formerly "Zeitschrift Für Experimentelle Psychologie"), 57*(2), 98–107.

Duyck, W., & Warlop, N. (2009). Translation priming between the native language and a second language: new evidence from Dutch-French bilinguals. *Experimental Psychology*, *56*(3), 173–179.

Duyck, W., & Brysbaert, M. (2008). Semantic Access in Number Word Translation. *Experimental Psychology (formerly "Zeitschrift Für Experimentelle Psychologie"), 55*(2), 102–112.

Duyck, W., & Brysbaert, M. (2004). Forward and backward number translation requires conceptual mediation in both balanced and unbalanced bilinguals. *Journal of Experimental Psychology: Human Perception and Performance, 30*(5), 889–906.

Duyck, W., & De Houwer, J. (2008). Semantic access in second-language visual word processing: evidence from the semantic Simon paradigm. *Psychonomic Bulletin & Review*, *15*(5), 961–966.

Eide, P. , Kemp, A., Silberstein, R.B., Nathan, P.J., & Stough, C. (2002). Test–retest reliability of the emotional Stroop task: examining the paradox of measurement change. *Journal of Psychology, 136*(5), 514–20.

Eilola, T. M., & Havelka, J. (2010). Behavioural and physiological responses to the emotional and taboo Stroop tasks in native and non-native speakers of English. *International Journal of Bilingualism*, *15*(3), 353–369.

Eilola, T. M., Havelka, J., & Sharma, D. (2007). Emotional activation in the first and second language. *Cognition and Emotion, 21*, 1064–1076.

Estes, Z, & Adelman, J. S. (2008). Automatic vigilance for negative words in lexical decision and naming: Comment on Larsen, Mercer, and Balota (2006). *Emotion, 8*(4), 441– 444.

Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics*, *24*(3), 369–383.

Gaillard, R., Naccache, L., Pinel, P., Clémenceau, S., Volle, E., Hasboun, D., Dupont, S., Baulac, M., Dehaene, S., Adam, C., & Cohen, L. (2006). Direct intracranial, fMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron, 50*(2), 191–204.

Grabe, W. (1991). Current developments in second language reading research. *Tesol Quarterly*, *25*(3), 375–406.

Grainger, J., & Frenck–Mestre, C. (1998). Masked priming translation equivalents in proficient bilinguals. *Language and Cognitive Processes, 13*, 601–623.

Green, D. (1998). Mental control of the bilingual lexico–semantic system. *Bilingualism: Language and Cognition, 1*, 67–81.

Hahne, A., & Friederici, A. D. (2001). Processing a second language: late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, *4*(2), 123–141.

Harris, C. L., Gleason, J. B., & Ayçiçegi, A. (2006). When is a First Language More Emotional? Psychophysiological Evidence from Bilingual Speakers. In A. Pavlenko (Ed.), *Bilingual minds: Emotional experience, expression, and representation* (pp. 257-283). Clevedon, United Kingdom: Mulilingual Matters.

Harris, C. (2004) Bilingual speakers in the lab: Psychophysiological measures of emotional reactivity. *Journal of Multilingual and Multicultural Development, 25*, 223–47.

Harris, C., Ayc¸ic¸egi, A., & Gleason, J. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first than in a second language. *Applied Psycholinguistics, 4*, 561–78.

Hirschfeld, G., Feldker, K., & Zwitserlood, P. (2012). Listening to "flying ducks": Individual differences in sentence-picture verification investigated with ERPs: ERPs and individual differences in imagery. *Psychophysiology*, *49*(3), 312–321.

Hu, G., & Jiang, N. (2011). Semantic integration in listening comprehension in a second language: Evidence from cross-modal priming. In P. Trofimovich & K. McDonough (Eds.), *Applying priming methods to L2 learning, teaching and research* (pp. 199–218). Amsterdam, Philadelphia: John Benjamins Publishing Company.

ILR reading skill scale (2014). Interagency language roundtable language skill level decription. Retrieved Febrary 27, 2014 from http://www.govtilr.org/Skills/ILRscale4.htm.

Jay, T., Caldwell-Harris, C. L. & King, K. (2008). Recalling taboo and non-taboo words. *American Journal of Psychology*, *121*, 83–103.

Jiang, N. (2012). *Conducting Reaction Time Research in Second Language Studies*. NY: Routledge.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language acquisition. *Language Learning, 57,* 1-33.

Jiang, N., & Forster, K. I. (2001). Cross-Language Priming Asymmetries in Lexical Decision and Episodic Recognition. *Journal of Memory and Language*, *44*(1), 32–51.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*(1), 47–77.

Jiang, N. (1999). Testing processing explanations for the asymmetry in masked cross-language priming. *Bilingualism: Language and Cognition*, *2*(1), 59–75.

Jin, Y.–S. (1990). Effects of concreteness on cross–language priming in lexical decisions. *Perceptual and Motor Skills, 70*, 1139–1154.

Kaup, B., Lüdtke, J., & Steiner, I. (2012). Word– vs. sentence–based simulation effects in language comprehension. In B. Stolterfoht & S. Featherston (Eds.), *Empirical Approaches to Linguistic Theory: Studies in Meaning and Structure*. (pp. 267–290). De Gruyter Mouton.

Kroll, J. F., & de Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives.* (pp. 201–224). Mahwah, NJ: Lawrence Erlbaum Publishers.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149–174.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Advancement of Science*, *207*(*4427*), 203–205.

Larsen, R. J., Mercer, K. A., & Balota, D. A. (2006). Lexical characteristics of words used in emotional Stroop experiments. *Emotion*, *6*(1), 62.

Larsen, Randy J., Mercer, K. A., Balota, D. A., & Strube, M. J. (2008). Not all negative words slow down lexical decision and naming speed: Importance of word arousal. *Emotion*, *8*(4), 445–452.

Lee–Ellis, S. (2009). The development and validation of a Korean C-Test using Rasch Analysis. *Language Testing*, *26*(2), 245–274.

Lee–Ellis, S. (2012). *Looking into Bilingualism through the heritage speaker's mind*. PhD Dissertation. University of Maryland.

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, *7*(4), 618–630.

MacKay, D.G., & Ahmetzanov, M, V. (2005). Emotion, memory, and attention in the taboo Stroop paradigm: An experimental analog of flashbulb memories. *Psychological Science, 16*, 25-32.

McKenna, F. P., & Sharma, D. (2004). Reversing the emotional Stroop effect: The role of fast and slow components. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*, 382–392.

Meteyard, L., Zokaei, N., Bahrami, B. and Vigliocco, G. (2008). Visual motion interferes with lexical decision on motion words. *Current Biology, 18*(17). R732–R733.

Naccache, L., Gaillard, R., Adam, C., Hasboun, D., Clémenceau, S., Baulac, M., Dehaene, S., & Cohen, L. (2005). A direct intracranial record of emotions evoked by subliminal words. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(21), 7713–7317.

Öhman, A., & Mineka, S. (2001). Fear, phobias and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*, 483–522.

Pavlenko, A. (2008) Structural and conceptual equivalence in acquisition and use of emotion words in a second language. *Mental Lexicon, 3*(1), 91–120.

Pavlenko, A. (2005) *Emotions and multilingualism*. Cambridge University Press.

Pecher, Diane, & Rolf A. Zwaan. (2005). *Grounding of cognition: The role of perception and action in memory, language, and thinking*. Cambridge University Press.

Pecher, D., Van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *Quarterly Journal of Experimental Psychology*, *62*, 1108–1114.

Pecher, D., Zeelenberg, R., & Raaijmakers, J. G. W. (1998). Does pizza prime coin? Perceptual priming in lexical decision and pronunciation. *Journal of Memory & Language, 38*, 401–418.
Rogers, T. T. and McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT Press, Cambridge, MA.
Rugg, M. D. (1995). ERP studies of memory. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event–related brain potentials and cognition* (pp. 132–170). Oxford: Oxford University Press.
Segalowitz, N., Trofimovich, P., Gatbonton, E., & Sokolovskaya, A. (2008). Feeling affect in a second language: The role of word recognition automaticity. *Mental Lexicon, 3*, 47–71.
Sholl, A., Sankaranarayanan, A., & Kroll, J. F. (1995). Transfer between picture naming and translation: A test of asymmetries in bilingual memory. *Psychological Science, 6*(1), 45–49.
Slabakova, R. (2006). Is there a critical period for semantics? *Second Language Research, 22*, 1–37.
Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*, 174–215.
Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156.
Strauss, G. P., Allen, D. N., Jorgensen, M. L., & Cramer, S. L. (2005). Test-retest reliability of standard and emotional Stroop tasks: an investigation of color-word and picture-word versions. *Assessment, 12*(3). 330-337.
Sutton, T. M., Altarriba, J., Gianico, J. L., & Basnight-Brown, D. M. (2007). The automatic access of emotion: Emotional Stroop effects in Spanish–English bilingual speakers. *Cognition & Emotion*, *21*(5), 1077–1090.
Taylor, L. J., & Zwaan, R. A. (2009). Action in cognition: The case of language. *Language & Cognition*, *1*(1), 45–58.
Vigliocco, G., Warren, J., Siri, S., Arciuli, J., Scott, S., & Wise, R. (2006). The role of semantics and grammatical class in the neural representation of words. *Cerebral Cortex, 16*, 1790–1796.
Vigliocco, G., Meteyard, L., Andrews, M. & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition, 1*(2), 215–244.
Weber–Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience 8*(3), 231–256.
Weinreich, U. (1953). *Language in contact*. NY: The Linguistic Circle of New York.
Wentura D., Rothermund K., Bak P. (2000). Automatic vigilance: the attention–grabbing power of approach- and avoidance-related social information. *Journal of Personality, and Social Psychology, 78*, 1024–1037.
Winskel, H. (2013). The emotional Stroop task and emotionality rating of negative and neutral words in late Thai–English bilinguals. *International Journal of Psychology, 48*(6), 1090–1098.

Wu, Y. J., & Thierry, G. (2012). How reading in a second language protects your heart. *The Journal of Neuroscience*, *32*(19), 6485–6489.

Zhao, X., Li, P., Liu, Y., Fang, X., & Shu, H. (2011). Cross–language priming in Chinese–English bilinguals with different second language proficiency levels. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society*. 801–806.

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: six replication attempts. *PLoS ONE*, *7*(12), e51382.

Zwaan, R. A. & L. J. Taylor. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology: General, 135*. 1–11.

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic–relatedness judgments. *Psychonomic Bulletin & Review, 10*, 954–958.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171.