Lexical Inference, Memory Representation, and Incidental Learning in Second Language Vocabulary Acquisition

A Dissertation

Submitted to the University of Tsukuba

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Linguistics

Akira HAMADA

2015

Abstract

Vocabulary learning is an inevitable task in successful second and foreign language (L2) acquisition. Although it requires a good balance of various learning approaches, evidence from past studies runs counter to theoretical accounts of word learning from text, compared to traditional paired-associate learning. In this dissertation, I examined how reading comprehension influences the incidental learning of words by Japanese students of English as a foreign language (EFL) through three sets of studies: generation and encoding of lexical inference, memory representations of words established by lexical inference, and incidental learning from text. Latent semantic analysis (LSA), a theory and method for extracting and representing the contextual-usage meaning of words, was used throughout each experiment to connect the usage-based model of language learning to discourse processing theories.

The first set of experiments explored whether and how Japanese EFL learners infer the meanings of unknown words and encode them in memory using information about word-context semantic similarity computed by LSA. In Experiment 1, an event-related brain potential study, university participants read target known and unknown words in two types of learning sentences that had higher- and lower-semantic similarity (HSS and LSS) with those meanings. Immediately after comprehending each sentence, they made plausibility judgments on the contextual-usage meaning of the words in test sentences. The N400 modulation elicited by the different target words showed that the Japanese EFL learners could anticipate upcoming words in the HSS learning sentences and integrate the inferred meaning into their memory. Interestingly, the effects of EFL reading proficiency were very small in the generation and encoding of lexical inferences.

Two more experiments were conducted to complement and expand on these findings using a reaction time methodology. The semantic similarity of a target word with

prior and subsequent contexts was isolated as a direction factor in contextual elaboration (i.e., forward and backward contextual elaboration). University participants read target known and unknown words in both a complete (Experiment 2A) and chunk-parsed sentence (Experiment 2B) with 2 (HSS and LSS) × 2 (forward and backward) conditions. After each sentence, they judged whether the target word as a prime was semantically related to the corresponding probe flashed on a computer screen. Both judgment-accuracy and reaction-time data proved consistent with results of Experiment 1 in terms of lexical inference generation and encoding during EFL reading. Furthermore, the results indicated that Japanese EFL learners were sensitive enough to backward contextual elaboration to generate and encode unknown word meanings, although it required more complicated cognitive processes compared to forward contextual elaboration.

The second set of experiments probed into how specific and robust memory of new words Japanese EFL learners are able to construct in their mental representations for the establishment of form-meaning connections. The first two experiments investigated the specificity of semantic representations of words inferred from HSS and LSS learning sentences with on-line and off-line methodological approaches. In Experiment 3A, two types of probe words representing specific and general lexical inferences were presented in the same two-word relatedness judgment test as in Experiment 2. The results found the effects of word-context semantic similarity; university participants inferred and encoded the synonyms and superordinate meanings of unknown words from the HSS and LSS learning sentences, respectively. Furthermore, the effects of English reading proficiency turned out to be significant, indicating that less-skilled learners were not able to narrow the possible meanings of those words down to specific ones in the HSS learning sentences while reading, although they still generated and encoded a relatively vague inference. The classification of unknown word meanings produced by other participants in Experiment 3B, which used an off-line lexical inference test, produced similar results as Experiment

3A.

In the next two experiments, I examined the relative accessibility of a word form represented in memory after its semantic representation was established by successful lexical inference. University participants in Experiment 4A read target unknown words embedded in short stories including either HSS or LSS learning sentences, and then determined if they had appeared in the stories as swiftly as possible. The word recognition test revealed that word-form memory turned out to be less accessible when the learners successfully inferred the meaning of an unknown word than when its meaning was unresolved. These different patterns of word recognition speed can be attributed to learners' attention allocation while constructing a mental representation of a text; accordingly, participants in Experiment 4B were engaged in task-induced lexical processing. They read the short stories to perform a recall task, in which some of them were asked to recall, in Japanese, what they had read (i.e., meaning focused), while others were required to recall it in English (i.e., form focused), after the word recognition trials. Whereas the reaction times in the meaning-focused condition were comparable with the Experiment 4A results, those in the form-focused condition did not differ between successful and unsuccessful lexical inferences. This suggests that controlling attention allocation through a form-focused task promotes the relatively simultaneous activation of a word form and its meaning in the mind, improving the establishment of form-meaning connections during reading.

The final set of experiments addressed the issue of incidental gains in knowledge of word meaning and usage through classroom-based reading activities. Experiment 5A was conducted to confirm the effects of the word-context semantic similarity computed by LSA in contextualized L2 vocabulary learning. Three groups of different vocabulary learning strategy users (beginner-level university participants) deliberately learned target unknown words with context and completed a Vocabulary Knowledge Scale test. The results provided evidence that the HSS learning sentences promoted knowledge gains in word meaning and usage. In addition, the LSS learning sentences consistently reduced learning outcomes, and such hindering effects appeared particularly in the less-frequent strategy of using context. Therefore, Experiment 5B explored two contrastive views in incidental vocabulary learning: (a) The lack of conscious intention to learn input reduces the knowledge gains that should be improved by HSS learning sentences, and (b) Task-induced comprehension of contextual information improves incidental knowledge gains. Other beginner-level university participants were given target words with HSS and LSS learning sentences in the lexical inference and multiple-choice glosses tasks. The same test was used and analyzed in the same way as in Experiment 5A. The participants' task performances showed that they inferred the target word meanings from the HSS learning sentences more successfully than from the LSS ones, but obtained similar scores between conditions in the multiple-choice glosses task. Nevertheless, the test results showed that the HSS learning sentences greatly contributed to the incidental knowledge gains in word meaning and usage. Thus, LSA predicted the outcomes of incidental word learning, which suggests that the learners acquired lexical knowledge from usage-based contextual information.

Together, these sets of experimental studies highlight the importance of deriving the contextual-usage meaning of words for incidental vocabulary learning. The complete findings support the usage-based model provided by the assumptions underlying the LSA theory and reveal its relationship to mental representation construction in comprehension. This dissertation, lastly, discusses pedagogical interventions for improving incidental L2 vocabulary learning from text, in terms of reading instructions, through developments in teaching materials using LSA, a task-induced focus-on-form approach, and multiple-componential and multi-test approach assessments of vocabulary knowledge in line with English education in Japan.

Acknowledgements

Many people provided advice, support, and encouragement for my completion of this PhD dissertation. My research interests in the integration of discourse processing and vocabulary learning in L2 were established over the course of my studies at the University of Tsukuba. I have been very blessed to learn from profound mentors whose assistance helped me shape my research directions and approaches. Special thanks are especially due to Professor Yuji Ushiro, my chief supervisor, for his generous encouragement and helpful guidance. He has continuously provided me with invaluable advice on navigating my research interests and allowed me much freedom and flexibility in conducting this research project. I have learned much, from the theories of second language acquisition to more specific fields concerned with reading and vocabulary learning in foreign languages, from his lectures and Reading Research Group, which provided me with a wealth of valuable knowledge.

The completion of this dissertation was also supported by the following vice supervisors. At its inception, Professor Hirosada Iwasaki steered my research interests to L2 vocabulary learning. I am also thankful to Professor Akira Kubota, whose broad knowledge helped me associate the present findings with theoretical framework in the right area. My deep gratitude goes to Professor Yukie Horiba at Kanda University of International Studies, who carefully reviewed this dissertation as an external supervisor. Her leading research on connecting L2 discourse processing to knowledge acquisition inspired me to tackle the complex issues addressed in this dissertation.

In conducting individual experiments, I have received help from senior associates in Professor Ushiro's seminar. They gave me very helpful advice on how to comprehend past research, design my own experiments, use statistical analysis, and discuss findings. These respected mentors include Dr. Chikako Nakagawa, Dr. Yuko Hoshino, Dr. Haruka

Shimizu, Dr. Akari Kai, Dr. Shuichi Takaki, Dr. Yusuke Hasegawa, and Dr. Shingo Nahatame. I also need to thank my schoolmate, Ms. Yukino Kimura, for encouraging me with her constant support throughout the present research. Although it is hard to list all of the associates who helped me, I am fortunate to be in a seminar that has such supportive people operating at the highest levels of academic and practical professionalism.

The present project has been supported by the Japan Society for the Promotion of Science from 2013 to 2015 (No. 25.487). Many thanks go to the students at the University of Tsukuba and Teikyo University of Science who joined my experiments. This research could not have been conducted without their kind cooperation. I am still a novice language teacher, but they always studied very hard and provided me with many hints related to my research.

Finally, I hope to express my most profound gratitude to my parents for their understanding, encouragement, and warm support. My father understood well the path that I should take, and my mother always watched me closely. I have the utmost respect for their belief in my future.

Akira HAMADA

Tsukuba December 1, 2015

Table of Contents

Chapter 4
Study 2: Memory Representations of Words Established by Lexical Inference

List of Tables

List of Figures

List of Appendices

Major Abbreviations and Acronyms

ANOVA	Analysis of Variance
EEG	Electroencephalogram
EFL	English as a Foreign Language
ERP	Event-Related Potential
H	Hypothesis
HSS	Higher Semantic Similarity
L1	First Language
L2	Second Language
LSA	Latent Semantic Analysis
LSS	Lower Semantic Similarity
MANOVA	Multivariate Analysis of Variance
RQ	Research Question
RT	Reaction Time
SLA	Second Language Acquisition

Publications and Presentation

The present dissertation covers multiple experimental studies reported on in the publications and presentations listed below:

Hamada, A. (2013). Lexical inference specificity and its activation level: Effects of contextual constraint and L2 reading proficiency. *ARELE: annual review of English language education in Japan*, *24*, 93–108. (Chapter 4)

Hamada, A. (2014). Using latent semantic analysis to promote the effectiveness of contextualized vocabulary learning. *JACET Journal*, *58*, 1–20. (Chapter 5)

Hamada, A. (2015a). Effects of forward and backward contextual elaboration on lexical inferences: Evidence from a semantic relatedness judgment task. *Reading in a Foreign Language*, *27*, 1–21. (Chapter 3)

Hamada, A. (2015b). Improving incidental L2 vocabulary learning with latent semantic analysis. *ARELE: annual review of English language education in Japan*, *26*, 61–75. (Chapter 5)

濱田彰. (2013, 8 月).「文脈内語彙学習における Form-Meaning Connection: 文脈の収束性と記銘タスクの効果」. 第 39 回全国英語教育学会北海道研究大会. (Chapter 4)

Hamada, A. (2015, July). *An ERP study on implicit learning of words from context by second language readers*. Poster session presented at the 22nd Annual Meeting Society for the Scientific Study of Reading, Hawaii. (Chapter 3)

Committee in charge:

Professor Yuji Ushiro, chief supervisor (University of Tsukuba)

Professor Hirosada Iwasaki, vice supervisor (University of Tsukuba)

Professor Akira Kubota, vice supervisor (University of Tsukuba)

Professor Yukie Horiba, external supervisor (Kanda University of International Studies)

Chapter 1

Introduction

1.1 Overview of Current Problems

The component-skills view has argued that language use is involved in multiple cognitive processes through the interaction of multiple subsets of linguistic knowledge. For example, readers comprehend written discourse by word decoding via vocabulary knowledge, and syntactic analysis via grammar knowledge (e.g., Grabe, 2009; Koda, 2005; Perfetti, 2007). Particularly, vocabulary knowledge is one of the most essential components in second language (L2) use because past L2 research has found its strong correlations with listening (Stæhr, 2009; van Zeeland & Schmitt, 2013), reading (Horiba, 2012; Jeon & Yamashita, 2014; Qian, 1999, 2002; van Gelderen et al., 2004; Yamashita & Shiotsu, 2015), speaking (Koizumi & In'nami, 2013), and writing (Schoonen et al., 2003; Webb, 2005). These observations clearly show that learning vocabulary is an inevitable part of second language acquisition (SLA).

There are, broadly, three aspects of lexical knowledge: form, meaning, and use (Nation, 2013). Although the basic goal of vocabulary learning is to build form-meaning connections in memory, the knowledge of word usage, such as grammar collocations and registries, is also required in language communication. Given that different aspects of word knowledge function variably in language use, successful L2 vocabulary learning requires a good balance of various learning methods (Nation, 2013; Nation & Webb, 2011). Additionally, theoretical and empirical accounts of vocabulary acquisition have emphasized the importance of contextualized learning (e.g., Hasegawa, 2012, 2013; Jiang, 2000; Webb, 2007a) because learning words only from definitions often results in poorer performance on tasks using words in context, compared to contextualized learning (e.g., Bolger, Balass, Landen, & Perfetti, 2008; McKeown, 1985).

In the study of contextualized learning, many researchers have largely focused on *incidental learning*, or by-products of reading activities, because they have agreed that discourse context is seen as a primary source for learning new words (e.g., Bolger et al., 2008; Paribakht & Wesche, 1999; Wesche & Paribakht, 2010). Typical incidental learning research is conducted based on the following causal circle:

> Learners with good skills at inferring from context develop larger vocabulary sizes. A large vocabulary size supports decoding skills and skill at accessing word meaning. These four aspects of vocabulary skills enable learners to achieve better comprehension of text. Better comprehension of text allows learners to process more input. Increased input and practice in processing input allows more inferring from context. (Nation, 2013, p. 351)

However, some studies have indicated that activities such as extensive reading made modest contribution to the knowledge gains in new L2 words (Day, Omura, & Hiramatsu, 1991; Horst, Cobb, & Meara, 1998; Pitts, White, & Krashen, 1989; Waring & Takaki, 2003). Similarly, others suggested that successful inference of new word meanings do not have a relationship to great retention of those words (e.g., Bensoussan & Laufer, 1984; Frantzen, 2003; Hulstijn, 1992; Laufer, 1997; Mondria & Wit-de Boer, 1991).

These contrastive views are primarily a result of a lack of cognitive approaches to addressing the issue of why and how vocabulary knowledge can be developed through reading (e.g., Bordag, Kirschenbaum, Tschirner, & Optiz, 2015; Inohara & Kusumi, 2012; Pulido, 2007). The theoretical background of incidental learning is often supported by the *usage-based model*, which argues that language can be learned from frequent input of information about how it is used in context (Ellis, 2002; Tomasello, 2003). Specifically, this usage-based approach views language acquisition as input-driven; namely, “a large and representative sample of language is required for learners to abstract a rational model that is a good fit to the language data" (Ellis & Wulff, 2015, p. 86). Regarding what kinds of input can promote vocabulary learning, Landauer and Dumais (1997) applied the algorithms of *Latent Semantic Analysis* (LSA) to simulating the vocabulary development in a first language (L1). In LSA, a computer learns a word meaning from its coincidence frequency with other words used in a particular context based on the statistical analyses of a large-scale corpus. If LSA has high validity for predicting L2 learners' word learning performance, it can become a powerful tool in creating informative inputs to improve incidental L2 vocabulary learning.

In incidental word learning, the primary goal of L2 learners is to comprehend a text. To understand a text's content successfully, readers have to construct a coherent memory representation of the text by processing written inputs with multiple cognitive processes (e.g., Graesser, Singer, & Trabasso, 1994; Just & Carpenter, 1980, 1992; Kintsch, 1998). Therefore, cognitive processes involved in reading comprehension are closely related to incidental L2 vocabulary learning, because the acquisition of new words during reading begins with building a memory representation of these words through inference (e.g., Bolger et al., 2008; Bordag et al., 2015; de Bot, Paribakht, & Wesche, 1997; Hamada, 2011; Paribakht & Wesche, 1999; Nassaji, 2003, 2006; Nation, 2013; Wesche & Paribakht, 2010). According to de Bot et al. (1997), when readers encounter an unknown word in a text, they sometimes try to gather lexical features included in the word from context, in order to comprehend the text. Theoretical accounts of word learning suggest that filling gaps in word knowledge through this input processing contributes to vocabulary growth. Given that the generation of lexical inference is part of discourse comprehension processes, it is important to investigate how memory traces of words are encoded and represented as the consequence of inference.

In line with this background, this dissertation addresses why input processing by lexical inference promotes incidental L2 vocabulary learning as shown in Figure 1.1. The primary goal of this research is to associate the usage-based model of language learning with the cognitive processes involved in memory representation construction during L2 reading, because the construction of text memory and the learning of words in a text should be interwoven. More specifically, the cognitive mechanism involved in word memory construction through lexical inference and the acquisition of new vocabulary knowledge are simultaneously examined in Japanese learners of English as a foreign language (EFL). This helps understand the factors and conditions that improve incidental L2 vocabulary learning. The findings will have pedagogical value for developing teaching materials, teaching procedures, and assessment methods aimed at cultivating vocabulary knowledge available for language communication.

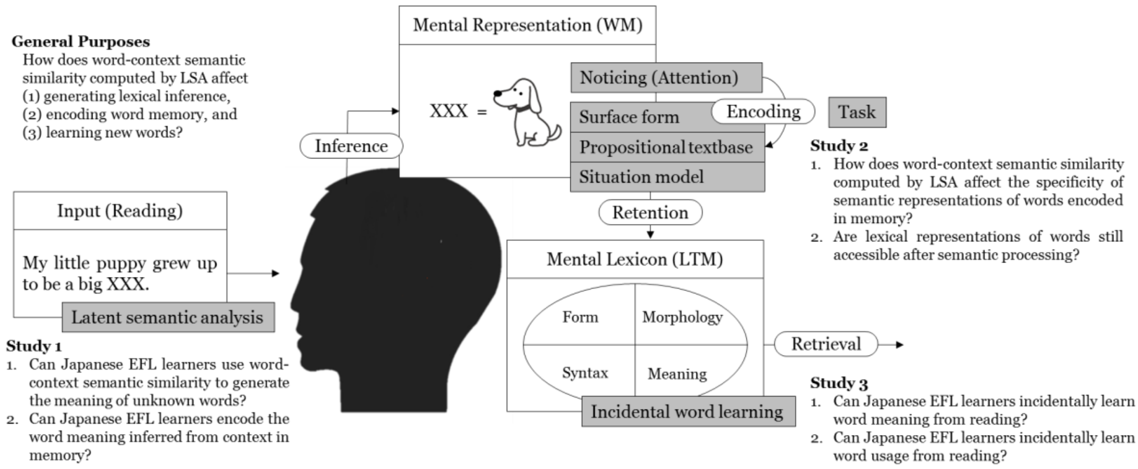


Figure 1.1. A schematic image of the issues involved in incidental L2 vocabulary learning from reading that are addressed in this dissertation. The rounded boxes indicate the cognitive processes from input processing to learning. The theories and models highlighted gray will be relevant to each mechanism. WM = working memory, LTM = long-term memory.

1.2 Organization of This Dissertation

This dissertation consists of seven chapters including the present one. After an overview of the issues regarding incidental L2 vocabulary learning in Chapter 1, Chapter 2 summarizes the literature closely related to the entire area of this research, from theories and empirical studies on (a) vocabulary acquisition, to (b) the principles, algorithms, and application of LSA in research on language processing and acquisition, (c) memory representation construction in reading, and (d) noticing hypothesis. Additionally, research methodologies are discussed in order to address the connection between the reading process and its products in terms of inference, encoding, and retention of new words. In addition to a review of past research, the connection of previous findings to the present research and the significance of this study are described.

The current research comprises three main studies, each of which consists of multiple experiments, as is briefly shown in Figure 1.2. Study 1, reported in Chapter 3, focused on the generation and encoding of lexical inference in on-line L2 sentence comprehension.[1] In Experiment 1, the inference and encoding of unknown words in sentential inputs was examined using event-related potentials (ERPs). These target words were placed at the ends of sentences, differing according to the word-context semantic similarity computed by LSA; that is, the prior context semantically elaborated on the meaning of those words. Experiment 2A employed a semantic priming method to examine the same semantic similarity factors for the target words, semantically elaborated by subsequent context, in order to complement the findings of Experiment 1. Experiment 2B was conducted as a follow-up study to reject an uninteresting possibility revealed in Experiment 2A.

[1] The technical term *on-line* means *during the course of reading*, but *off-line* does not (van den Broek, Fletcher, & Risden, 1993).

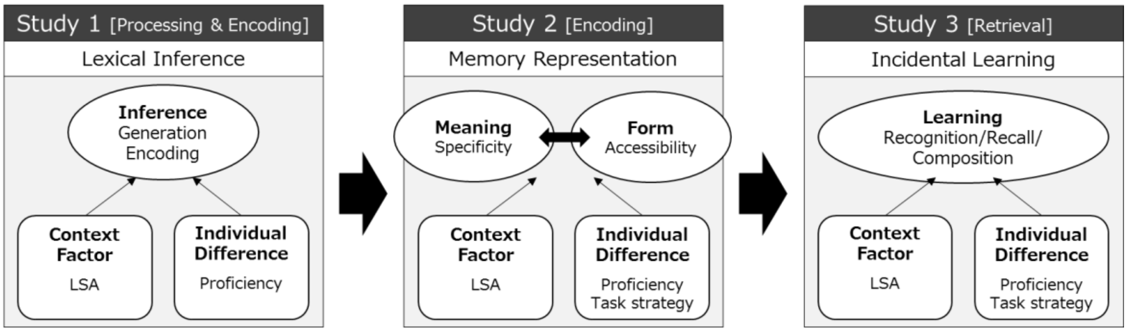


Figure 1.2. The overall scheme of the current research experiments.

Chapter 4 describes Study 2, on the memory representations of words constructed by lexical inference. Experiments 3A and 3B examined the specificity of semantic representations of target words encoded in memory using on-line and off-line methods, respectively. The word-context semantic similarities and participants' English reading proficiency were dealt with as an interaction effect on lexical inference specificity. In Experiment 4A, the relative accessibility to lexical representations of target words (i.e., spelling) was examined using a word recognition test. Recognition latencies for target words were compared between participants' successful inferences of the meaning of those words and those that were unsuccessful. Experiment 4B investigated whether task-induced lexical processing affects the accessibility to lexical representations stored in memory. Specifically, reading tasks were given to participants to direct their attention to the meanings and form aspects of target words.

Chapter 5 reports two experiments conducted as Study 3. They examined whether the word-context semantic similarity quantified by LSA has high validity for predicting L2 learners' word learning performance. Generally, Japanese undergraduates were given 20 target words with contexts whose propositions had higher or lower semantic similarities to those words. In Experiment 5A, they intentionally learned those words to reveal how contextualized vocabulary learning is promoted by word-context semantic similarity. Experiment 5B, in turn, provided participants with lexical inference and multiple-choice glosses tasks to determine if they incidentally acquired new vocabulary knowledge. A Vocabulary Knowledge Scale (VKS) test was used to assess the gains in knowledge of word meaning and usage.

Chapter 6, a general discussion, summarizes the main findings from the three studies and discusses the theoretical issues related to the process and products of incidental L2 vocabulary learning with reference to theories in linguistics, psychology, and second language acquisition. The integrated results of these experiments are discussed from the perspectives of (a) initial stage of incidental L2 vocabulary learning, (b) inference specificity and form-meaning connections, and (c) LSA theory and gains in word meaning and usage knowledge.

In Chapter 7, concluding remarks offer more definite observations of the current research and pedagogical implications related to promoting lexical inference skills and incidental vocabulary learning in EFL classrooms in Japan. Finally, some limitations and suggestions for future studies are introduced. The complete materials used in this set of experiments are provided in the Appendices.

Chapter 2

Literature Review

2.1 Incidental L2 Vocabulary Learning

Language learning involves the acquisition of a novel linguistic system connecting lexical forms to their meanings and rules of functional use. In the context of SLA, ordinary learners already know the referents, or meanings, of new word forms in their L1. Therefore, their primary task is to associate the conceptual meanings stored in their minds with the novel words (Wolter, 2006). In order to use language in real life situations, they are further required to comprehend its function about when and for what purpose they can use it (e.g., N. Ellis, 2008; R. Ellis, 2008; Gass, Behney, & Plonsky, 2013; Larsen-Freeman & Long, 1991; Robinson, 2011; Tomasello, 2003; van Patten, Williams, Rott, & Overstreet, 2004). Although L2 vocabulary learning is a subsequent goal of broader goals that should be achieved in class, ultimate attainment in vocabulary acquisition involves form-meaning-function mapping in the same way as in general SLA (Jiang, 2000, 2002; Nation, 2013).

Since traditional paired-associate learning, such as repeated memorization of the word form and meaning, was replaced by communicative language teaching including task-based and form-focused approaches, the role of context in vocabulary learning has been emphasized (e.g., Bolger et al., 2008; Hamada, 2011; Hasegawa, 2012, 2013; Webb, 2007a, 2008). In language communication, words seldom appear in isolation; instead, each word meaning is woven into a context to build a coherent message. Let us consider example (1), in which the target word is italicized (adopted from Webb, 2007a, p. 80).

(1) They continue to *mourn* for years after the death of their friend.

The contextual information provides various lexical properties of *mourn*, such as its orthography, part of speech, syntactic features, meaning, and word usage. This suggests that vocabulary acquisition and development should progress when learners comprehend unknown words in meaningful contexts (e.g., Bolger et al., 2008; de Bot et al., 1997; Jiang, 2000; Hasegawa, 2012, 2013). Actually, foreign language textbooks show students a particular topic or situation (e.g., sending an email, booking a hotel, or shopping) and present lists of new words that appear in the context (Nation, 2013).

A vast amount of L1 acquisition and SLA literature has accumulated insights into vocabulary learning from text. For example, Cunningham and Stanovich (1991) proved that the amount of exposure to written texts becomes the stronger predictor of children's L1 vocabulary knowledge compared to their age, general knowledge, and phonological decoding ability. In incidental L2 vocabulary acquisition, research in extensive reading suggested that reading habits promote the acquisition and development of L2 vocabulary knowledge consistent with Krashen's (1989) input hypothesis. It is important to note that one of the focal questions addressed in previous research concerned the possible factors contributing to knowledge gains for novel words. The chronological research progress in L1 and L2 from the 1980s to 2010s can be roughly overviewed as follows:

[1] 1980s: The relationship between reading comprehension (or extensive reading) and the learning rates of new words was explored in both L1 (Cunningham & Stanovich, 1991; Nagy, Anderson, & Herman, 1987; Nagy, Herman, & Anderson, 1985) and L2 (e.g., Day et al., 1991; Horst et al., 1989; Krashen, 1989; Pitts et al., 1989). Although these studies suggested that the exposure to novel words resulted in the incidental acquisition of those words, they did not explore the mental processes concerning lexical processing or incidental learning.

[2] 1990s: Think-aloud studies explored the mental processes of lexical inferences based

on the assumption that the inference is the prerequisite for establishing a knowledge representation of a new word during reading (de Bot et al., 1997; Haastrup, 1991; Huckin & Bloch, 1993; Paribakht & Wesche, 1999). The results were further applied to find predictors affecting successful lexical inferences and incidental form-meaning connections, mainly focusing on the context factors (Laufer, 1997; Li, 1988; Mondria & Wit-de Boer, 1991; Paribakht & Wesche, 1999) and individual differences (Cain, Oakhill, & Lemmon, 2004; Nassaji, 2003). However, some researchers showed that successful lexical inferences were not necessarily correlated with the retention of target word meanings (Frantzen, 2003; Laufer, 1997; Mondria & Wit-de Boer, 1991). As the counter results were discussed based on the role of attention or noticing (Ellis, 1994; Hulstijn, 1992; Hulstijn & Laufer, 2001), further research tested the effects of pedagogical interventions on the levels of lexical processing and the connection with learning outcomes (Fraser, 1999; Hulstijn, Hollander, & Greidanus, 1996; Laufer & Hulstijn, 2001; Paribakht & Wesche, 1997; Wesche & Paribakht, 2000).

[3] 2000s: Research tried to determine how much exposure to target words is necessary for L2 learners to acquire word knowledge (e.g., Chen & Truscott, 2010; Rott, 1999, 2007; Waring & Takaki, 2003; Webb, 2007b, 2008). Although the absolute number of word encounters was not identified, the research showed that multiple exposures explained the development of lexical knowledge. Another approach examined what aspects of lexical knowledge were acquired via lexical inferences (Bolger et al., 2008; Hamada, 2011; Webb, 2007a, 2007b, 2008).

[4] 2010s: Because there was little research that explained why words can be learned during reading, the latest L1 and L2 studies explored the mental processes of lexical inferences and word learning. Particularly, they applied well-established models of reading comprehension and memory construction in order to simulate the cognitive mechanism involved in incidental word learning (e.g., Bolger et al., 2008; Bordag et

al., 2015; Hamada, 2012; Wesche & Paribakht, 2010). Furthermore, the number of studies focusing on the initial stage of vocabulary acquisition is growing thanks to the development of on-line measurements of word knowledge (Borovsky, Elman, & Kutas, 2012; Borovsky, Kutas, & Elman, 2010; Elgort, 2011; Elgort, Perfetti, Rickles, & Stafura, 2015; Elgort & Warren, 2014).

In summary, whereas past research established various factorial designs to identify the facilitators of incidental vocabulary learning, some empirical studies also showed the null or small effects of reading on incidental L2 vocabulary learning. Regarding these two contrastive findings, Inohara and Kusumi (2012) claimed that, until the 2010s, there were few cognitive approaches addressing why and how reading changes readers' vocabulary. Particularly, they emphasized the importance of hypothesizing the vocabulary acquisition processes during text reading to predict learning outcomes and provide pedagogical implications.

The theoretical and methodological development of discourse processing and vocabulary learning research allows us to examine the cognitive processes involved in incidental L2 vocabulary learning according to each perspective. From a theoretical viewpoint, many SLA researchers have proposed various kinds of theories, models, and hypotheses of language learning. To reveal the cognitive processes involved in incidental L2 vocabulary learning, the present study adopted *LSA theory* (e.g., Landauer & Dumais, 1997) and *noticing hypothesis* (Schmidt, 1990), which are reviewed in the subsequent sections. The LSA theory is associated with the usage-based model in vocabulary learning in which learners should obtain word knowledge by experiencing word usage in context (e.g., Ellis, 2002). Noticing is considered as a language learning facilitator. Although "incidental learning refers to the mode in which participants are not forewarned of an upcoming test for a particular type of information" (Hulstijn, 2005, p. 132), researchers indicated that the robust form-meaning-function mapping of words should require readers' input processing with the conscious attention to find out the lexical features of the words from context (Bolger et al., 2008; Ellis, 1994; Leung & Williams, 2011). In Section 2.5, I will review the methodological issues to examine these theoretical frameworks.

2.2 Latent Semantic Analysis

An important theme in language acquisition research is to explain its cognitive mechanisms. From numerous theories, models, and hypotheses developed in the field of SLA research, the present research has employed the theories related to inductive learning because incidental L2 vocabulary learning is inductive, occurring without the explicit presentation of word definitions or usage (Hulstijn, 2005). Induction ability is important because L2 learners must adapt a relatively limited set of linguistic inputs to a potentially infinite variety of situations in language communication (Gass et al., 2013). Given that L2 vocabulary learning usually consists of mapping novel forms to familiar meanings, it is necessary to understand how learners induce the rules about form-meaning-function connections from a preexisting language representation in their mental lexicon.

In psycholinguistics, researchers developed the LSA theory of knowledge representation and its inductive acquisition using statistical language analysis with a large-scale corpus. LSA was developed for automatic indexing and retrieval, in which a computer learns a word meaning, represented by the degree of word-word, word-text, and text-text semantic similarity (hereafter, *LSA similarity*), from its co-occurrence frequency with other words used in a particular context (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990). Landauer and Dumais (1997) applied this method to establish human knowledge structures and measure human verbal concepts to predict the cognitive mechanism of language acquisition. As spoken and written languages are sources of input for human language acquisition, a large-scale corpus compiled from the language data represents language input for LSA. It then constructs a high-dimensional semantic space through mathematical algorithms and statistical analyses to represent human knowledge of the concepts. As shown in Figure 2.1, the meanings of each word and passage are represented in a semantic space. This allows LSA to produce measures of word-word, word-passage, and passage-passage semantic distances. The distances are regarded as the semantic similarity between concepts.

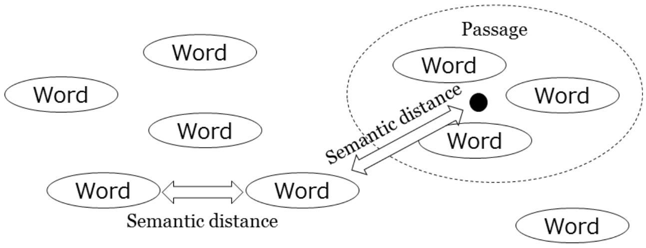


Figure 2.1. A diagram of a high-dimensional semantic space in LSA (simplified into two dimensions), adapted from Inohara and Kusumi (2011, p. 103). Words represent each word meaning as a concept. The black dot is a centroid of multiple words, representing the concept of a passage composed of the words included in it.

Because the focus of this study is on word learning based on information about its contextual usage, it is necessary to assess how the theoretical reasoning and procedure underlying LSA are related to the learning processes. Therefore, Section 2.2.1 will introduce the fundamental rules of LSA and Section 2.2.2 will describe the mathematical and statistical analyses to capture the process of word knowledge acquisition from context. Finally, in Section 2.2.3, some empirical studies applying LSA will be reviewed, focusing on lexical processing and the consequential acquisition of word knowledge.

2.2.1 The Principles of LSA

The learning mechanism behind LSA is similar to the usage-based model because both assume that we learn language based on information on how frequently the word appears in a particular context (e.g., Ellis, 2002; Ellis & Ferreira-Junior, 2009; Ellis, O'Donnell, & Römer, 2013; Inohara & Kusumi, 2011, 2012; Landauer & Dumais, 1997). The core of the usage-based model is that our linguistic knowledge emerges in memory because of multiple exposures to a significant amount of information about its usage in different contexts (Ellis, 2002; Tomasello, 2000a, 2000b, 2003). Landauer and Dumais (1997) assume that "some powerful [induction] mechanism exists in the minds of children that can use the finite information they receive to turn them into competent users of human language" (p. 212). Previous L1 and L2 empirical studies described this induction ability that allows students to acquire new words through the following two broad points.

- Intention reading: It is the ability to learn communicative intentions in order to acquire the appropriate use of linguistic symbols or forms (Tomasello, 2000a, 2000b). According to Tomasello (2003), this skill enables some induction processes such as analogy, which is employed when the new linguistic inputs play similar functional roles to the students' existing knowledge.
- Pattern finding: It is the ability to form conceptual categories of similar concepts through the statistical learning of linguistic inputs' frequencies (e.g., Ellis, 2002; Ellis & Ferreira-Junior, 2009; Ellis et al., 2013). Many researchers have suggested that this skill is necessary to induce the rules of language usage from different patterns of context-based inputs (Chaffin, 1997; Landauer & Dumais, 1997; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Tomasello, 2003).

Ellis and Wulff (2015) commented on these two learning mechanisms in the context of SLA. First, language learning is regarded as associating form with meaning or function. To construct form-meaning associations in memory, learners engage in rational cognitive processing, in which they induce the rules of form-meaning mapping from context (i.e., the pattern-finding process). This process is supported by intention reading, in which learners notice cues for understanding the meaning and usage of language with the help of the usual social and pedagogical interactions (e.g., R. Ellis, 2008; Gass et al., 2013). Accordingly, the knowledge of a certain form-meaning pair is a reflection of the learner's accumulated and abstracted experiences involving repeated uses of particular expressions (e.g., Tomasello, 2000a). These theoretical accounts suggest that in relation to incidental L2 word learning, the usage-based approach includes the following two constructs:

- Vocabulary knowledge is developed based on the induction processes (e.g., Landauer & Dumais, 1997). The sources for vocabulary learning are patterns of language usage, including collocational and frequency information (e.g., Crossley, Salsbury, Titak, & McNamara, 2014; Ellis, 2002; N. Ellis, 2008; Ellis & Ferreira-Junior, 2009; Ellis et al., 2013; Kidd, Lieven, & Tomasello, 2010).
- Lexical items, or linguistic symbols, themselves do not have meanings; instead, word meanings in a context are determined by their usage, including grammatical patterns (Langacker, 2008). Language is always interwoven into context, so that language use varies according to contextual factors (Tyler, 2010). Success in incidental L2 word learning from reading depends on whether learners can extract the contextual-usage meanings of words by reading the L2 text.

The LSA theory adopts two assumptions about this usage-based language learning: the principles of direct and indirect relations of co-occurrence (Inohara & Kusumi, 2011; Landauer & Dumais, 1997). In the first principle, words co-occurring in the same context should share similar semantic properties (Landauer, Foltz, & Laham, 1998). For example, in the sentence *The dog jumped up and licked his face*, the target word *lick* co-occurs with other content words such as *dog*, *jump*, and *face*. The usage-based model suggests that when learners have frequent exposure to the word *lick* in this context, they acquire its semantic relations or contextual-usage meaning (Ellis & Ferreira-Junior, 2009; Kidd et al., 2010; Tomasello, 2003). For the second principle, indirect relations are adapted to resolve the issue included in the direct relations of co-occurrence. For example, the target word *lick* is not in the following sentence: *Her little puppy grew up to be a big dog*. Although there must be a semantic relation between *lick* and *puppy*, the direct relations alone are insufficient to capture human language learning. Landauer and Dumais (1997) indicated that in this case, because of the induction processes in language learning, we can learn the indirect semantic relations between words. In LSA, the semantic similarity between *lick* and *puppy* is evaluated by the indirect co-occurrences through the mediation of *dog*.

The first principle is consistent with empirical findings that the words semantically similar to a target word can be linguistic cues for meaning generation by inference and incidental learning (de Bot et al., 1997; Huckin & Bloch, 1993; Nassaji, 2006; Paribakht & Wesche, 1999; Wesche & Paribakht, 2010). In addition, the second principle should be related to the context diversity effects on incidental vocabulary learning (Bolger et al., 2008; Chen & Truscott, 2010; Elgort et al., 2015; Pellicer-Sánchez, 2015; Webb, 2008). According to Bolger et al. (2008), direct encounters with a target word in a single context establish only context-dependent word memory; however, the word memory is abstracted when the learner processes the word in different contexts many times. Vocabulary learning based on the direct relations of co-occurrence can be defined as the explicit instruction of word meanings by a teacher and the paired-associated learning of a word form and its meaning (Inohara & Kusumi, 2012). Contrary to this, the use of the indirect relations of co-occurrence requires numerous inputs, which cannot be gained only from explicit instructions, although vocabulary acquisition simulated by LSA is mostly based on the indirect relations of co-occurrence (Landauer & Dumais, 1997). In summary, the learning mechanism in the LSA theory is compatible with usage-based learning of language, motivating the present research to apply the LSA theory to incidental L2 vocabulary learning from reading. Although research on incidental L2 vocabulary learning has not explained how word knowledge can be gained from reading, the theories of LSA can provide a solution to this issue.

2.2.2 The Algorithm of LSA

Based on the results of machine learning, LSA outputs the strength of semantic similarities between concepts described by words and contexts; it is represented as the cosine of the angle (i.e., LSA similarity) formed by high-dimensional vectors (see Figure 2.2). The theoretical range of the LSA value is from –1.00 to 1.00, and semantic similarity strengthens as the value approximates 1.00 because the angle formed by two vectors approaches zero.

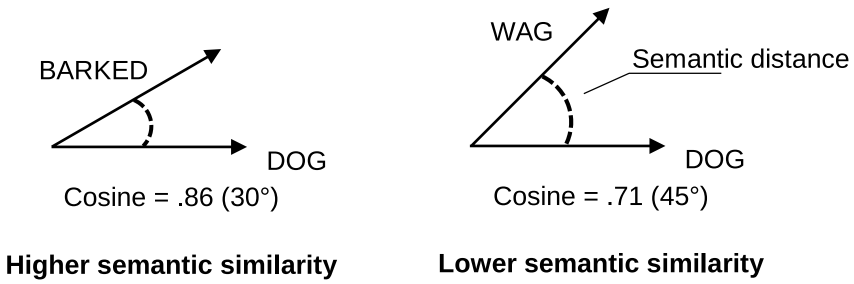


Figure 2.2. A simplified vector space of semantic similarity computed by LSA (Dennis, 2007). Each arrow represents a vector of the corresponding word's meaning.

Deerwester et al. (1990), Landauer and Dumais (1997), and Landauer et al. (1998) presented an LSA example using a small corpus compiled from nine titles about human-computer interaction and mathematical graph theory. The former and the latter titles as listed in Figure 2.3 are conceptually unrelated with each other, and the content words used in at least two of the titles were analyzed in this LSA. They were then transformed into word-by-context matrix $\{X\}$. Each cell archives the frequency to determine if each word appears in each passage denoted by its column. As a result, the original matrix has nine9 columns and 12 rows. In this case, *human* does not co-occur with either *user* or *minors* in the same passage. Therefore, Spearman's coefficients as the word semantic similarities are –.38 between *human* and *user*, and –.29 between *human* and *minors*; these assume to run counter to our intuition.

Example of text data: Nine titles of technical memoranda

c1: *Human* machine *interface* for ABC *computer* applications
c2: A *survey* of *user* opinion of *computer system response time*
c3: The *EPS user interface* management *system*
c4: *System* and *human system* engineering testing of *EPS*
c5: Relation of *user* perceived *response time* to error measurement

m1: The generation of random, binary, ordered *trees*
m2: The intersection *graph* of paths in *trees*
m3: *Graph minors* IV: Width of *trees* and well-quasi-ordering
m4: *Graph minors*: A *survey*

$\{X\}$ =

	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	1	0	0	1	0	0	0	0	0
interface	1	0	1	0	0	0	0	0	0
computer	1	1	0	0	0	0	0	0	0
user	0	1	1	0	1	0	0	0	0
system	0	1	1	2	0	0	0	0	0
response	0	1	0	0	1	0	0	0	0
time	0	1	0	0	1	0	0	0	0
ESP	0	0	1	1	0	0	0	0	0
survey	0	1	0	0	0	0	0	0	1
trees	0	0	0	0	0	1	1	1	0
graph	0	0	0	0	0	0	1	1	1
minors	0	0	0	0	0	0	0	1	1

Figure 2.3. A word-by-context matrix $\{X\}$, formed from the titles of five articles about human-computer interaction (c1-c5) and four articles about mathematical graph theory (m1-m4), adapted from Landauer et al. (1998, p. 265). The words italicized are the content words used in at least two of the titles.

Next, singular value decomposition (SVD), like a factor analysis, is applied to the matrix, "in which each cell frequency is weighted by a function that expresses both the word's importance in the particular passage and the degree to which the word type carries information in the domain of discourse" (Landauer et al., 1998, p. 263). In SVD, an $m \times n$ matrix is decomposed into three matrices as shown in Figure 2.4. According to Inohara and Kusumi (2011), two matrices $\{U\}$ and $\{V^*\}$ describe the original row and column entities as orthogonal vectors, each of which represents the semantics of words (i.e., left singular vector) and contexts (i.e., right singular vector). The third is diagonal matrix $\{D\}$ of a singular value, which is regarded as the word's importance.

$\{X\} = \{U\}\{D\}\{V^*\}$

$\{W_{12\times 9}\} =$

0.22	–0.11	0.29	–0.41	–0.11	–0.34	0.52	–0.06	–0.41
0.20	–0.07	0.14	–0.55	0.28	–0.50	–0.07	–0.01	–0.11
0.24	0.04	–0.16	–0.59	–0.11	–0.25	–0.30	0.06	0.49
0.40	0.06	–0.34	0.10	0.33	0.38	0.00	0.00	0.01
0.64	–0.17	0.36	0.33	–0.16	–0.21	–0.17	0.03	0.27
0.27	0.11	–0.43	0.07	0.08	–0.17	0.28	–0.02	–0.05
0.27	0.11	–0.43	0.07	0.08	–0.17	0.28	–0.02	–0.05
0.30	–0.14	0.33	0.19	0.11	0.27	0.03	–0.02	–0.17
0.21	0.27	–0.18	–0.03	–0.54	0.08	–0.47	–0.04	–0.58
0.01	0.49	0.23	0.03	0.59	–0.39	–0.29	0.25	–0.23
0.04	0.62	0.22	0.00	–0.07	0.11	0.16	–0.68	0.23
–0.03	0.45	0.14	–0.01	–0.30	0.28	0.34	0.68	0.18

$\{D\} =$

–3.34	–2.54	2.35	1.64	1.50	1.31	0.85	0.56	0.36

$\{V^*_{9\times 9}\} =$

0.20	0.61	0.46	0.54	0.28	0.00	0.01	0.02	0.08
–0.06	0.17	–0.13	–0.23	0.11	0.19	0.44	0.62	0.53
0.11	–0.50	0.21	0.57	–0.51	0.10	0.19	0.25	0.08
–0.95	–0.03	0.04	0.27	0.15	0.02	0.02	0.01	–0.03
0.05	–0.21	0.38	–0.21	0.33	0.39	0.35	0.15	–0.60
–0.08	–0.26	0.72	–0.37	0.03	–0.30	–0.21	0.00	0.36
0.18	–0.43	–0.24	0.26	0.67	–0.34	–0.15	0.25	0.04
–0.01	0.05	0.01	–0.02	–0.06	0.45	–0.76	0.45	–0.07
–0.06	0.24	0.02	–0.08	–0.26	–0.62	0.02	0.52	–0.45

Figure 2.4. Complete SVD of matrix {X}, adapted from Landauer et al. (1998, p. 266).

The final step of LSA is to reduce dimensionality using the singular value to an optional dimension, which creates an approximate matrix of its original. In this example, Landauer et al. (1998) used the first two values, shaded columns of the three matrices in Figure 2.4, for dimensionality reduction from nine to two dimensions. Figure 2.5 displays the reconstructed two-dimensional approximate matrix {*X'*}. Every value in each cell represents how well the corresponding word can contribute to expressing its context like an eigenvalue in a factor analysis (Inohara & Kusumi, 2011). For example, *human* is more typically used in c1 to c4 (0.16 to 0.47; human-computer interaction topics) than in m1 to m4 (–0.16 to –0.05; mathematical graph theory topics). Furthermore, the *human-user* semantic similarity increased to .94 and the *human-minors* semantic similarity decreased to –.83. Thus, even though these words do not appear in the same context, LSA is able to determine that the terms *human* and *user* can occur in contexts of similar sense, which realizes the second principle of the indirect relations in LSA.

{*X'*} =

	c1	c2	c3	c4	c5	m1	m2	m3	m4
Human	0.16	0.40	0.38	0.47	0.18	–0.05	–0.12	–0.16	–0.09
interface	0.14	0.37	0.33	0.40	0.16	–0.03	–0.07	–0.10	–0.04
computer	0.15	0.51	0.36	0.41	0.24	0.02	0.06	0.09	0.12
User	0.26	0.84	0.61	0.70	0.39	0.03	0.08	0.12	0.19
System	0.45	1.23	1.05	1.27	0.56	–0.07	–0.15	–0.21	–0.05
response	0.16	0.58	0.38	0.42	0.28	0.06	0.13	0.19	0.22
Time	0.16	0.58	0.38	0.42	0.28	0.06	0.13	0.19	0.22
ESP	0.22	0.55	0.51	0.63	0.24	–0.07	–0.14	–0.20	–0.11
Survey	0.10	0.53	0.23	0.21	0.27	0.14	0.31	0.44	0.42
Trees	–0.06	0.23	–0.14	–0.27	0.14	0.24	0.55	0.77	0.66
Graph	–0.06	0.34	–0.15	–0.30	0.20	0.31	0.69	0.98	0.85
Minors	–0.04	0.25	–0.10	–0.21	0.15	0.22	0.50	0.71	0.62

Figure 2.5. The two-dimensional matrix reconstructed from matrix {X} adapted from Landauer et al. (1998, p. 267).

In an LSA, large corpora are used as language input for learning word meaning. Accordingly, LSA similarities can reflect representations of human language knowledge and performances ranging from developmental acquisition of vocabulary (e.g., Crossley, Salsbury, McCarthy, & McNamara, 2008; Inohara & Kusumi, 2012; Landauer & Dumais, 1997) to word-categorization (e.g., Landauer & Dumais, 1997), semantic priming (e.g., Inohara & Kusumi, 2011), and reading comprehension (e.g., Wang, Pomplun, Chen, Ko, & Rayner, 2010; Wolfe, Magliano, & Larsen, 2005).

2.2.3 The Application of LSA to Text Processing and Acquisition

Landauer et al. (1998) suggested the merit of the application of word and context meaning representations quantified by LSA in the following two ways. First, LSA can be a practical tool for obtaining approximate estimates of the contextual usages of words or meaning similarities among word and text segments. Second, it can be used as a model of the cognitive processes and representations underlying substantial portions of the acquisition and utilization of knowledge. In the present dissertation, it is mainly used as a tool for extracting the contextual-usage meaning of words in order to assess the context quality for incidental L2 vocabulary learning; however, this approach will support the assumption of the LSA theory and the usage-based model if LSA values are correlated with the learners' vocabulary learning performances from reading. Therefore, this section will outline the ways to apply LSA to text processing and language learning based on previous research that is reviewed below.

Text processing. Researchers have provided theoretical and empirical evidence that semantic similarity between text elements influences the establishment of coherent connections among text elements, which are required for successful text comprehension. In the Construction-Integration model (Kintsch, 1998), the semantic relatedness of words is generated and networked into larger units of text elements during the construction of a textbase representation. Wolfe et al. (2005) showed that higher LSA similarity between a sentence and the prior discourse facilitated the processing and retention of the sentence.

This effect was also found at the word-level processing; a subsequent word is processed faster and retained if it has high semantic relatedness with a prior word (Till, Mross, & Kintsch, 1988). In L2 reading comprehension, Nahatame (2012) demonstrated that LSA similarity affected less-skilled learners' gap-filling ability in local discourse. Particularly, when the word-sentence semantic similarity at the sentence level was higher, they were likely to give correct answers in a cloze completion test. According to Till et al. (1988), these results suggest that the construction of word meanings and their relations during text comprehension are operated by semantic activation based on the semantic similarity between a focal word and the prior context.

The semantic activation process of words contributes to anticipation in language comprehension (e.g., Kutas & Hillyard, 1980). Using LSA, Wang et al. (2010) reanalyzed the difference in the LSA similarities between the predictable and unpredictable words used in Rayner, Ashby, Pollatsek, and Reichle (2004), categorized based on a cloze probability in a sentence (i.e., the strength of contextual constraint or word predictability). The results showed that the LSA similarities were significantly higher between the prior sentence (e.g., *Most cowboys know how to ride a*…) and predictable words (e.g., *horse*) compared to unpredictable words (e.g., *camel*), with a medium effect size[2] (r = .35).[3] Furthermore, their eye-tracking experiment provided evidence that the LSA similarity was related with the identification of upcoming words. In other words, L1 readers are so sensitive to word-sentence semantic similarity that they can anticipate the meaning of upcoming words in discourse comprehension. Given that the upcoming words in discourse are unknown until the readers visually process them, these findings can be applied to the generation of lexical inference (Bolger et al., 2008; Borovsky et al., 2010; Elgort et al., 2015; Hamada, 2012).

---

[2] In this dissertation, effect sizes are interpreted based on Plonsky and Oswald (2014).
[3] The effect sizes were calculated by the author based on the t value and its degree of freedom reported in Wang et al.'s (2010, p. 1377) t test.

Unlike the strength of contextual constraint as defined by a cloze probability, the advantage of using LSA values in the present research is that it should reflect the mental process involved in text processing and word learning. On the other hand, the degree of contextual constraint calculated by the means of a cloze test will differ according to some individual differences among learners (e.g., L2 proficiency). Nation and Webb (2011) claimed that this technique vitiates ecological validity, relating to how well the findings from experiments reflect real classrooms. Although the use of context in L2 lexical inference has been examined in terms of its availability (e.g., Hamada, 2011; Hasegawa, 2012; Nakagawa, 2006; Ushiro et al., 2013; Wesche & Paribakht, 2010), it is important to clarify whether and how L2 learners use word-context semantic similarity to generate the meanings of unknown words during text processing.

Language learning. As the LSA theory is a knowledge representation model of human language, it has been applied to the exploration of text processing because reading comprehension is a knowledge-based activity. However, compared to the application of LSA for text processing, the number of LSA studies on language learning is limited for both L1 and L2. Among the limited research, one approach is the corpus analysis, used in Crossley et al. (2008), in which particular elicitation tasks (e.g., free conversation) are given to individual participants and the spoken data from each participant are gathered over the course of a year. In their study, LSA similarity was seen as a lexical knowledge quality in terms of the ability to produce semantically coherent speech between text segments. In other words, a significant increase in the LSA values reflects the participant's lexical growth because of long-term exposure to input. Although Crossley et al. showed that the semantic networks of L2 words as represented by LSA developed as learners progressed in their SLA, this approach does not account for how the learners acquired L2 words from contextualized input.

In another example, Inohara and Kusumi (2012) applied the LSA theory to predict the effects of reading habits on L1 vocabulary growth. Specifically, the experiment tested whether the response frequency and patterns of an L1 word association test are predicted by the LSA values computed from either a newspaper-based corpus or a novel-based one. The rationale behind their examination was that if the words input into an existing mental lexicon are classified as either newspaper-based or novel-based based on the participants' reading habits, the patterns of word association should vary according to co-occurrence information between words stored in their mental lexicon (*refrigerator*$_{\text{stimulus}}$ → *electrical appliances*$_{\text{response}}$ in newspaper-based knowledge vs. *refrigerator*$_{\text{stimulus}}$ → *open*$_{\text{response}}$ in novel-based knowledge). This shows the causal relationships between reading activities and vocabulary acquisition predicted by the usage-based model; however, it is not yet clear why and how LSA similarities are correlated to incidental vocabulary learning, particularly in an L2. Specifically, it is necessary to examine how L2 learners use the word-context semantic similarities computed by LSA to extract word information while reading and to determine if the LSA similarities predict the outcomes of incidental L2 vocabulary learning from context.

2.3 Memory Representation Construction in Reading

The incidental acquisition of new words during reading begins with establishing a memory representation of those words through inference from a context (Bolger et al., 2008; Bordag et al., 2015; de Bot et al., 1997; Elgort et al., 2015; Hulstijn, 1992; Paribakht & Wesche, 1999). Inference generation is one of the central reading processes in building a text memory (e.g., Graesser et al. 1994; Kintsch, 1998; McKoon & Ratcliff, 1992). In the processing of new words, an inference can be defined as meaning generation (Bordag et al., 2015), or deriving word meaning from a context and representing it in the reader's text memory. Despite the important role of meaning generation in incidental vocabulary learning, the relationships between its cognitive processes and products, including newly learned L2 lexical knowledge, have not been clarified (Bordag et al., 2015; Nassaji, 2006; Paribakht & Wesche, 1999; Pulido, 2007). Accordingly, I will provide an overview of the relationships between language processing and memory as its consequences.

2.3.1 Reading Comprehension and Text Memory

A standard goal of comprehending a text is to construct a mental representation of the text (e.g., Graesser et al., 1994). More importantly, the text memory constructed from discourse can be retained as linguistic knowledge (e.g., Wesche & Paribakht, 2010) as well as world knowledge such as concepts and principles (e.g., van den Broek, 2010). However, establishing a verbatim mental representation is not identical with learning from a text because it does not imply "that one is able to use the information provided by the text in other ways, not just for reproduction" (Kintsch, 1994, p. 294). Similarly, word learning from texts should be beyond simply constructing a memory of the text. So far, researchers have tried to connect the existing reading models to the nature of vocabulary acquisition (e.g., Bordag et al., 2015; Elgort et al., 2015; Horiba & Fukaya, 2006, 2012, 2015; Nassaji, 2003, 2006; Pulido, 2007). This section presents an overview of the relationships between text processing and memory in order to interweave them with L2 vocabulary acquisition through text comprehension.

Multilevel mental representations. In the time course of reading comprehension, semantic information is continuously extracted from individual words in a text and encoded in the mind in order to form a mental representation (Kintsch, 1998). Encoding a text is equal to comprehending it, but the form of mental representations is multilevel, ranging from the surface level to the deepest understanding (Fletcher, 1994; Kintsch, 1988, 1994, 1998; van Dijk & Kintsch, 1983). Regarding the knowledge acquisition of concepts from texts, reading researchers agree that readers achieve the construction of a *situation model*, in which text memories of events are integrated with their background knowledge (e.g., Horiba & Fukaya, 2006, 2012, 2015; Kintsch, 1994; Ushiro et al., 2015; van den Broek, 2010). This ultimate level of mental representations is distinguished from a *propositional textbase* in that the meaning of the text per se is extracted and interconnected within a network of text propositions, because text ideas can be recalled from a textbase memory, but simple reproduction of a text does not mean learning. While semantic information extracted from every word is represented in these levels of mental representations, the most superficial mental representation (*surface form*) only preserves the exact words and syntactic structures in the text. Figure 2.6 gives a schematic example of a multilevel mental representation.

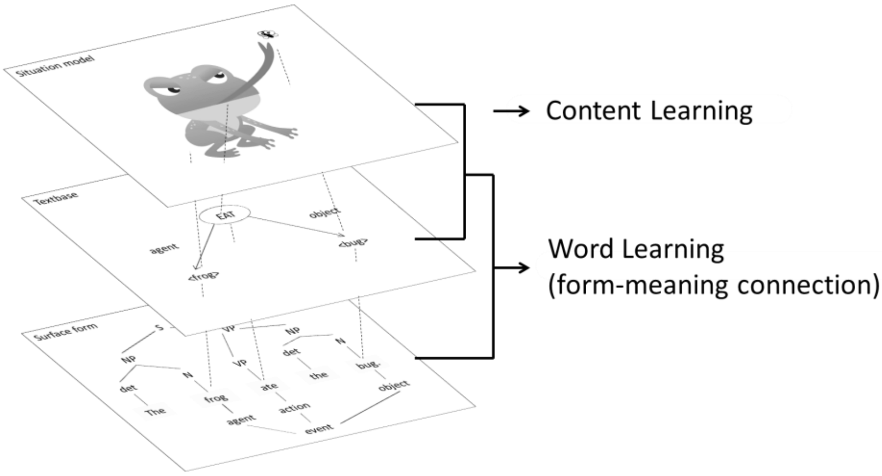


Figure 2.6. Three levels of mental representations (adapted from Fletcher, 1994, p. 590).

Learning from a text requires readers to build a complete and elaborate situation model; however, it is a more complicated matter in word learning. The task of word learning is to establish robust form-meaning connections. If learners can comprehend the meaning of an unknown word from a context, they then have to associate the meaning with the focal word form represented in memory at the same time as shown in Figure 2.6 (Leung & Williams, 2011; van Patten et al., 2004). However, Kintsch's model suggests that surface memory of elements such as letters, words, and syntactic constructions is abstracted and not maintained after successful text comprehension because it is characterized as the maintenance of semantic representations converted from lexical and syntactic information. For example, research in L1 reading provided evidence that verbatim text memory can be fragile compared to its semantically abstracted representation based on readers' sentence recognition memory (Fletcher & Chrysler, 1990; Kintsch, Welsch, Schmalhofer, & Zimny, 1990). These findings point to the assumption that the memory of a string of letters in unknown words will steadily decay when the word meanings are generated and encoded in mental representations because readers are no longer required to retain the word-form memory. Thus, the consequences of discourse comprehension suggest that the link between a word form and its meaning is not easily established while reading.

Empirical research provided further evidence that the relative accessibility of word-form memory differs by whether readers can establish the semantic memory of the words or not. Gerrig and his colleagues assumed that when readers encounter something unknown in a text, they expect that the unknown information will be revealed as the text unfolds; therefore, the related information is highly activated in memory (Gerrig, Love, & McKoon, 2009; Love, McKoon, & Gerrig, 2010). This assumption was examined using a probe recognition test, in which L1 participants read stories and then responded as swiftly and accurately as possible to the question of whether a character name that flashed on a computer screen had appeared in the stories. The name belonged to a protagonist, but his or her role in the stories remained unclear. The results showed that recognition latencies to the target words were shorter when the character's name appeared without mentioning his or her narrative function than when the character's role was clarified by subsequent context.

Similarly, Dopkins and his colleagues demonstrated that recognition latencies to the non-antecedents of an anaphor were slower after readers processed the anaphor than before they did (Dopkins & Ngo, 2005; Dopkins & Nordlie, 2011; Nordlie, Dopkins, & Johnson, 2001). As stated above, Fletcher (1992, p. 199) explained that "the generally accepted explanation for [these findings] was that the surface form of a sentence is held in short-term memory until its meaning is understood, and then only the meaning is stored in long-term memory." Also, Gernsbacher (1990) described that as the construction of mental representations proceeds through the text, the available information represented in working memory is continuously refreshed, with some information remaining the focus of attention and other information being suppressed.

Although the differences in relative accessibility to unsure information in narratives are not simply applicable to establishing form-meaning connections in word learning, the past studies suggest that a memory representation as a consequence of text processing is closely related to subsequent learning. In other words, the processing of a text to form its representation in memory ultimately produces a certain outcome of learning, so that it is necessary to integrate general theories of discourse processing with its consequences such as learning from texts.

Constructing mental representations for learning. The language processing in the context of learning from texts requires the integration of comprehended information into the readers' prior knowledge (e.g., Kintsch, 1994, 1998; van den Broek, 2010). Text representations are multidimensional, consisting of lexemes, words, syntactic structures, and propositions (Graesser, Millis, & Zwaan, 1997), so that corresponding knowledge interacts from lower-level to higher-level language processing despite the differences between L1 and L2 reading (Grabe, 2009; Horiba, 2000; Koda, 2005).

One of the most essential cognitive processes in learning from texts is to modify an existing knowledge structure (e.g., McNamara, Kintsch, Songer, & Kintsch, 1996; Rapp, 2008; Rapp & Kendeou, 2007, 2009; van den Broek, 2010) by connecting individual text elements with meaningful relations and incorporating them into background knowledge.

These types of processing are generally supported by inference generation (Kintsch, 1994, 1998) because the roles of inference are to bridge the gap between different parts of the text (e.g., Gernsbacher, 1990), to connect known with unknown information in the text in order to achieve a coherent mental representation (e.g., Graesser et al., 1994; McKoon & Ratcliff, 1992), and to integrate the text with readers' prior knowledge (e.g., Kintsch, 1988, 1994; van den Broek, 2010). Given the importance of inference generation in learning from texts, the types of processing should also play a significant role in learning words from reading.

2.3.2 Lexical Inference and Word Memory

Although comprehending a text always starts with the identification of the first words that appear in a sentence (e.g., Just & Carpenter, 1980), L2 learners often encounter unknown words due to their low quality or lack of word knowledge (Hu & Nation, 2000; Qian, 2002). These words create gaps in learners' comprehension of the text meaning and prevent them from understanding the explicit contents of the text. To close the gaps in text comprehension created by unknown words, theoretically, L2 learners have to infer the possible meanings of unknown words from a given context. In this section, I will review the previous research that examined how word memory can be integrated into the mental lexicon during L2 reading.

Word memory established by inference. Inferential information activated during reading can be encoded in long-term memory (Fincher-Kiefer, 1995, 1996; Klin, 1995; Klin, Murray, Levine, & Guzmán, 1999; Ushiro et al., 2012). According to Chaffin (1997), the inferred meaning of unknown words typically includes information about either their synonyms or hypernyms. For example, the lexical inferences activated by the sentence *The surfers were attacked by a dangerous sind in the sea* might be that "shark" is a synonym of *sind* (a target pseudoword) or that "fish" is a hypernym of *shark*. Following a hierarchical lexical network between hyponyms (e.g., *shark*) and hypernyms (e.g., *fish*), the former has more semantically specific characteristic than the latter (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990; Rosch et al., 1976). Thus, the generation of synonymous meanings can be defined as a *specific lexical inference,* while a superordinate inference can be regarded as a *general lexical inference*. The specificity of lexical inference is affected by readers' proficiency. For example, Fukkink, Blok, and de Glopper (2001) demonstrated that children with lower L1 proficiency than adults could not narrow inferable meanings down to specific ones. The children's lexical inferences typically produced the widely interpreted meaning of unknown words, which easily matched contextual information in meaning.

Many SLA researchers have examined whether the word memory built by lexical inference is specific or not, by asking L2 learners to infer the meanings of unknown words from the information contained in contextual messages. The results indicated that their meaning identification is likely to be unsuccessful (Bensoussan & Laufer, 1984; Frantzen, 2003; Huckin & Bloch, 1993; Hulstijn, 1992; Hulstijn et al., 1996; Laufer, 1997; Mondria & Wit-de Boer, 1991). A critical problem that inhibits contextualized word learning from lexical inferences is inadequate comprehension of context due to their limited cognitive capacity available in L2 processing. For example, Griffin (1992), Hasegawa (2013), and Ma, Chen, Lu, and Dunlap (2015) showed that context contributes to vocabulary gains for skilled L2 learners but not for less-skilled learners. Laufer and Shmueli (1997), Prince (1996), and Webb (2007a) used short and simple sentences to define the usage of target words; however, their results were the same as the aforementioned studies, concluding that contextualized learning is not superior to decontextualized one. Remarkably, although Webb (2007a) examined various aspects of lexical knowledge that seemed more likely to be acquired through context (i.e., grammatical functions, syntagmatic association, and paradigmatic association) as well as form-meaning knowledge, a single glossed sentential input had little effect on vocabulary knowledge acquisition compared to paired-associate learning. These findings suggest that the comprehensibility of context is not a unique factor affecting the learning of any lexical knowledge of words.

Another possibility is related to the influences of vocabulary learning strategies used by individual learners. Many researchers have suggested that the learning outcomes of new words are reflected in the strategy types preferred by each learner (e.g., Gu & Johnson, 1996; Mizumoto & Takeuchi, 2009; Nation, 2013). Such aptitude-treatment interaction may also determine the results of incidental L2 vocabulary learning from text. For example, Hamada (2011) demonstrated that what aspects of word memory are built from inferences can be attributed to the learner’s strategy. Specifically, the strategy using word structure led to the learning of word forms only, whereas the good use of context during lexical inference promoted gains in word meaning and syntactic knowledge. Thus, it is necessary to consider the possible interaction between context and strategy influences the outcomes of incidental L2 vocabulary learning through lexical inferences.

More importantly, although past studies have postulated that context may play an important role in constructing word memory traces as predicted in the usage-based model, prior studies could not establish its effectiveness because of deficiencies in the research designs. Whereas the assumption is that contexts are conductive to the lexical properties of a target word, the meaning of an unknown word is transparent in some sentences but it will be opaque in others (Huckin & Bloch, 1993; Webb, 2008). Nevertheless, prior studies did not consider the quality of individual contexts and could not make an accurate assessment of those effects (Nation & Webb, 2011). To reinforce incidental L2 vocabulary learning from context, therefore, it is important to evaluate and manipulate how well a given context functions. This perspective must be associated with the process of lexical inference (see the next section) and the LSA theory based on the usage-based model.

Lexical processing by inference. The other approach to examining incidental L2 vocabulary learning is the development of cognitive models of lexical inference to predict and explain learning outcomes as a consequence of a particular process. First, de Bot et al. (1997) provided lemma construction, which is a cognitive process model for meaning specification, focusing on the role of declarative and procedural components used in lexical inference. In this model, when L2 readers are trying to comprehend a text, the word identification process starts by matching visual inputs with the lexeme (i.e., phonological and orthographical knowledge), which activates lemma information (i.e., syntactic and semantic knowledge) stored in their mental lexicon. Thus, novel lexemes make the readers notice that they do not know a corresponding lemma of the word. To fill in the gaps of their knowledge about the word, they attempt to find and use various knowledge sources from contextual information and prior background knowledge; simultaneously, the word information assembled from the knowledge sources can be interlinked as vocabulary knowledge (Wesche & Paribakht, 2010).

Similar to lemma construction, Huckin and Bloch (1993) conceptualized the lexical inference processes from their case study (see Figure 2.7). In their cognitive processing framework, L2 learners monitor their cognitive processes in meaning generation through generator, evaluator, and metalinguistic modules. When they encounter an unknown word in a text, they try to generate hypotheses about its possible meaning and verify the validity of inferences based on their linguistic and nonlinguistic knowledge. Moreover, they meta-cognitively control the processes involved in decision making while finding contextual information, making inferences, and evaluating the outcomes.

Both models show that when L2 learners are not able to collect enough information to generate the meaning of unknown words, they try to look for additional contextual information. Others have supported this assumption and concluded that lexical inferences, which are problem-solving types of processing, are so consciously deliberate that they can be neither smooth nor successful (e.g., Bensoussan & Laufer, 1984; Haastrup, 1991;

Nassaji, 2006; Paribakht & Wesche, 1999). In other words, there are a great number of potential difficulties that L2 learners have to deal with when they are exposed to a new word in reading comprehension. However, as described in the previous section, earlier SLA research did not focus much on the contextual factors affecting the process of lexical inference.

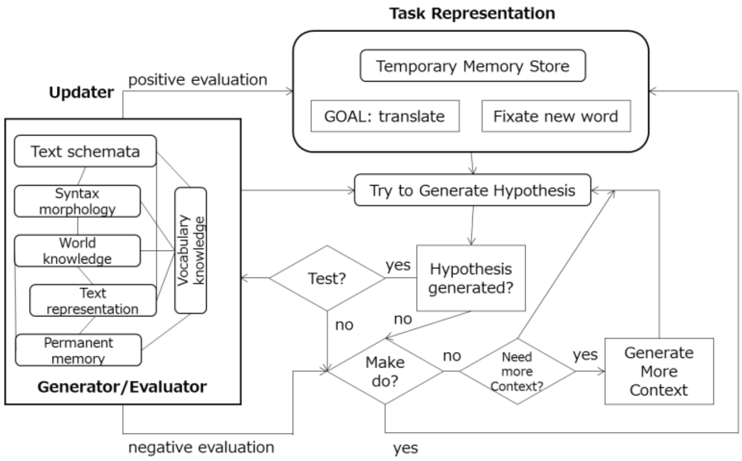


Figure 2.7. Cognitive processing framework of L2 meaning generation (adapted from Huckin & Bloch, 1993, p. 170). Whereas positive evaluation indicates that readers rated their generated hypothesis about the word meaning as appropriate, negative evaluation indicates the case where the hypothesis was rejected.

On the other hand, psycholinguistic research demonstrated that a highly contextual elaboration allows readers to anticipate a specific upcoming word during comprehension in both L1 (Camblin, Gordon, & Swaab, 2007; DeLong, Urbach, & Kutas, 2005; Otten & van Berkum, 2008, 2009; Szewczyk & Schriefers, 2013; van Berkum, Brown, Zwiserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004) and L2 (Martin et al., 2013; van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011). In Otten and van Berkum (2008), after reading a prior context, such as "Sylvie and Joanna really feel like dancing and flirting tonight," participants were implicitly tested on anomalous word detection in a subsequent sentence (e.g., "Therefore they go to a *stove* [*disco*]..."). In this case, the prior context semantically elaborated the possible meaning of upcoming words (i.e., *disco*); thus, the participants quickly detected anomalies through context-based word anticipation. Although some researchers have regarded anomaly detection as a sign of the difficulty of integrating words into mental representations, most psychophysiological studies have suggested that readers are using context to generate predictive inferences for an upcoming word meaning (DeLong et al., 2005; van Berkum et al., 2005).

Focusing on the processing of unknown words, Borovsky et al. (2010) examined contextual elaboration effects. They compared the comprehensibility of a nonword (e.g., *marf*) after presenting two types of prior contexts ("He tried to put the pieces of the broken plate back together with *marf*" vs. "She walked across the large room to Mike's messy desk and returned his *marf*"). Processing the word *marf* was facilitated by the former compared with the latter sentence, providing evidence that the prior context promoted the generation of lexical inference that the word *marf* referred to *glue*. This result suggests that elaborative contexts activate the meaning of the upcoming words, and the activated meaning is quickly integrated into the word form. In fact, L2 reading research also has demonstrated that meaning-oriented cues in the same sentence as the target unknown word are of primary importance (e.g., de Bot et al., 1997; Nassaji, 2006; Paribakht & Wesche, 1999; Wesche & Paribakht, 2010). Taken together, it is hypothesized that lexical inference using highly elaborative information is readily available to L2 learners while they are reading a text in L2.

Evidence suggests that the contextual elaboration of a prior context (hereafter, *forward contextual elaboration*) facilitates lexical processing by anticipation. However, in the context of lexical inferences, unknown words are not always elaborated by a prior context. That is, readers often encounter an unknown word that is semantically elaborated by a subsequent context (hereafter, *backward contextual elaboration*). Let us consider the sentence “Joe picked up the *asdor* and began to strum a tune” (Chaffin, Morris, & Seely, 2001, p. 226). In this example, the likely meaning of nonword *asdor* (i.e., *guitar*) is semantically elaborated by the contextual information, *began to strum a tune*, but not the prior context *Joe picked up*. To successfully comprehend this sentence, readers have to activate the meaning of *asdor* by integrating the subsequent information with the under-constructed mental representation that includes the unknown word. Using eye-tracking measures, Chaffin et al. (2001) found that L1 adult readers frequently made regressive eye-movements when an elaborative context was a primary cue for the meaning of target words. This suggests that the readers were so sensitive to the contextual elaboration that they were able to integrate the contextual message, *began to strum a tune* with the inferable meaning of the word. In other words, the meaning of *asdor* could be represented in the mind after backward elaborative contexts were added to readers’ text memory.

Nevertheless, backward lexical inferences may be difficult in even L1 reading, especially when language skills (Cain et al., 2004) and cognitive capacity (Daneman & Green, 1986) are constrained. In particular, these studies suggest that the limited capacity of working memory affects backward lexical inferences because readers must keep an encountered unknown word in working memory until highly elaborative contexts appear. Huckin and Bloch (1993) also demonstrated that if L2 learners had not yet collected enough contextual information to infer the meaning of an encountered unknown word, they often skipped over it. In this case, Hulstijn (1993) showed that learners are very afraid of poor text comprehension, and wonder as to what kind of thing the words refer. Thus, we can infer that the effect of contexts on lexical inferences may vary, but it is possible that a think-aloud task overestimates learners’ use of contextual information.

2.4 Noticing Hypothesis

So far, I have overviewed the memory representation construction of words in L2 reading and its positive relationship to incidental gains on word knowledge. Nevertheless, evidence for L2 vocabulary learning from reading run counter to Krashen's (1989) input hypothesis that most parts of an L2 can be acquired unconsciously. SLA research explores whether L2 can be learned by simple exposure to it or not (e.g., Hama & Leow, 2010; Leung & Williams, 2011, 2012), and recent studies have highlighted the role of conscious attention in input processing for learning (Robinson, 1995).

The role of attention in SLA was proposed by Schmidt's (1990, 1994, 1995, 2001) noticing hypothesis; attention, at one of the levels of noticing, is necessary, or at least a facilitator to integrate input into representations of learners' language system (i.e., intake). Regarding incidental word learning in both L1 and L2, meta-analysis studies indicate that explicitly deriving word meaning promotes larger gains in lexical knowledge (Fukkink & de Glopper, 1998; Swanborn & de Glopper, 1999) compared with when learners' attention is not directed to target unknown words (e.g., Huang, Eslami, & Willson, 2012).

Although the techniques to quantify attention are disputable (Godfroid, Boers, & Housen, 2013), psycholinguistic research often applies eye-tracking measures to capture the allocation of attentional resources in text comprehension (Rayner, 1998, 2009). In L2 vocabulary acquisition research, Godfroid et al. (2013) assumed from the attention-memory model proposed by Robinson (1995), that the initial representation encoded in memory becomes robust through additional and time-consuming processing such as cautious elaborative rehearsal. In their eye-tracking session, Dutch-speaking EFL learners read paragraph-unit texts that contained a target unknown word. Immediately afterward, the individual sentences from the texts, each with a target word deleted, were presented, and the participants were asked to select a correct word form from multiple options to fill in the blank. A multiple regression analysis showed that the total fixation durations in processing the target words explained the incidental gains in word meaning knowledge. First fixation and gaze duration were not strong predictors, but they positively correlated with the learning gains.

Additionally, Williams and Morris (2004) showed that the increased time of second fixation durations (i.e., reanalysis of target words) had positive effects on the recognition of novel words. As the second fixation durations for words are regarded as readers' semantic integration between word and context (Rayner, 1998), Godfroid et al. (2013) explained that linguistic items to which learners pay more attention are further enhanced by elaborative rehearsal in working memory. Pellicer-Sánchez (2015) further examined the relationships between fixation times and vocabulary test scores, showing that longer fixation times were associated with meaning recall but not form and meaning recognition. Because the meaning recall test requires the establishment of form-meaning links, these sets of past studies suggest that attention plays an important role in finding the rules of form-meaning connections.

While noticing and attention allocation will become a facilitator in incidental L2 vocabulary learning, it should be noted that L2 learners' cognitive capacity is limited in the processing of a text (e.g., Grabe, 2009; Horiba, 2000; Koda, 2005). Therefore, learners' allocation of their attention resource is often manipulated by text factors such as text difficulty. One assumption is that reading easy texts helps learners to pay attention to target words while they are trying to comprehend the whole content of the texts (Pulido, 2007). The other is that when texts are syntactically complex, readers may need to take up every single word in order to construct a coherent mental representation (Bordag et al., 2015). Early studies provided controversial results for the first assumption; we have not yet obtained a research consensus on whether texts that are easy to read in terms of lexical inference lead to incidental vocabulary gains (Li, 1988) or not (Mondria & Wit-de Boer, 1991; Pulido, 2007). For example, in think-aloud protocols, learners ignored and skipped over unknown words because the easy texts were comprehensible even if those words remained unknown (Fraser, 1999; Huckin & Bloch, 1993; Paribakht & Wesche, 1999). In contrast, consistent results were obtained from reaction time (RT) and eye-tracking research, showing that L2 learners directed their attention to words whose meaning was difficult to infer, which predicted the subsequent retention of word knowledge (e.g., Bordag et al., 2015; Godfroid et al., 2013; Pellicer-Sánchez, 2015). Thus, the necessity of input processing and additional elaboration increases the amount of attention allocated by L2 learners (e.g., Ellis, 1994; Hulstijn & Laufer, 2001; Laufer & Hulstijn, 2001).

As further examples, learners' attention to different aspects of word features is controlled using lexical processing strategies (Barcroft, 2002; Hamada, 2011) and tasks (Barcroft, 2003, 2009; Horiba & Fukaya, 2006, 2012, 2015; Wesche & Paribakht, 2000). First, Barcroft's (2002) type of processing-resource allocation (TOPRA) model predicts that the differential allocation of limited attention resources results in knowledge gains in different aspects, as visualized in Figure 2.8. The heavy outer lines in this model represent the limitation on the overall attention resource per learner. The inside lines may move as the amount of attention allocated for a particular processing increases or decreases; thus, the gains in corresponding aspects of word knowledge can be promoted. For example, the more attention resources to semantic processing learners allocate, the larger the space for semantic learning is. As the amount of processing of word forms and mapping decreases, the gains in word knowledge and form-meaning mapping become smaller.

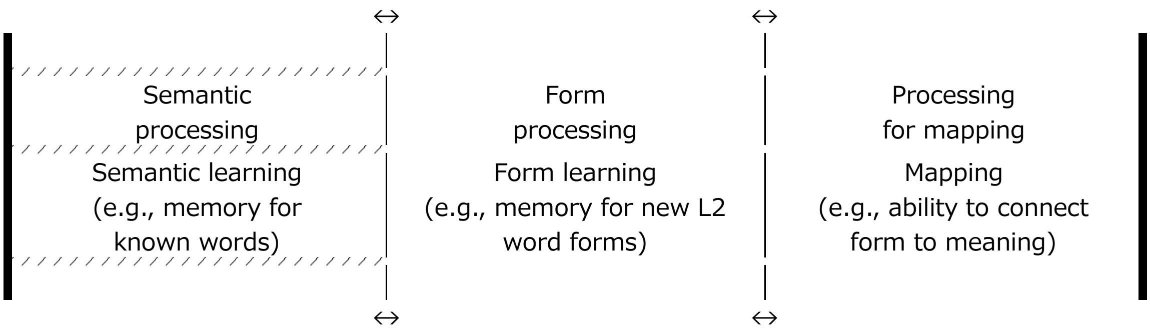


Figure 2.8. Relationships between language processing type and learning in the TOPRA model (adapted from Barcroft, 2003, p. 549).

Regarding the different effects of lexical processing strategies on incidental word learning, Barcroft (2002) required Spanish ESL learners to intentionally learn new words through a semantic elaboration task or structural elaboration task. The result was consistent with his TOPRA model because it showed that the semantic elaboration task reduced the recall production of target word forms. As described in Section 2.3.2, Hamada (2011) also indicated that different inference strategies led to the learning of different aspects of vocabulary knowledge. Using a think-aloud method, he classified the lexical inference strategies used by Japanese EFL learners into the following three types: word forms, background knowledge, and word forms and background knowledge. The results of a free word association test showed that whereas word-based association was produced by the learners who used the word form strategy (e.g., $\textit{ware}_{\text{stimulus}} \rightarrow \textit{wear}_{\text{response}}$), knowledge-based association was found by those who used background knowledge (e.g., $\textit{ware}_{\text{stimulus}} \rightarrow \textit{truck}_{\text{response}}$). Thus, the aspects of lexical knowledge acquired depend on the type of lexical processing strategies used.

Barcroft (2009) demonstrated that the TOPRA model is applicable to incidental L2 vocabulary learning in an experiment of reading with certain tasks. In his experiment, Spanish ESL learners were read a text with 10 target words for comprehension in either an incidental condition or an incidental-semantic condition. While the participants in the incidental condition (i.e., control group) read the text for meaning only, their counterparts in the incidental-semantic condition (i.e., a semantic elaboration task) generated a synonym of the target words in their L1 as they read. Next, L1-to-L2 and L2-to-L1 translation tests, both of which required form-meaning connections of target words, were implemented, but word form knowledge was assumed to be relatively important in the L1-to-L2 test because the participants had to reproduce the spellings represented in their memory. The results showed that both test scores were higher in the incidental condition than in the incidental-semantic condition, and the negative effects of semantic processing were large in word form learning ($d = 0.88$) and the building of form-meaning connections ($d = 1.21$).[4] These findings were replicated by Kida (2010) with Japanese EFL learners, where a structure elaboration task was added as a factorial design, supporting the prediction of the TOPRA model.

The other strategies[5] to control L2 learners' attention allocation have been studied by Hulstijn and colleagues, who employed (a) multiple-choice glosses (MCG: Hulstijn, 1992; Hulstijn & Laufer, 2001) and (b) dictionary use (Hulstijn, 1993; Hulstijn et al., 1996). MCG promotes the cognitive processes of evaluation, or hypothesis testing of word meaning inferred from context (Hulstijn & Laufer, 2001). The effect of MCG was also explained by Rott (2005); this technique elaborates the inferred meaning of new words by directing learners' attention to the word itself. Similarly, dictionary use helps such evaluation and elaboration. Hulstijn (1993) showed that L2 learners consulted a dictionary even after they inferred the meaning of target words. Moreover, evaluating and elaborating the meaning of unknown words represented in mind facilitated word meaning learning (e.g., Hulstijn et al., 1996).

In the field of SLA, many researchers provided evidence of the role of noticing and attention allocation in vocabulary learning as reviewed so far. Nevertheless, these studies did not inform "on the nature of the relationships that exist among [the] comprehension, intake, and incidental acquisition" of new words (Pulido, 2007, p. 156). For example, although Pellicer-Sánchez (2015) concluded that the amount of attention predicted form-meaning links only, it remains unclear why attention allocated in text comprehension did not improve form and meaning recognition performance.

---

[4] The effect sizes were reanalyzed by the author based on the descriptive statistics reported in Barcroft (2009, p. 95, Table 2).

[5] Given that the definition of task: (a) The primary focus is on meaning in communicative situations, (b) Real-world cognitive processes in language use are required, and (c) A communicative outcome is produced (e.g., R. Ellis, 2008), the MCG and dictionary use as well as the synonym generation task used in Barcroft (2009) and Kida (2010) cannot be regarded as exact tasks in SLA research. However, in practice, these activities are called *task* in this dissertation.

Regarding what nature of mental representations of new words can be constructed through task-induced reading processes, Horiba and Fukaya (2006, 2012, 2015) implicitly required L2 learners to establish different mental representations using an L1 recall task (*meaning-focused* condition) and L2 recall task (*form-focused* condition). In the meaning-focused condition, the participants needed to direct their attention to encoding discourse propositions for the subsequent recall task by regulating their plans and strategies to recall while comprehending a text. In turn, attention had to be excessively directed toward a surface form of the text in the form-focused condition because the participants had to recall the text contents in the L2. The results of a vocabulary test showed that the form-focused condition promoted the incidental learning of word meaning as well as word form compared to the meaning-focused condition. Other studies on focus-on-form tasks also showed positive results that the establishment of form-meaning connections is improved through attention allocation to language form by a task while learners are comprehending given input (de la Fuente, 2006; Hulstijn & Laufer, 2001; Laufer, 2006). These findings suggest a strong relationship between L2 vocabulary acquisition and discourse processing in cognition, as illustrated in Figure 2.9.

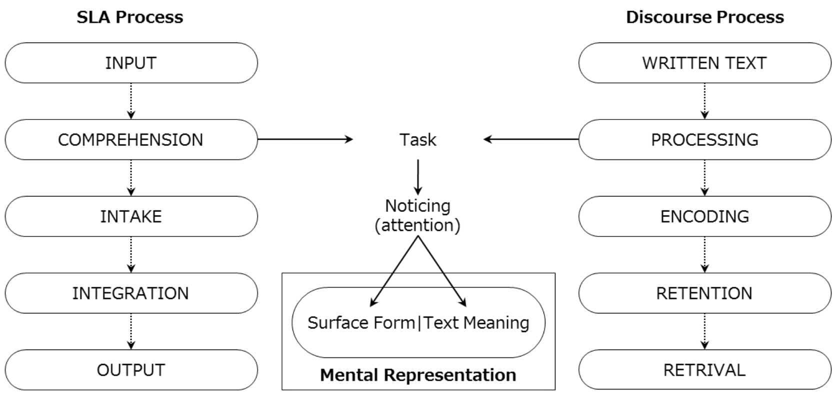


Figure 2.9. The possible relationships between the theories underlying the SLA process and discourse processing. Noticing as a facilitator driving the process of L2 vocabulary acquisition (adapted from Gass et al. 2013).

However, there are some limitations in the previous studies in that they provided word definitions in the L1 as glosses during the reading phase. Both L1 (Bolger et al., 2008; McKeown, 1985) and L2 (Griffin, 1992; Hasegawa, 2013; Webb, 2007a) studies indicated that word definitions can complement the lack of contextual experience to infer the meaning of unknown words but they do not improve the accuracy of using the word in a sentence. Although the establishment of form-meaning connections is a basic goal of vocabulary learning, it is necessary to take a close look at the role that contextualized incidental learning plays in vocabulary development, because decontextualized deliberate learning usually contributes to increasing form-meaning links (e.g., Nation, 2013; Schmitt, 2010). Also, from a pedagogical perspective, L1 glosses are not always available in L2 reading, particularly outside classrooms (e.g., Wesche & Paribakht, 2010).

2.5 Research Methods in the Inference, Encoding, and Learning of Words

Despite of the large amount of research on incidental L2 vocabulary learning and its replications, the methodologies used to explore the generation, encoding, and learning of the contextual-usage meaning of words were varied and sometimes incongruent with the research purposes. For example, traditional recall and recognition tests are not able to measure the initial knowledge represented in an under-constructed mental representation (Bordag et al., 2015; Borovsky et al., 2012; Perfetti, Wlotko, & Hart, 2005). Also, claims about the cognitive processes involved in lexical inference have relied heavily on learners' protocol data elicited by think-aloud procedures (Deschambault, 2012). In this section, I will present an overview of the methodological paradigms in terms of the indices of inference, encoding, and learning of unknown words achieved through reading a text.

2.5.1 Think-Aloud Protocol and Alternatives

The cognitive processes of lexical inferences during reading have been examined by various research methods. In L2 reading, many studies used a *think-aloud* method to verbalize the learners' thoughts while attempting to generate the meaning of unknown words from contexts (e.g., Bensoussan & Laufer, 1984; de Bot et al., 1997; Fraser, 1999; Haastrup, 1991; Hamada, 2011; Huckin & Bloch, 1993; Nassaji, 2006; Paribakht & Wesche, 1999; Wesche & Paribakht, 2010). The protocol elicited by a think-aloud task has been analyzed to categorize the types of knowledge sources used by L2 learners when generating word meanings. Previous studies showed that the most important knowledge source is a meaning-oriented cue embedded in the same sentence as the target word rather than a word-based cue embedded in an unknown word itself such as word-form analogy and morphology (e.g., Fraser, 1999; Nassaji, 2006; Wesche & Paribakht, 2010). In other words, the immediate informative linguistic cues of an unknown word are the first place L2 learners strategically look for information about its meaning.

Although the procedure is widely accepted as a method for observing learners' cognitive processes as they complete a task, there is debate on whether thinking aloud is a useful tool for grasping on-line processes (e.g., Ericson & Simon, 1980; Magliano, Trabasso, & Graesser, 1999). According to Ericson and Simon (1980), an appropriate think-aloud procedure requires readers to verbalize only thoughts that are consciously available during reading. This suggests that think-aloud protocols reflect what kinds of information are available to working memory, accessible to consciousness, and verbalizable. Because of the nature of the think-aloud methodology, Magliano et al. (1999) suggested using a combination of think-aloud protocols and RT measurements to study on-line comprehension processes. For example, the think-aloud methods are limited to gaining information on the passive processes of reading comprehension that cannot be verbalized. In addition, Deschambault (2012) showed that the protocols elicited by the task are influenced by social factors such as pragmatics. To solve these methodological problems and validate the think-aloud data reported in previous studies, therefore, it is important to collect independent behavioral measures that reflect what thoughts are available during silent reading (Magliano et al., 1999).

From the viewpoint of SLA, Fukuta (2015) argued the potential methodological biases in studying the relationships between attention and learning using the think-aloud methods. Some researchers supported incidental L2 learning without attention (e.g., Leung & Williams, 2011, 2012; using retrospective interview), while others rejected it (e.g., Hama & Leow, 2010; using a think-aloud task); this discrepancy depends on what measures of attention were selected. Given that think-aloud protocols excessively include the conscious reports about what information is available, it is possible to overestimate the outcomes of incidental learning by considering them together with the outcomes of deliberate learning (e.g., Rebuschat, 2013). This discussion highlights the importance of employing a multi-componential and multi-test approach in order to assess the knowledge acquired from reading (Pellicer-Sánchez, 2015). Since the partial knowledge of words being constructed is not stable and robust enough to verbalize (e.g., Borovsky et al., 2012), as an alternative way, more knowledge-sensitive measures should be required to assess the early stage of incidental L2 vocabulary learning (Bordag et al., 2015).

2.5.2 Reaction Time

One of the research methodologies to provide on-line measures of cognitive processes during comprehension is the use of RT data (e.g., Jegerski, 2014; Jiang, 2012; Magliano et al., 1999; van den Broek et al., 1993). RT research explores the difference in RTs according to the stimuli provided in a particular test (Jiang, 2012), instead of measuring the absolute speed in performance on the test. In the selection of a test, following the existing theories and models, for example, of reading comprehension (Magliano et al., 1999; van den Broek et al., 1993), it is essential to consider the rationales of what mental processes are required to complete the test and the logics about what can be explained through the differences in RTs elicited by the test. This section will review these two issues, focusing on the *self-paced reading method*, *semantic related judgment test*, and *probe recognition test*. These tests are relevant to the generation and encoding of lexical inference in language comprehension.

Self-paced reading method. Reading times reflect the cognitive processes of on-line comprehension because the data are collected while participants are comprehending a text (Nahatame, 2015). Therefore, one of the advantages of using this method is that it can rule out the uninteresting interpretation that the word memory detected by a test was built because of the memory reconstruction as a consequence of taking the test. Moreover, this method can be applied to examine whether a text meaning is encoded in memory or not by manipulating testing time (Bordag et al., 2015).

In the self-paced reading method, experimenters record a participant's reading times for each designated segment of a sentence. According to Jegerski (2014), the rationale behind this method is that the inflation and deflation of reading times for targets reflect the interference with and facilitation of comprehension caused by the targets, respectively. For example, van Berkum et al. (2005, Experiment 3) used a noncumulative self-paced reading method (i.e., word-by-word reading) to test context effects on the anticipation of an upcoming word as a text unfolds. Dutch native speakers read the following discourse passages (p. 446):

(2) De inbreker had geen enkele moeite de geheime familiekluis te vinden.

[The burglar had no trouble locating the secret family safe.]

(3a) Deze bevond zich natuurlijk achter een groot$_{neu}$ maar onopvallend schilderij$_{neu}$.

[Of course, it was situated behind a big-Ø$_{neu}$ but unobtrusive painting$_{neu}$.] (consistent)

(3b) Deze bevond zich natuurlijk achter een grote$_{com}$ maar onopvallende schilderij$_{neu}$.

[Of course, it was situated behind a big-e$_{com}$ but unobtrusive painting$_{neu}$.] (inconsistent)

The prior context meaning was manipulated to constrain the predictable meaning of the target noun (i.e., *painting*). Then, the reading times during the processing of the target adjectives (*big*-Ø$_{neu}$ vs. *big-e*$_{com}$) were compared. In the experiments, the inflectional suffix of the target adjectives was (not) morpho-syntactically consistent with the predictable noun (i.e., *painting*$_{neu}$). Therefore, if readers anticipated the upcoming word meanings by inferences, the reading times for the inconsistent adjectives would be longer as a result of an interference effect because the adjectives were completely unexpected for the readers. Although the processing time for the target adjectives did not differ between the two conditions, van Berkum et al. found that the interference effect of inconsistency appeared later during the sentence reading.

Regarding the encoding of word meanings generated from context, Bordag et al. (2015) used a self-paced reading task with moving windows to record reading times during the processing of sentences with target unknown words. In their study, reading times were recorded when L2 learners read either semantically plausible (e.g., *I certainly won't use the broken RAKE anymore*) or implausible (e.g., *I certainly won't use the empty RAKE anymore*) sentences with the target word meaning (e.g., *rake*). The test sentences were preceded by learning sentences that included the same unknown words, which was expected to promote the identification of a possible meaning of each word. Accordingly, if L2 learners encoded the appropriate meanings of the target words from the learning sentences, the reading times would be longer for the implausible sentences than for the plausible sentences because the learners exhibited difficulties in integrating the meanings encoded in their mental representations with the semantically incongruent sentences. Bordag et al. obtained results similar to those of van Berkum et al. (2005), suggesting that the interference effect elicited by plausibility appeared for the later words (e.g., *anymore*) but not the target words (e.g., *rake*). This suggests that the meaning inferred and encoded from contextual information was relatively weak and unavailable in L2 reading.

Although the nature of reading times validates the research results on the generation and encoding of lexical inference, Nahatame (2015, p. 45) indicated that “it is difficult to use reading times to reveal the specific inference detected.” In other words, the results of a self-paced reading method provide little information about how semantic memory is constructed and encoded in mental representations during on-line comprehension. For example, the plausibility effect (i.e., the inflation of reading times for *rake*) suggested in Bordag et al. (2015) might be elicited by other meanings such as *pencil*. This deficit in the self-paced reading method requires complementary tests in which the generation and encoding of an intended meaning is obligatory.

Semantic relatedness judgment test. To examine the semantic representation of words in the mental lexicon, Elgort et al. (2015) and Jiang (2002, 2004) used the semantic relatedness judgment test. Generally, the test is presented after text comprehension so that the RTs obtained from the posttest reflect the mental representations readers constructed while trying to understand the text. Particularly, the semantic relatedness judgment test assesses what information was encoded in memory during text processing.

In this test, prime and probe words are sequentially presented to a participant, who is asked to determine if the two words are semantically related. Then, the test assumes that the RTs will be faster for a pair of words with the same meaning than for semantically different word pairs. Importantly, this priming-based measure reflects the implicit knowledge acquired from learning because the test taps participants for the implicit knowledge of a word instead of its conscious use (Elgort, 2011; Elgort & Warren, 2014).

For a first example, Elgort et al. (2015) showed that the word meaning generated from context was encoded in memory. They used the semantic relatedness judgment test after L2 participants encountered 90 target unknown words in three different learning sentences. These sentences were designed to constrain the possible meanings of the target words. Let us consider the following stimuli used in their experiment (p. 526):

Learning sentence:	The World Trade Center memorial foundations capture the idea of permanence versus *evanescence*.
Semantic probe:	INSTABILITY (semantically related)

One day after the learning phase, the participants made yes-no judgments on the semantic relation between the prime (i.e., *evanescence*) and the probe (i.e., *instability*). The rationale for interpreting the RT data was that if the participants encoded the meaning inferred from the learning sentences in their memory, the RTs should be faster for semantically related pairs than for semantically unrelated pairs. The results showed that a semantic relatedness effect was found for skilled L2 learners but not for less-skilled learners; whether the inferred meaning is encoded in memory or not depends on the participants' L2 proficiency. However, although the finding suggests the fragility of the word's semantic memory constructed by less-skilled L2 learners, Elgort et al. did not examine the possible differences in lexical inference between the two proficiency groups. Thus, it is still vague whether lexical inference contributes to L2 vocabulary development.

Bordag et al. (2015) employed a similar priming paradigm to measure the word knowledge implicitly learned from learning sentences. Like Elgort et al. (2015), they used a posttest after the participants read 10 texts including target words. Instead of judging whether two paired words were semantically related or not, they asked to the participants to determine if the probe words were real or not (i.e., lexical decision test). Using the same rationale for interpreting the RT data, they produced results consistent with the less-skilled learners' performances (Elgort et al., 2015), where RTs were slower in response to semantically unrelated pairs compared semantically related pairs. Although Bordag et al. did not consider the proficiency effect, they concluded that the interference effect in the test could be attributable to the frail representations of the word meanings generated by inferences. This finding is remarkable because, as described above, the participants were able to infer the meanings of unknown words in the learning phase.

As stated in Bordag et al. (2015), RT research has only recently been applied to assess the initial stage of incidental L2 word learning. Also, the latest studies are potentially limited because priming tests have been used only as a delayed posttest to measure the learning outcomes, separated from text processing. Despite the combination of self-paced reading RT methods (Bordag et al., 2015), it is still unclear whether the semantic memory of words decays by the time of the test or cannot initially be encoded. An additional possibility is the problem of form-meaning connections. Specifically, the posttest requires L2 learners to establish robust form-meaning connections during the learning phase; otherwise, they fail in accessing the semantic memory of words when looking at the newly learned word forms in the test. If it is reasonable, further research should examine which is more critical in incidental L2 vocabulary learning, encoding the information of words derived from context or establishing form-meaning connections in L2 reading comprehension.

Probe recognition test. In the present research, a probe recognition test was used to investigate the encoding of word-form information, or the establishment of a lexical representation in building a form-meaning connection during L2 reading. In this test, participants read a passage and immediately indicated whether a probe word had appeared in the previous passage. Thus, the RT data become indices of difficulty for the probe response, indicating the relative accessibility of a particular memory represented by the probe (e.g., Dopkins & Nordlie, 2011; Gernsbacher, 1990; Gerrig et al., 2009; Love et al., 2010; Nordlie et al., 2001). Note that the probe recognition test is conducted to determine if a particular semantic concept is available in working memory. For example, as introduced in Section 2.3.1, Gerrig and his colleagues used probe words as the semantic concepts that would be constructed from texts (e.g., a character name).

However, Fletcher and Chrysler (1990) showed that the probe recognition test is able to distinguish between semantic and surface memory in discourse when a probe word that appears in the text is used. Table 2.1 shows an example of a text and recognition test item used in their experiments. After reading the text, the participants were asked to determine which of two words appeared in the target sentence. The reasoning behind this procedure is that if they were able to choose the original word (i.e., *carpet*) even though the distractor (i.e., *rug*) was a close synonym used interchangeably in the text, it would suggest that they represented the surface memory of the word rather than its meaning (Fletcher, 1992). Given that Muramoto (2000) and Nahatame (2015) avoided presenting probe words in the L2 when they are interested in semantic recognition memory in L2, it is highly possible that the participants' performance on the probe recognition test reflected the surface memory of words available in the mental representation.

Table 2.1

A Text and Recognition Test Item Used in Fletcher and Chrysler (1990, p. 178)

Text George likes to flaunt his wealth by purchasing rare art treasures. He has a Persian rug worth as much as my car and it's the cheapest thing he owns. Last week he bought a French oil painting for $12,000 and an Indian necklace for $13,500. *George says that his wife was angry when she found out that the necklace cost more than the carpet*. His most expensive "treasures" are a Ming vase and a Greek statue. The statue is the only thing he ever spent more than $50,000 for. It's hard to believe that the statue cost George more than five times what he paid for the beautiful Persian carpet.
Surface Test Item: George says that his wife was angry when she found out that the necklace cost more than the (carpet/rug).

Note. The critical sentence is italicized.

2.5.3 Event-Related Brain Potentials

As described in the previous section, past RT research design did not provide information about the relationships between inference generation and encoding because the RT measures only took the place of traditional posttests such as recall and recognition. As there is still debate about whether the inference generated from context is subsequently encoded in long-term memory even in L1 reading research (Klin, 1995; Klin et al., 1999), it will be inappropriate to use RT-based tests in order to measure the fragile knowledge incidentally acquired from reading. According to Borovsky et al. (2012), these types of measurements may not be enough to capture earlier stages of word learning when learners' knowledge has not yet become tangible enough for behavioral tests (e.g., Borovsky et al., 2012). Alternatively, many researchers suggest that ERPs can be used to examine the on-line processes involved in word processing and learning in both L1 (e.g., Balass, Nelson, & Perfetti, 2010; Frishkoff, Perfetti, & Collins-Thompson, 2010; Perfetti et al., 2005) and L2 (e.g., Elgort et al., 2015; Ojima, Nakata, & Kakigi, 2005).

Recording electroencephalograms (EEGs) has been performed in the broad domain of mind, brain, and behavior sciences. The EEGs elicited by a particular event (e.g., visual linguistic stimuli) are defined as ERPs and used to examine the on-line comprehension processes in psychophysiological research on language comprehension (e.g., Kutas & Federmeier, 2011; Swaab, Ledoux, Camblin, & Boudewyn, 2012). According to *The Oxford handbook of event-related potential components* edited by Kappenman and Luck (2012), "the ERP waveform appears on the scalp as a series of positive and negative peaks that vary in polarity, amplitude, and duration as the waveform unfolds over time" (p. 4). Therefore, in ERP recordings, EEGs are obtained from multiple scalp electrodes from the onset of a particular event (Nittono, 2005); then, specific *ERP components* can be visible after multiple EEGs are overlaid together to form an average ERP waveform (Kappenman & Luck, 2012). Figure 2.10 illustrates the procedure of ERP recording and averaging.

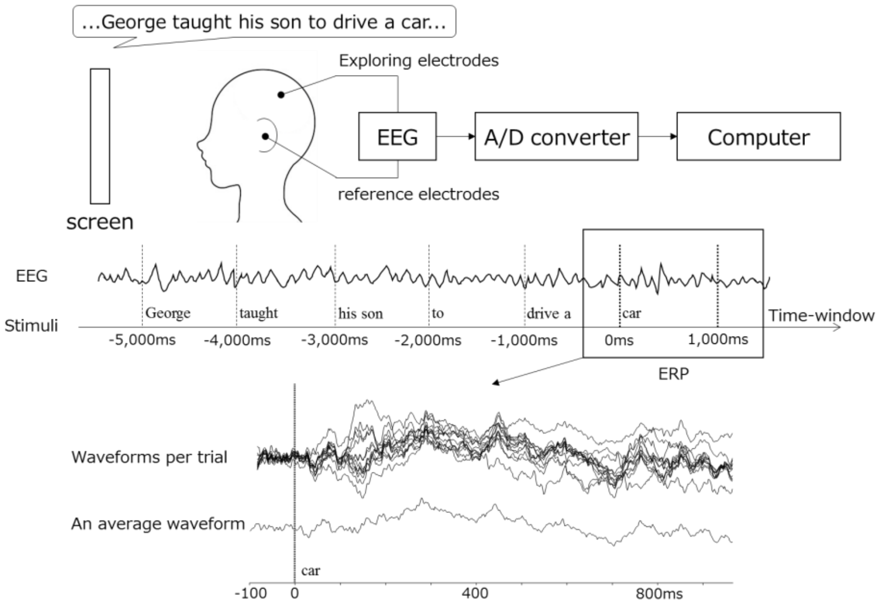


Figure 2.10. The schematic procedure of ERP recording and averaging (adapted from Nittono, 2005, p. 7). The example of ERP data is obtained from Experiment 1 reported in this dissertation.

Typical ERP research interprets the ERP components with relevance to the traits of the stimulus presented in order to infer the language comprehension processes. For example, a great number of researchers suggested that an N400 amplitude is a language-related ERP component associated with semantic processing (e.g., Borovsky et al., 2010, 2012; Kutas & Federmeier, 2011; Kutas & Hillyard, 1980; Perfetti et al., 2005; Swaab et al., 2012; van Berkum et al., 2005). The N400 is a negative shift in the EEG waveform and usually largest between 380 and 440 msec after the stimulus onset. Importantly, the N400 is sensitive to semantic information of stimuli in lexical, sentence, and discourse processing. It usually becomes larger in the processing of orthographically legal and pronounceable pseudowords than in the processing of real words, reflecting the relative difficulty of retrieving semantic memory of words (e.g., Kutas & van Petten, 1994). On the other hand, many ERP studies found an N400 modulation elicited by critical words whose possible meaning could be anticipated from the sentential or discourse context

(Borovsky et al., 2010; Camblin et al., 2007; DeLong et al., 2005; Martin et al., 2013; Otten & van Berkum, 2008, 2009; Szewczyk & Schriefers, 2013; van Berkum et al., 2005). Although the N400 amplitudes elicited by unknown words should be large because of processing difficulty, they can be modulated if readers represent the words' meanings.

Some bilingual studies described the similarities and differences in the N400 between L1 and L2 processing (e.g., Ardal, Donald, Meuter, Muldrew, & Luce, 1990; Martin et al., 2013). The findings are summarized in the following three points:

- The N400 is larger when a critical word is (a) unknown, (b) infrequent, (c) pseudo, or (d) incongruent with its prior context.
- The N400 is typically observed at midline central and parietal electrodes in semantic processing. The N400 observed in frontal electrodes is especially for word learning.
- The timing of the N400 evoked is later for languages in which the comprehender has low proficiency.

The nature of the N400 is useful to examine meaning generation, encoding, and learning of unknown words during reading comprehension. Although it is beyond the scope of this study to apply the ERP indices to incidental L2 vocabulary learning (i.e., the use of ERPs in a delayed posttest), Elgort et al. (2015) demonstrated that after experiencing a target unknown word in multiple, different contexts, skilled L2 learners could use the semantic knowledge encoded in memory to understand a subsequent test sentence. Specifically, they read three different supportive sentences including a target unknown word (e.g., "The World Trade Center memorial fountains capture the idea of permanence versus *evanescence*"). One day later, they read two types of test sentences; one was a congruous context with the meaning of *evanescence* (e.g., "The feelings of joy and sorrow eventually go away; we've all experienced their *evanescence*"), and the other was an incongruous context with the same word (e.g., “She pulled up in front of the house taking up all the space in the driveway without *evanescence*”). If the participants did not acquire the meaning of *evanescence* from the learning sentences, this would suggest the large N400 amplitudes did not differ between the two test sentences. However, if they did incidentally acquire its meaning, the N400 amplitudes would be larger in the incongruous test sentence than in the congruous one. Generally, the knowledge assessed by the index of ERPs is implicit and procedural, but not explicit and declarative (Morgan-Short, 2014). Therefore, it is appropriate to measure the initial knowledge of words, or the word memory encoded and represented in the mental representation.

2.6 Solutions to Problems of Past Research

The research reviewed so far on incidental L2 vocabulary learning has mostly looked for ways to improve its outcomes, which provides evidence to determine the factorial and pedagogical design related to discourse processing in L2. As a primary factor in the learning of new words via reading, the quality of contextual information has been found capable of predicting the success in meaning generation by inference. Regarding the use of contextual information, it is important to consider individual differences in L2 reading proficiency and lexical processing strategies as moderator variables. In addition, attention allocation control by task implementation allows L2 learners to mediate input into intake, promoting the subsequent incidental learning of words.

Although the overall research supports the general assumption that reading is an effective way to increase L2 vocabulary knowledge, some empirical studies showed only small effects of reading. One problem is that there is a lack of research investigating the theoretical frameworks of learning from texts in incidental L2 word acquisition. The present research takes the usage-based model as a theory of language learning in which learners gain lexical knowledge from the contextual usages of words. To address this assumption, LSA can be a useful tool because it has been developed based on the principles positing that “meaning is acquired by solving an enormous set of simultaneous equations that capture the contextual usage of words” (Landauer, McNamara, Dennis, & Kintsch, 2007, p. xi). Furthermore, it is important to combine the learning from texts with the theories of memory representation construction in discourse processing. Based on the Schmidt’s (1990, 1994, 1995, 2001) noticing hypothesis in language learning, the present research considered learners’ attention allocation during reading for the construction of form-meaning connections in memory.

The other problem is relevant to the methodological issue of how to capture the word knowledge acquired as a consequence of lexical inference in discourse. As pointed out by Pellicer-Sánchez (2015), research on vocabulary learning from reading has mostly revealed the quantitative and qualitative knowledge of words that are incidentally learned only after reading. Although recent L2 studies have used on-line measures such as eye-tracking to test the processing of unknown words during reading, the posttests used have a simple recall and multiple-choice format requiring overt knowledge to associate a word form with its meaning. On the other hand, the present research combined on-line and off-line measures to reveal the relationship between the word memory established while reading and the outcome retained as lexical knowledge. ERP and RT research is becoming accepted as a multi-componential approach to investigate implicit knowledge of language. Although these multiple approaches should provide empirical findings for the reading and learning theories described above, they appear not to have been applied to vocabulary research in a systematic way.

The main purpose of this study, therefore, is to provide a fuller explanation of why new words are incidentally learned from L2 reading, using a combination of on-line and off-line assessments related to how L2 readers construct mental representations of words and integrate word memory into their existing mental lexicon. Specifically, three general issues, summarized below, were addressed here:

- Issue 1: How are the meanings of unknown words generated by lexical inference and encoded in memory during L2 reading?
- Issue 2: How are form-meaning connections established during L2 reading?
- Issue 3: How is knowledge of word meaning and usage incidentally acquired?

In this dissertation, these research agendas are set as subsequent goals related to the research questions addressed in each experiment. Furthermore, the present research replicates the theoretical and empirical results obtained from laboratory-based experiments through two classroom-based investigations to confirm their ecological validity. Based on the findings, this dissertation ultimately aims to suggest vocabulary-teaching procedures as well as to establish a theoretical model of memory construction of words during L2 reading comprehension.

Chapter 3

Study 1: Generation and Encoding of Lexical Inference in L2 Reading

3.1 Experiment 1: ERPs in Lexical Inference and Encoding

3.1.1 Design and Research Questions

Experiment 1 used ERPs to test participants' generation and encoding of lexical inference while they read sentences in L2. Specifically, the first goal was to determine if Japanese EFL learners can use their comprehension of sentential information to rapidly infer and generate the meaning of unknown words embedded in a context. The second goal was to test whether the meaning generated by inference was encoded in the learners' mental lexicon. To achieve these purposes, this ERP experiment focused on the N400 amplitudes that can be an index of the semantic processing of words. The three research questions (RQs) addressed in Experiment 1 are as follows:

RQ1: Do Japanese EFL learners generate the meaning of unknown words based on the word-context semantic similarity computed by LSA?

RQ2: Do Japanese EFL learners rapidly encode the inferred meaning of unknown words in their mental lexicon?

RQ3: Is learners' L2 reading proficiency a predictor of the generation and encoding of lexical inference?

Regarding the processing of unknown words, two levels of word-context semantic similarity (HSS and LSS) and two types of target words (Known and Unknown) were crossed as a factorial design. According to the prediction of the usage-based model and the LSA theory, the target words embedded in HSS contexts are highly compatible with a contextual message compared to those in LSS contexts. Therefore, Japanese EFL learners might glean lexical properties from the HSS contexts better than from the LSS contexts. The paradigm to test this context effect is similar to the one used by Borovsky et al. (2010), as reviewed in Section 2.3.2. While unknown words are known to elicit an N400 effect (e.g., Kutas & Federmeier, 2011), the N400 amplitudes can be modulated by context effects (e.g., Otten & van Berkum, 2008). Particularly, if the semantic processing of unknown words is facilitated by inferences, the N400 amplitudes should be lower than if it was still difficult for readers (e.g., Borovsky et al., 2010).

The factorial design for examining encoding inferences was similar to the one used by Borovsky et al. (2010): 2 (Context: HSS and LSS) × 2 (Word: Known and Unknown) × 2 (Plausibility: Plausible and Implausible). Two types of test sentences for word encoding were created in terms of the plausibility of word usage between a transitive verb and a target word (see details in Section 3.1.2.2). For example, the meaning of *computer* is compatible with the transitive verb *move*, but it is highly atypical with *greet*. Because the N400 is sensitive to context-dependent semantic anomalies (e.g., Kutas & Hillyard, 1980), the N400 amplitudes should be lower when *computer* (presented as a target unknown word) co-occurs with *move* compared to *greet* if the inferred meaning of the word is rapidly encoded in memory.

3.1.2 Method

3.1.2.1 Participants

Forty-four right-handed, Japanese EFL learners participated in this ERP session (17 female and 27 male; average age = 20.1, range = 18–23). None had neurological or psychiatric disease. All had normal or corrected-to-normal vision. The participants were undergraduate or graduate students at the University of Tsukuba, majoring in the humanities, social sciences, education, engineering, biology, or medicine. At the time of the experiment, all had studied English for at least six years in Japan.

Their English reading proficiency was estimated using the reading subsections of the pre-first (6 items) and second grade (20 items) in the Eiken tests (see Appendices 1 and 2; Obunsha, 2010a, 2010b). The test scores (M = 15.27, 95% CI [13.71, 16.84], SD = 5.15, Cronbach's α = .84) were not statistically different among the subsequent experiments, $F < 1$. Their proficiency level was assumed to be from A2 to B2 based on Common European Framework of Reference (CEFR) for Languages' can-do descriptors (Council of Europe, 2001). Participants gave informed consent before the experiment that was approved by the research ethics committee of the University of Tsukuba and gained ¥2,000 for their participation.

3.1.2.2 Materials

Learning sentence creation. For each target word, a learning sentence pair was created, resulting in 170 learning sentences with 85 target words (examples in Table 3.1). These sets were collected from Borovsky et al. (2010), Griffin and Bock (1998), and van Assche et al. (2011), and were slightly modified especially for Japanese EFL learners by replacing low-frequency basewords with high-frequency synonyms ranging from 1,000 to 3,000 word level based on the *JACET list of 8,000 basic words* (JACET, 2003). The strength of semantic similarities between each target word and contextual meanings was computed by LSA (http://lsa.colorado.edu/), based on the semantic space of "General reading up to first year college" (i.e., the corpus referenced by LSA; see Dennis, 2007).

Between each learning sentence pair, one with relatively high LSA similarity was categorized as HSS context, and the other was regarded as LSS context. There was a large difference in the strength of semantic similarity between the HSS (M = .39, 95% CI [.36, .42], SD = .12) and LSS contexts (M = .27, 95% CI [.24, .29], SD = .10), $t(63)$ = 7.00, $p < .001$, d = 1.09, M_{diff} = .12 (95% CI [.09, .15]).[6] The length of learning sentences

[6] As described in the next section, finally, 64 learning sentence pairs were used in this experiment.

between the conditions did not differ, $t(63) = 1.40$, $p = .168$, $d = 0.15$, $M_{\text{diff}} = 0.34$ (95% CI [–0.15, 0.84]). All the learning sentence pairs are presented in Appendix 3.

Table 3.1

Sample Sets of Learning and Test Sentences per Condition

Learning sentences:			
HSS context	I installed new word processor software on my		computer/crench
LSS context	His seat in the small classroom was next to the		computer/crench
Test sentences:			
Plausible	P1: He moved the	P2: She used the	computer/crench
Implausible	I1: She surprised the	I2: He greeted the	computer/crench

Note. The target words were either a real English word (e.g., *computer*) or an English-like pseudoword (e.g., *crench*). Test sentences included either a plausible (P1 and P2) or an implausible (I1 and I2) verb for the usage of the previously presented target word.

Test sentence creation. Each learning sentence pair was accompanied by four test sentences for the plausibility judgment test, created in consultation with a native speaker of English. As shown in Table 3.1, two of the test sentences in each set included a transitive verb semantically congruent with the usage of the target word, and the other two sentences had a transitive verb that was semantically incongruent (see also Appendix 4). As in Borovsky et al. (2010), the test sentences consisted of a pronoun + transitive verb + determiner + target word. To present all the combinations of test sentences in a trial, two of the four possible test sentences (i.e., PP, PI, IP, and II) were presented after each learning sentence by a random selection.

The relative plausibility of the test sentences was verified with 32 Japanese undergraduate and graduate students different those who participated in the main experiment. They were asked to determine if the target known words embedded in each test sentence were used appropriately by a yes-no judgment. Based on Borovsky et al. (2010), the experiment adopted the test sentences that obtained 75% agreement, resulting in the elimination of 21 learning sentence pairs for the stimuli. Finally, 64 learning sentence pairs and corresponding test sentences were used in the ERP experiment.

3.1.2.3 Procedure

During the experiment, the participants were seated in a soundproof, electrically shielded recording booth. Stimuli were presented on a 27-inch computer screen (SX2761 model; Eizo, Japan) approximately 100 cm in front of the participant, in black, Times New Roman 72-point font on a white background. At this distance, 20.0 characters were equal to 1° of visual angle, so that eye movements were almost not necessary to see a word presented on the screen. The participants were asked to read sentences for comprehension even when unfamiliar words appeared. After they read each learning sentence, they then read two test sentences and provided a yes-no judgment of each plausibility.

Stimuli were presented one word at a time by a rapid-serial-visual-presentation technique (RSVP) in order to avoid artifacts caused by eye movements. The start of each trial was announced by the word "Ready?" appearing at a center of the screen (1,500 msec duration). All but the sentence final words appeared for 500 msec with a stimulus-onset-asynchrony (SOA) of 200 msec. The final critical word appeared for 1,400 msec, and it was immediately followed by the two test sentences, to prevent participants from using any encoding strategies for unknown words. After a row of fixation crosses with numbers (i.e., **1** and **2**) for 1,500 msec, words appeared with timing identical to the learning sentences. The participants were instructed to respond to whether a meaning of a test sentence was plausible or not by clicking a corresponding mouse button. Swift responses were not required to avoid excessive muscle activity. Every time 1/4 of the trials ended, the participants were able to have as much rest as they wanted. This small block took approximately five minutes to complete, so that the ERP session led to approximately 25 minutes in total. Figure 3.1 displays the sequence of each trial.

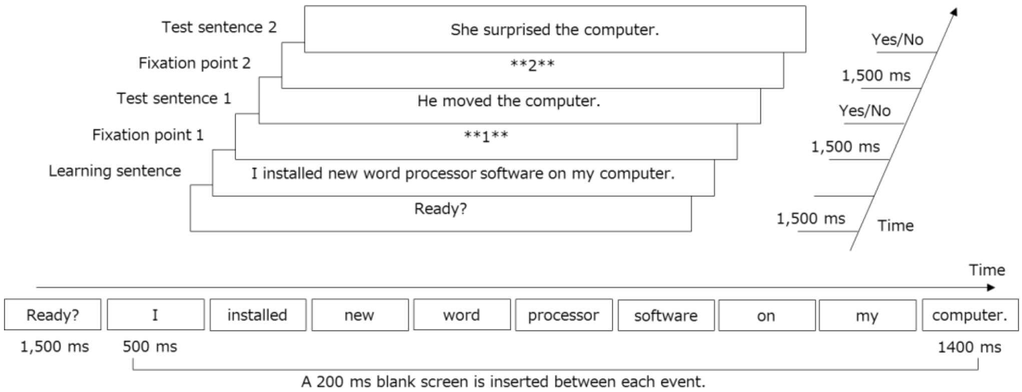


Figure 3.1. The sequence of events in each trial during the EEG recording (upper) and the RSVP and SOA in the learning and two test sentences (lower).

The 64 learning sentences were presented in a random order, and two of four possible test sentences (i.e., PP, PI, IP, and II) were randomly selected with all possible combinations throughout the experiment, counterbalanced across versions using the Latin square method. SuperLab 4.5 for Windows (Cedrus, the U.S.) was used to present stimuli and record responses. To familiarize participants with the experimental procedure, they completed eight practice sets before the session.

3.1.2.4 ERP Recording and Data Analysis

The EEG was recorded from 19 tin electrodes embedded in an Electro-Cap (Electro-Cap International, Eaton, OH), placed according to the standard 10-20 system locations (see Figure 3.2), with a Biotop 6R12 amplifier system (NEC Medical Systems Corp., Japan). Recordings were referenced to the on-line linked earlobes. Impedances were kept below 5kΩ. The EEG activity was digitally sampled at a rate of 500 Hz and on-line filtered between 0.5 Hz and 200 Hz. Epochs ranged from –100 to 1,100 msec after the onset of the target words embedded in the learning and test sentences. Trials with artifacts by eye movements, blinks, and excessive muscle activity per participant were manually screened, resulting in 2% removal of all observations (range = 1%–4% per condition). After that, baseline correction was performed using the average EEG activity in the 100 msec epoch preceding the onset of the target words. Because the presentations of the three target words (i.e., one was in a learning sentence, and two were in test sentences in each trial) were adequately separated from each other, any problematic waveform overlap was not observed.

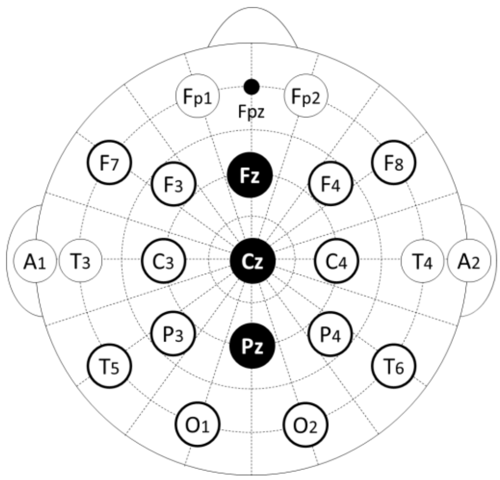


Figure 3.2. International 10-20 system of electrode placement. Fifteen electrodes of thick-lined circles were entered into statistical analyses. Three central line electrodes (i.e., Fz, Cz, and Pz; black circles) were mainly focused in the analysis of the N400.

3.1.3 Results

3.1.3.1 Inference Generation

ERP results. The time-window for the analysis at the target words was determined on the basis of visual inspection of the grand average waveforms; the context effect was seen in the time-window of 400 to 700 msec (see Figure 3.3). The different negative waveforms suggest that the N400 effect appeared differently according to the degree of word-context semantic similarity. However, it is necessary to verify whether the ERPs were indeed defined as the N400 based on the combination of polarity, shape, and scalp distribution (Kappenman & Luck, 2012), because the negativity found in this experiment was relatively late compared to the typical N400 waveforms (e.g., Kutas & Hillyard, 1980). This problem will be further discussed in Section 3.1.4.

Table 3.2 presents the descriptive statistics of the N400 amplitudes after grand averaging. A two-way MANOVA was performed on the mean amplitudes at the Fz, Cz, and Pz electrodes, and a significant interaction between Context and Word was found, $F(3, 41) = 3.51$, $p = .023$, $\eta^2 = .20$ (see also Table 3.3). Figure 3.4 illustrates the different effects of Context on the mean N400 amplitudes between known and unknown words.

In the Fz analysis, whereas the Context effect on the N400 was not reliable in the processing of known words ($p = .093$, $d = 0.26$, $M_{\text{diff}} = 0.71$ [–0.12, 1.54]), the effect became medium in the processing of unknown words ($p < .001$, $d = 0.50$, $M_{\text{diff}} = 2.22$ [1.12, 3.32]). Similar results were obtained in the analysis of the Cz electrode; the Context effect was small in the processing of known words ($p = .059$, $d = 0.38$, $M_{\text{diff}} = 0.71$ [–0.02, 1.08]). Although it was small in the processing of unknown words, the N400 amplitude statistically increased when the unknown words were presented in the LSS context ($p = .008$, $d = 0.43$, $M_{\text{diff}} = 1.36$ [0.38, 2.34]). At the Pz electrode, the Context effect was also null in the processing of known words ($p = .353$, $d = 0.19$, $M_{\text{diff}} = 0.25$ [–0.29, 0.79]), and it was medium in the processing of unknown words ($p = .002$, $d = 0.56$, $M_{\text{diff}} = 1.83$ [0.71, 2.95]). The N400 amplitudes between known and unknown words presented in the HSS context did not differ at Fz ($p = .570$, $d = 0.26$, $M_{\text{diff}} = –0.33$ [–1.51, 0.84]), Cz ($p = .551$, $d = 0.26$, $M_{\text{diff}} = 0.25$ [–0.60, 1.11]), and Pz ($p = .383$, $d = 0.26$, $M_{\text{diff}} = 0.36$ [–0.46, 1.19]).

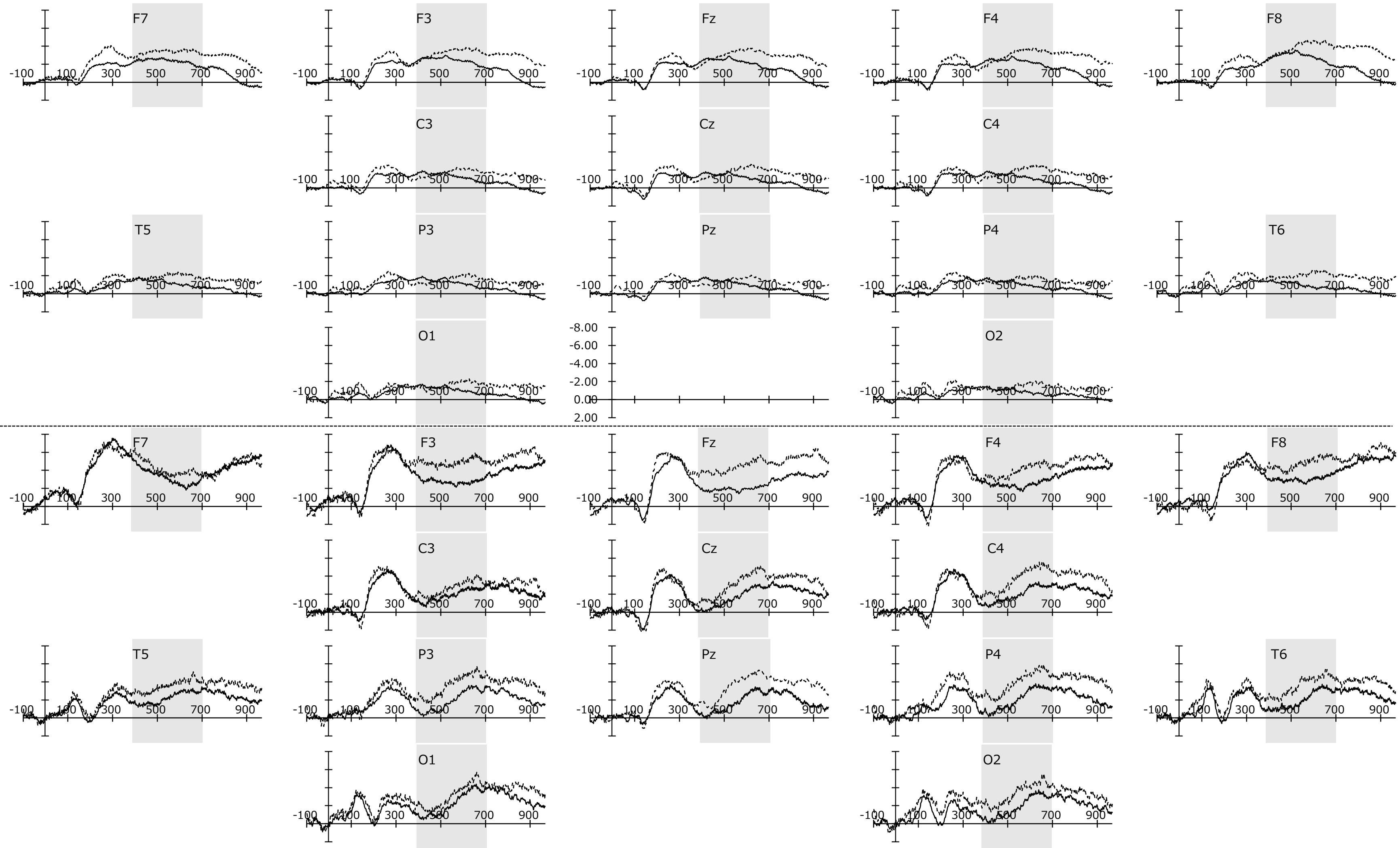


Figure 3.3. Grand average ERP amplitudes (μV) elicited by known (upper) and unknown (lower) words in the learning sentences between the HSS (solid line) and LSS (dotted line) conditions. The analyzed time-windows of 400–700 msec are highlighted gray.

Table 3.2

*Grand Average of the N400 Amplitudes (μV) Elicited by Context and Word Effects (*N *= 44)*

Target words	HSS learning sentence			LSS learning sentence		
	M	95% CI	*SD*	*M*	95% CI	*SD*
			Midline-frontal electrode (Fz)			
Known	–2.29	[–2.93, –1.65]	2.11	–3.00	[–3.97, –2.03]	3.19
Unknown	–1.96	[–3.19, –0.72]	4.06	–4.18	[–5.63, –2.73]	4.77
			Midline-central electrode (Cz)			
Known	–1.29	[–1.61, –0.96]	1.08	–1.82	[–2.32, –1.31]	1.67
Unknown	–1.54	[–2.36, –0.72]	2.71	–2.90	[–3.98, –1.81]	3.57
			Midline-parietal electrode (Pz)			
Known	–1.18	[–1.51, –0.86]	1.07	–1.44	[–1.91, –0.96]	1.58
Unknown	–1.54	[–2.32, –0.77]	2.54	–3.37	[–4.54, –2.20]	3.84

Table 3.3

Multivariate and Univariate ANOVAs for the N400 Observed From Three Electrodes

Source	Multivariate $F(3, 41)$			Univariate $F(1, 43)$ Fz			Cz			Pz		
	F	*P*	η^2	*F*	*p*	η^2	*F*	*p*	η^2	*F*	*p*	η^2
Context (C)	6.65	.001	.33	17.91	< .001	.40	12.96	< .001	.30	10.02	.003	.23
Word (W)	4.75	.006	.26	0.90	.348	.02	3.43	.071	.08	10.09	.003	.23
C × W	3.51	.023	.20	5.02	.030	.12	1.96	.169	.05	7.53	.009	.18

Note. Multivariate *F* rations were generated from Pillai's statistic. Multivariate η^2 = 1 – Wilks' Λ.

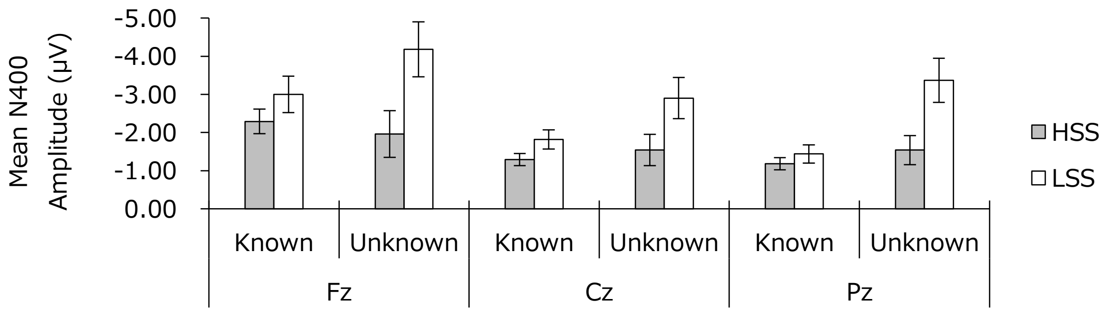


Figure 3.4. Means with ±*SEM* bars of the N400 amplitudes (μV) elicited by the target known and unknown words in the HSS and LSS learning sentences at Fz, Cz, and Pz.

Effects of proficiency. As shown in Figure 3.5, the N400 amplitudes elicited by the unknown words at Fz, Cz, and Pz did not correlate with participants' English reading proficiency except for the condition that the target unknown words were presented in the HSS learning sentences at Pz (N = 44, r = –.33, p = .028). This was an only significant predictor in the simultaneous multiple regression model, F(1, 42) = 5.18, p = .028, adjusted R^2 = .09, and the overall result showed that the difficulty in meaning generation from the HSS sentences, as measured by the N400 amplitudes, was not relevant to participants' reading proficiency.

Note that the adjusted R^2 obtained here must not be reliable because of the small sample size.[7] Cook's distance was referred to find outliers of the data set; the maximum value was .15, less than 1.00. Normality, homoscedasticity, and linearity of residuals were regarded as good by a visual inspection of corresponding plots.

[7] Prerequisites for performing a multiple regression analysis were confirmed as accurately as possible. *Multicollinearity*: None of the variables correlated strongly with each other (r < .70), and the tolerance values of each factor were not less than 1.00. These suggested that there were no multicollinearities among them. Independence of residuals: The result of the Durbin-Watson statistic was 1.91 (not less than 1.00 or more than 3.00). This showed that there were no correlations among any combinations of variables' residuals. *Outliers*: A leverage method was used to find any outliers of the data set; the maximum value of a leverage was .05, which was less than a criterion of .07 (= 2 × {1 [the number of predictors] + 1} / 44 [a sample size]). *Normality*, *homoscedasticity*, and *linearity of residuals*: Although some residuals of the variables submitted into the regression model were not homogeneous, the normality and linearity of the residuals were regarded as good.

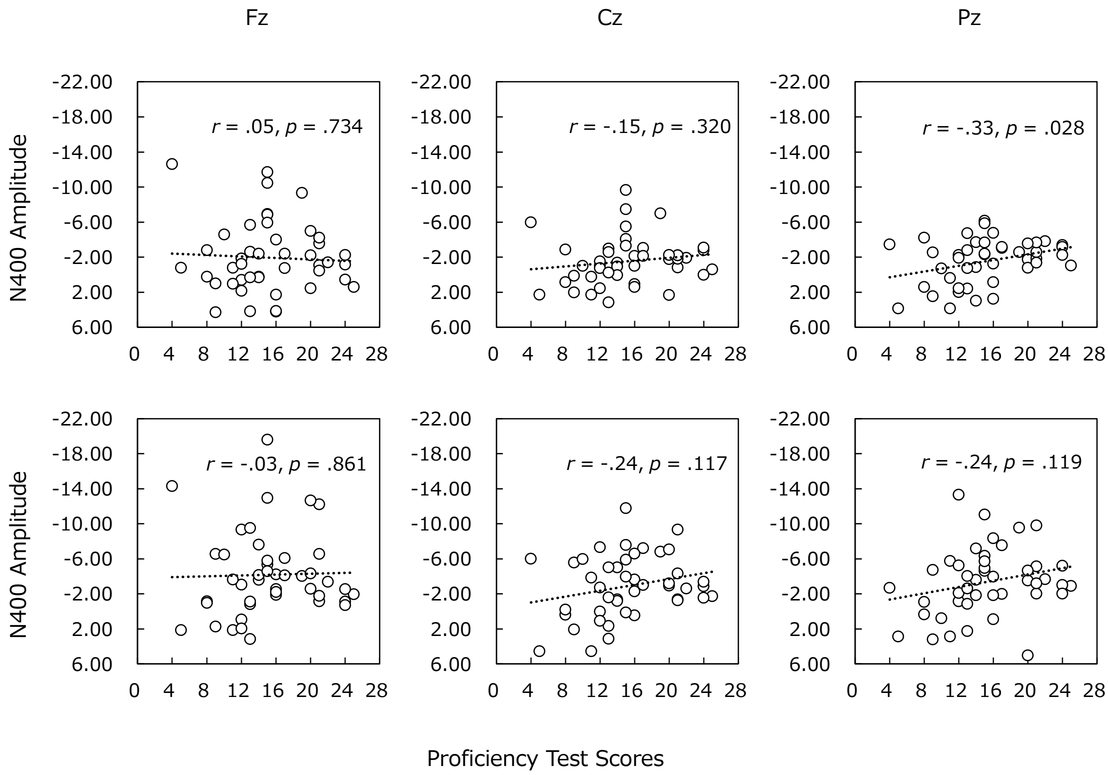


Figure 3.5. Scatterplots between the proficiency test scores and N400 amplitudes in the HSS (upper) and LSS (lower) conditions with an approximate line (N = 44).

3.1.3.2 Word Encoding

ERP results. In the same manner as in the previous data analysis, a time-window was set in the analysis of the plausibility effects on the processing of target words. As shown in Figures 3.6 and 3.7, the second large negativity occurred in the 300–600 msec latency window. A three-way MANOVA was performed on the mean amplitudes at the Fz, Cz, and Pz electrodes, and a significant three-way interaction of Context × Word × Plausibility was found, $F(3, 41) = 5.32$, $p = .003$, $\eta^2 = .28$. The descriptive statistics and the overall results of the multivariate and univariate tests are shown in Tables 3.4 and 3.5, respectively.

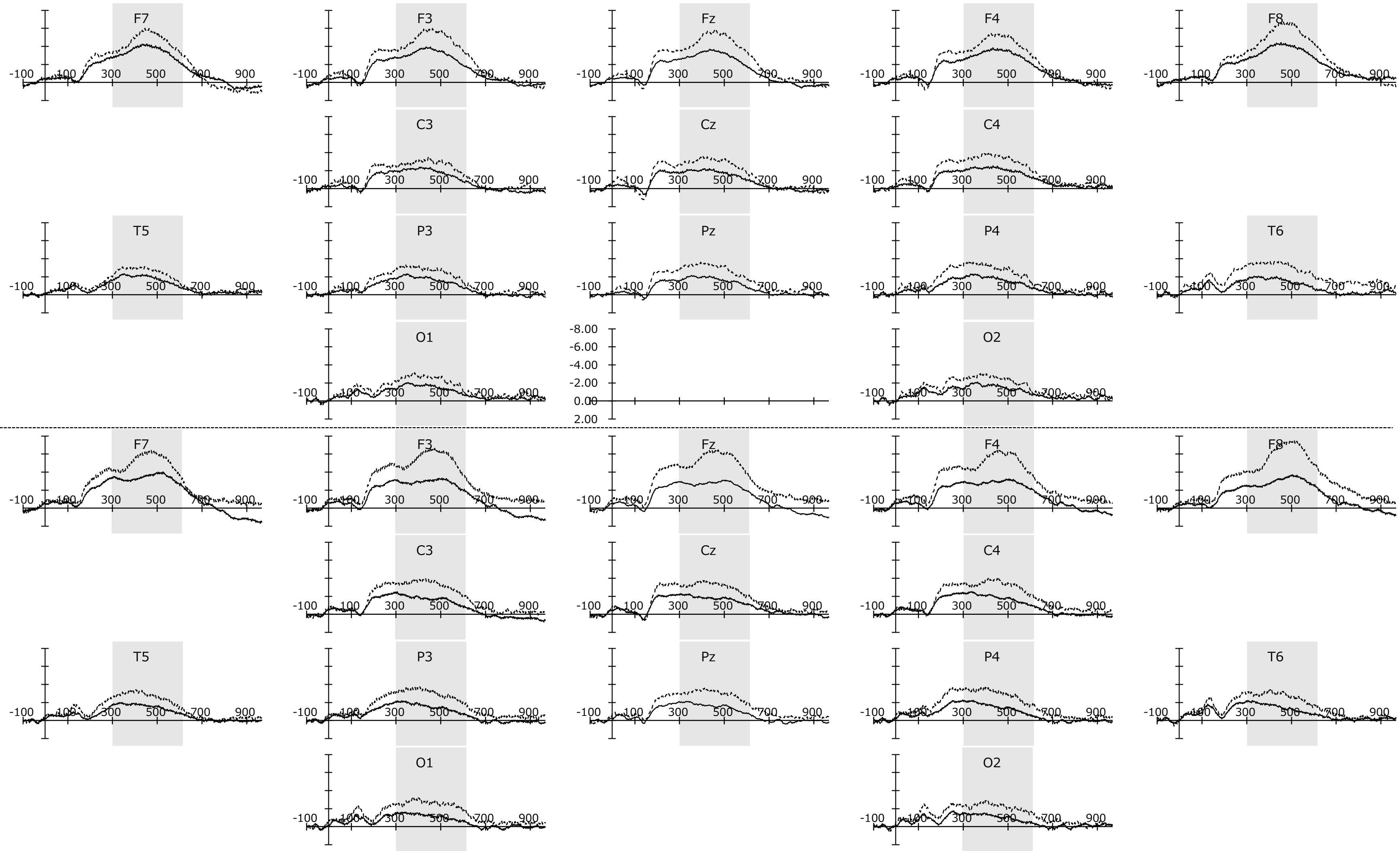


Figure 3.6. Grand average ERP amplitudes (μV) elicited by known words in the plausible (solid line) and implausible (dotted line) test sentences between the HSS (upper) and LSS (lower) conditions. The analyzed time-windows of 300–600 msec are highlighted gray.

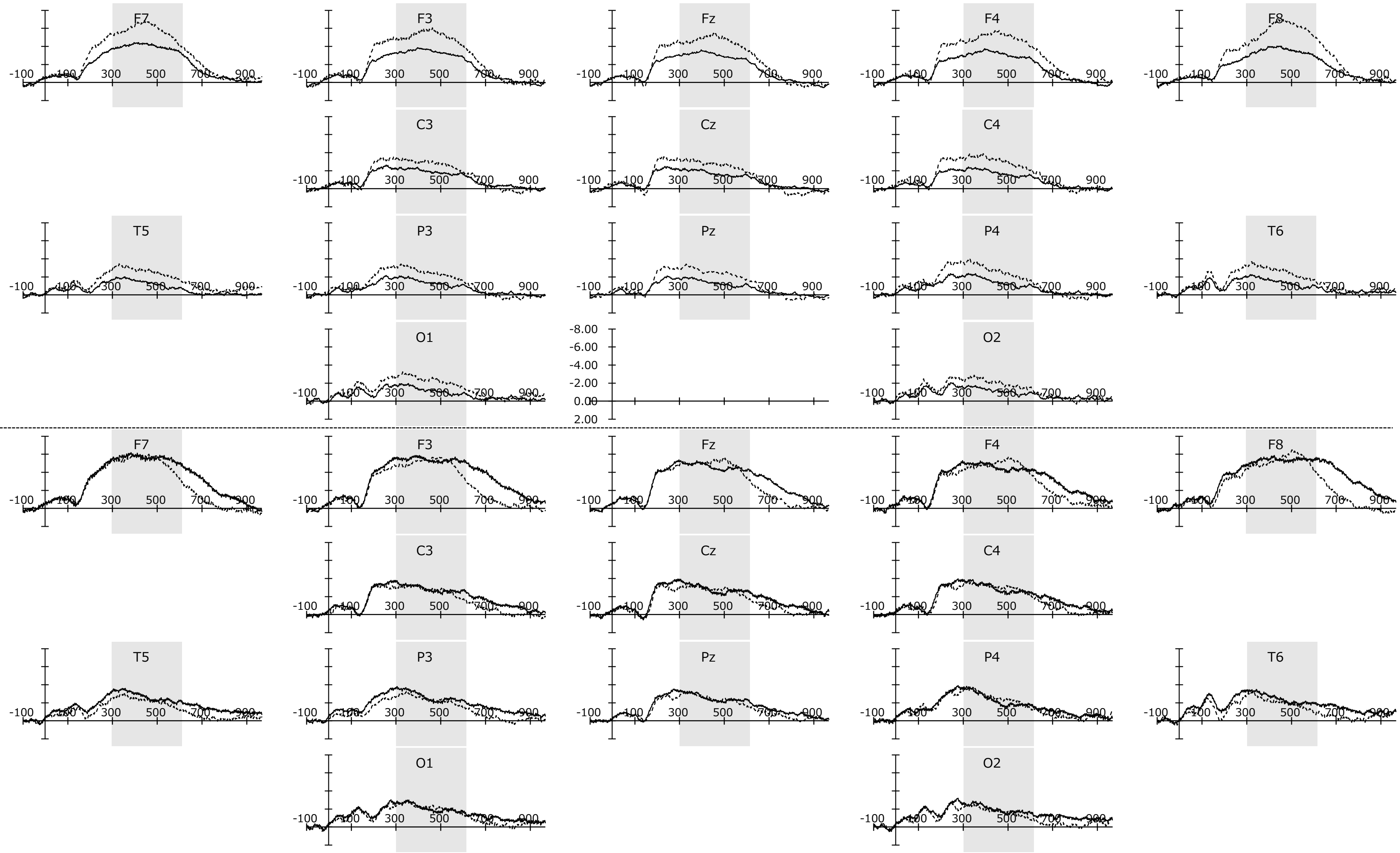


Figure 3.7. Grand average ERP amplitudes (μV) elicited by unknown words in the plausible (solid line) and implausible (dotted line) test sentences between the HSS (upper) and LSS (lower) conditions. The analyzed time-windows of 300–600 msec are highlighted gray.

Table 3.4

*Grand Average of the N400 Amplitudes (µV) Elicited by Plausibility Effects (*N *= 44)*

	HSS learning sentence			LSS learning sentence		
Test sentences	*M*	95% CI	*SD*	*M*	95% CI	*SD*
			Midline-frontal electrode (Fz)			
Known						
Plausible	–2.97	[–3.87, –2.07]	2.96	–2.70	[–3.43, –1.98]	2.40
Implausible	–4.62	[–5.70, –3.55]	3.54	–5.43	[–6.59, –4.27]	3.82
Unknown						
Plausible	–3.07	[–3.81, –2.33]	2.44	–4.64	[–5.81, –3.47]	3.85
Implausible	–4.65	[–5.85, –3.45]	3.95	–4.92	[–6.25, –3.59]	4.38
			Midline-central electrode (Cz)			
Known						
Plausible	–1.78	[–2.27, –1.29]	1.61	–1.77	[–2.20, –1.34]	1.41
Implausible	–2.93	[–3.58, –2.27]	2.15	–3.15	[–3.80, –2.50]	2.12
Unknown						
Plausible	–1.69	[–2.10, –1.29]	1.33	–2.85	[–3.58, –2.12]	2.39
Implausible	–2.70	[–3.41, –2.00]	2.33	–2.80	[–3.43, –2.17]	2.06
			Midline-parietal electrode (Pz)			
Known						
Plausible	–1.69	[–2.10, –1.29]	1.33	–1.59	[–1.98, –1.19]	1.29
Implausible	–2.96	[–3.52, –2.40]	1.84	–3.00	[–3.52, –2.48]	1.71
Unknown						
Plausible	–1.36	[–1.71, –1.00]	1.18	–2.56	[–3.21, –1.90]	2.16
Implausible	–2.53	[–3.14, –1.92]	2.00	–2.50	[–3.01, –2.00]	1.66

Table 3.5

Multivariate and Univariate ANOVAs for the N400 Observed From Three Electrodes

	Multivariate $F(3, 41)$			Univariate $F(1, 43)$								
				Fz			Cz			Pz		
Source	F	p	η^2	F	p	η^2	F	p	η^2	F	p	η^2
Context (C)	2.50	.073	.15	7.58	.009	.02	6.26	.016	.01	4.75	.035	.01
Word (W)	5.53	.003	.29	3.82	.057	.01	0.45	.504	.01	0.34	.564	.01
Plausibility (P)	22.26	< .001	.62	49.54	< .001	.15	36.42	< .001	.05	64.53	< .001	.06
C × W	1.85	.154	.12	4.21	.046	.01	4.64	.037	.01	5.50	.024	.01
C × P	1.89	.147	.12	0.08	.774	.01	1.94	.170	.01	0.06	.055	.01
W × P	2.39	.083	.15	7.10	.011	.03	4.33	.043	.01	5.84	.020	.01
C × W × P	5.32	.003	.28	13.33	< .001	.02	5.18	.028	.01	7.47	.009	.01

Note. Multivariate F rations were generated from Pillai's statistic. Multivariate $\eta^2 = 1 - \text{Wilks' } \Lambda$.

The N400 effects were found in both the HSS and LSS learning sentences when the target known words were presented in the implausible test sentences (see Figure 3.8). Post hoc analyses showed the following results for the plausibility effects:

- In the Fz analysis, the N400 magnitudes were larger in the implausible test sentences than in the plausible test sentences, in both the HSS condition, $F(1, 43) = 15.91$, $p < .001$, $\eta^2 = .05$ ($M_{\text{diff}} = –1.65$, 95% CI [–2.48, –0.82]), and the LSS condition, $F(1, 43) = 47.01$, $p < .001$, $\eta^2 = .02$ ($M_{\text{diff}} = –2.73$, 95% CI [–3.53, –1.93]).
- In the Cz analysis, they were larger in the implausible test sentences than in the plausible test sentences, in both the HSS condition, $F(1, 43) = 16.94$, $p < .001$, $\eta^2 = .04$ ($M_{\text{diff}} = –1.15$, 95% CI [–1.71, –0.59]), and the LSS condition, $F(1, 43) = 16.95$, $p < .001$, $\eta^2 = .05$ ($M_{\text{diff}} = –1.38$, 95% CI [–2.06, –0.70]).
- In the Pz analysis, they were larger in the implausible test sentences than in the plausible test sentences, in both the HSS condition, $F(1, 43) = 26.38$, $p < .001$, $\eta^2 = .06$ ($M_{\text{diff}} = –1.27$, 95% CI [–1.77, –0.77]), and the LSS condition, $F(1, 43) = 25.99$, $p < .001$, $\eta^2 = .07$ ($M_{\text{diff}} = –1.42$, 95% CI [–1.98, –0.86]).

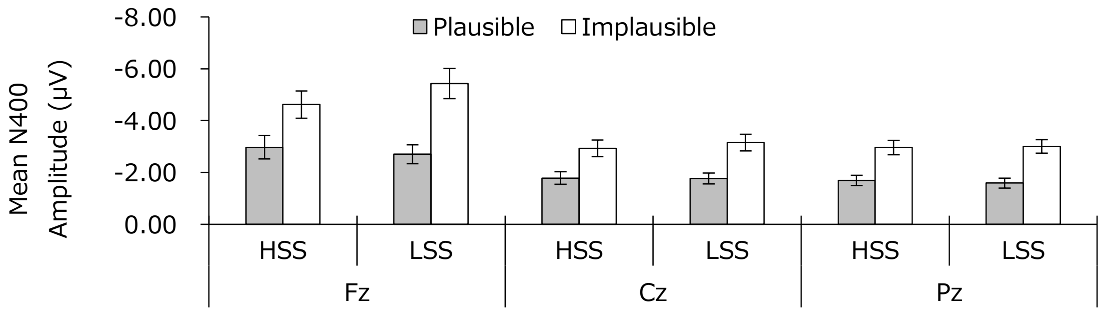


Figure 3.8. Means with ±*SEM* bars of the N400 amplitudes (μV) elicited by the target known words in the plausible and implausible test sentences at Fz, Cz, and Pz.

Figure 3.9 shows the different plausibility effects of Plausibility on the mean N400 amplitudes elicited by the target unknown words between the HSS and LSS conditions. The N400 magnitudes did not reduce in the plausible test sentences after the participants read the words in the LSS contexts, compared to when the target known words were presented. They were still, however, different in the HSS condition between the plausible and implausible test sentences, as shown by post hoc analyses:

- In the Fz analysis, the N400 magnitudes were larger in the implausible test sentences than in the plausible test sentences in the HSS condition, $F(1, 43) = 22.26$, $p < .001$, $\eta^2 = .02$ ($M_{\text{diff}} = –1.58$, 95% CI [–2.25, –0.90]), but not in the LSS condition, $F(1, 43) = 0.35$, $p = .558$, $\eta^2 < .01$ ($M_{\text{diff}} = –0.28$, 95% CI [–1.24, 0.68]).
- In the Cz analysis, they were larger in the implausible test sentences than in the plausible test sentences in the HSS condition, $F(1, 43) = 12.72$, $p = .001$, $\eta^2 = .03$ ($M_{\text{diff}} = –1.01$, 95% CI [–1.59, –0.44]), but not in the LSS condition, $F(1, 43) = 0.06$, $p = .886$, $\eta^2 < .01$ ($M_{\text{diff}} = 0.05$, 95% CI [–0.66, 0.76]).
- In the Pz analysis, they were larger in the implausible test sentences than in the plausible test sentences in the HSS condition, $F(1, 43) = 18.50$, $p < .001$, $\eta^2 = .05$ ($M_{\text{diff}} = –1.18$, 95% CI [–1.73, –0.62]), but not in the LSS condition, $F(1, 43) = 0.07$, $p = .851$, $\eta^2 < .01$ ($M_{\text{diff}} = 0.06$, 95% CI [–0.54, 0.65]).

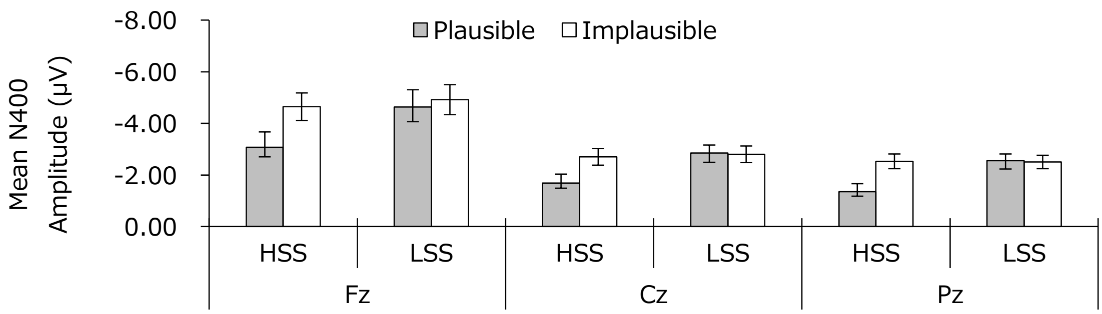


Figure 3.9. Means with ±*SEM* bars of the N400 amplitudes (μV) elicited by the target unknown words in the plausible and implausible test sentences at Fz, Cz, and Pz.

Effects of proficiency. The N400 amplitudes elicited by the unknown words at Fz, Cz, and Pz did not correlate with participants' English reading proficiency across conditions ($N = 44$, $r_{range} = .11$ to .30, ps > .05). A simultaneous multiple regression model was not significant, $F(12, 31) = 0.99$, $p = .484$, adjusted $R^2 = .28$, suggesting that the plausibility effect did not appear according to the learners' reading proficiency.

3.1.3.3 Plausibility Judgment Test

Accuracy. Table 3.6 shows the mean accuracy rates of correct responses per condition. Regarding the Context effects on the judgment accuracy between known and unknown words, the results of three-way ANOVA showed that there were significant interactions of Context × Word, $F(1, 43) = 24.19$, $p < .001$, $\eta^2 = .02$, and of Word × Plausibility, $F(1, 43) = 44.03$, $p < .001$, $\eta^2 = .08$ (see also Table 3.7).

Table 3.6

Mean Accuracy Rates With 95% CIs and SD*s of the Plausibility Judgment Test (*N *= 44)*

	HSS learning sentence			LSS learning sentence		
Test sentences	*M*	95% CI	*SD*	*M*	95% CI	*SD*
Known word						
Plausible	.88	[.85, .92]	.11	.90	[.87, .93]	.10
Implausible	.87	[.84, .90]	.10	.87	[.84, .89]	.10
Total	.88	[.85, .90]	.08	.88	[.86, .91]	.08
Unknown word						
Plausible	.74	[.69, .79]	.16	.64	[.59, .69]	.16
Implausible	.86	[.83, .90]	.12	.80	[.76, .84]	.12
Total	.80	[.77, .83]	.10	.72	[.70, .75]	.08

Table 3.7

ANOVA Results for Main Effects and Interaction Effects of Context, Word Type, and Plausibility on the Judgment Accuracy in the Test

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
Context (C)	1	0.11	0.11	27.93	< .001	.01
Error (C)	43	0.17	0.00			
Word (W)	1	1.25	1.25	100.07	< .001	.16
Error (W)	43	0.54	0.01			
Plausibility (P)	1	0.29	0.29	8.21	.006	.04
Error (P)	43	1.51	0.04			
C × W	1	0.17	0.17	24.19	< .001	.02
Error (C × W)	43	0.31	0.01			
C × P	1	0.00	0.00	0.11	.747	.01
Error (C × P)	43	0.24	0.01			
W × P	1	0.62	0.62	44.03	< .001	.08
Error (W × P)	43	0.61	0.01			
C × W × P	1	0.01	0.01	1.95	.170	.01
Error (C × W × P)	43	0.32	0.01			

Note. Between-cells variance = 1.43.

The accuracy of the plausibility judgment on the known word usage did not differ per condition, $F(1, 43) = 0.79$, $p = .379$, $\eta^2 < .01$ (see Figure 3.10). In contrast, the judgment accuracy for unknown word usage was higher in the HSS context (.80) than in the LSS context (.72), $p < .001$, $d = 0.88$, $M_{\text{diff}} = .08$ (95% CI [.06, .11]). The other result was that the Plausibility effect was found in the processing of unknown words, $F(1, 43) = 22.96$, $p < .001$, $\eta^2 = .22$, but not in that of known words, $F(1, 43) = 2.96$, $p = .093$, η^2

= .02. The relative difficulty in the plausibility judgment was slightly higher when the usage of unknown words was implausible than when it was plausible.

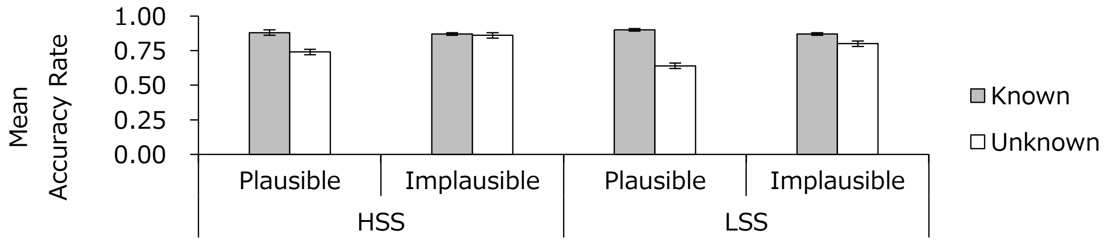


Figure 3.10. Mean judgment accuracy for known and unknown words with ±*SEM* bars modulated by the Context and Plausibility effects (N = 44).

Effects of proficiency. Figure 3.11 visualizes the results of Pearson's correlations. The plausibility judgment performance for unknown words correlated with participants' English reading proficiency in the HSS condition (N = 44, r = .38, p = .012) but not in the LSS condition (N = 44, r = .18, p = .245). A stepwise multiple regression model indicated that only the judgment accuracy for the unknown words presented in the HSS sentences was significant in explaining their reading proficiency, $F(1, 42) = 6.95$, p = .012, adjusted R^2 = .12.

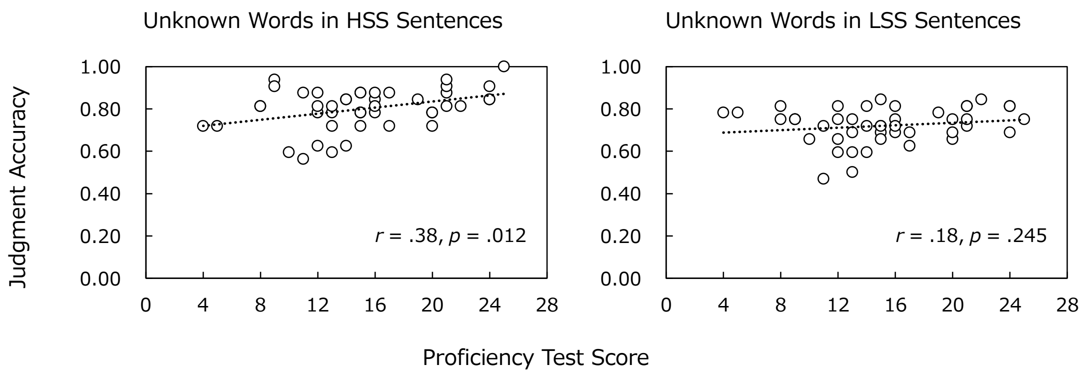


Figure 3.11. Scatterplots of the proficiency test scores and judgment accuracy rates with an approximate line (N = 44).

3.1.4 Discussion

In the current ERP experiment, I examined whether Japanese EFL learners can use the information about the word-context semantic similarity computed by LSA to generate the meaning of unknown words while reading a sentence in L2. The second interest was that the contextual usage meaning of the words derived from context can be encoded in memory. To probe whether Japanese university students indeed comprehended the meaning of unknown words by generating lexical inference, the strength of word-context semantic similarity was manipulated so that the HSS learning sentences had enough information for new word meanings to be derived and the LSS ones did not. In addition, to test whether they could integrate inferential information about the meaning of the words into their memory, the participants made plausibility judgments of how to use word meanings in the new sentences. The N400 effects were further analyzed in terms of the correlation with the participants' L2 reading proficiency.

The rationale underlying RQ1 was that whereas the N400 is larger when processing an unknown word, its magnitude should be modulated if its meaning is identified by inference (e.g., Borovsky et al., 2010). The ERP results showed that the target unknown words embedded in the LSS learning sentences elicited larger N400 amplitudes than those in the HSS learning sentences. Therefore, the decrease in the N400 effects in the HSS learning sentences shows that the participants derived the word meanings from context, resulting in the relative ease of the semantic processing of the target words as reflected by the N400 modulation. I will discuss this central finding below from the viewpoints of context-based lexical prediction, individual differences, and methodological issues.

Before doing so, it is important to settle a general concern that the ERP component elicited by the target words in this experiment can be interpreted as the N400 because the time-window of 400–700 msec adopted in the ERP analysis was later than the ordinary one reported in previous research. In L1 research, the time-window of 300–500 msec was often used in the N400 analysis (e.g., Borovsky et al., 2010; Camblin et al., 2007; DeLong et al., 2005; Kutas & Hillyard, 1980; Otten & van Berkum, 2008, 2009; van Berkum et al., 2005; Wicha et al., 2004). Prior studies with proficient bilinguals conventionally used the same time-window when comparing brain waveforms between L1 and L2 readers, based on research using a similar experimental paradigm. For example, Elgort et al. (2015) and Martin et al. (2013) employed the 300–500 msec time-window as the N400 elicited by critical nouns. Similarly, Ardal et al. (1990) temporarily defined the N400 as the average peak deviation in the 300–600 msec time-window. However, it should be noted that in context effects on lexical predictive inference, Martin et al. (2013) also reported that N400 peak latencies were delayed in L2 (320–520 msec) compared to L1 (220–420 msec). Additionally, Ardal et al. (1990) showed that N400 peak latencies were significantly delayed in the processing of semantic incongruity during L2 reading.

According to Kappenman and Luck (2012), although it is not usually possible to identify and define a specific ERP component from the observed ERP waveforms, "the best way to identify a specific component is to take a *converging evidence* approach" (p. 17) in terms of a combination of polarity, latency, scalp distribution, and sensitivity to experimental manipulations. As mentioned in Section 2.5.3, the N400 is defined as a negative-going brain waveform between 200 and 600 msec, largest over the centro-parietal sites, with a right-hemisphere bias (Kutas & Federmeier, 2011). In these respects, the ERPs observed in the current experiment were negative going and sensitive to the differences between the known and unknown words and between the HSS and LSS learning sentences. Furthermore, the N400 effects appeared slightly larger in the right hemisphere, such as at F8, C4, P4, T6, and O2, than in the left hemisphere, such as at F7, C3, P3, T5, and O1 (see the lower column of Figure 3.3). The latency and the centro-parietal bias were not found in the ERPs data; however, they often varied across experiments. For example, the N400 effects saliently appeared in Cz and Pz in Borovsky et al. (2010) and Kutas and Hillyard (1980); Fz, Cz, and Pz in Otten and van Berkum (2008, 2009) and van Berkum et al. (2005); and Fz and Cz in Martin et al. (2013). Therefore, in practice, given that the present experiment was not designed to determine the difference in the N400 effects between native and bilingual readers, it should be appropriate to adopt the other time-window (i.e., 400–700 msec) to define the obtained EEG as the N400 component.

The N400 modulation elicited by target nouns in highly supportive sentences is a result consistent with L1 readers (e.g., Camblin et al., 2007; DeLong et al., 2005; Otten & van Berkum, 2008, 2009; van Berkum et al., 2005; Wicha et al., 2004), and it also replicates the findings of L2 readers' comprehension (Martin et al., 2013). These suggest that the participants in Experiment 1 used context-based lexical anticipation because the experimental design was the same as in the previous research, in which the stimulus sentences had the ability to activate a specific concept of upcoming words (e.g., van Berkum et al., 2005). In other words, the meaning inferred by lexical anticipation facilitated the semantic processing of upcoming words during sentence comprehension. Borovsky et al. (2010) showed that the N400 effects elicited by word anticipation also appeared when the target nouns were unfamiliar to readers.

However, the anticipation effect on the N400 modulation was not so large in terms of effect sizes, similar to the case in Martin et al. (2013), which demonstrated the deficit in bilinguals' semantic processing compared to L1 readers. This suggests that L2 learners indeed derived enough contextual information to generate the meaning of unknown words, but this cognitive process was limited. In relation to RQ3, a possible interpretation is that individual differences in L2 reading proficiency affected word anticipation. Nevertheless, although a large N400 effect can be an index of the relative ease of semantic processing (e.g., DeLong et al., 2005), the N400 magnitudes elicited by the target unknown words did not reliably correlate with L2 reading proficiency scores. This indicates that the comprehensibility of the target unknown words in sentence processing did not vary according to the L2 reading proficiency. These claims have both consistent and inconsistent views with previous L2 research: (a) the generation of lexical inference requires higher L2 proficiency (e.g., Bensoussan & Laufer, 1984; Nassaji, 2006; Ushiro et al., 2013; Wesche & Paribakht, 2010), and (b) the context effects surpasses the deficit in L2 proficiency (Elgort et al., 2015; Hamada, 2012; Ma et al., 2015; Webb, 2008).

These discrepancies regarding the effects of L2 proficiency on the generation of lexical inference can be attributed to the different methodologies used in each research. Most L2 studies employed think-aloud protocols as indices of learners' generation of lexical inference (Deschambault, 2012). Whereas verbal protocols reflect only conscious processing about thoughts that can be verbalized (e.g., Magliano et al., 1999), the ERPs measure unconscious, neurological processing that is not reportable (e.g., Borovsky et al., 2012; Kutas & Federmeier, 2011; Swaab et al., 2012). Thus, it is possible that the proficiency effects did not appear in the present ERP experiment compared to the previous think-aloud studies. However, this experiment did not directly compare the difference in the N400 effects between L1 and L2 readers like Martin et al. (2013), and their ERP study found that the N400 effects appeared later in bilinguals' sentence comprehension. Thus, it is more reasonable to consider that the proficiency effects on lexical inference should interact with various factors such as the stimuli, individual differences, and research methodologies. In Experiment 1, at least, both skilled and less-skilled university students generated the meaning of unknown words as reflected by the N400 modulation found between the HSS and LSS learning sentences.

Regarding RQ2, both ERP waveforms and plausible judgment accuracy showed that the meanings generated by lexical inference were encoded in memory. First, the N400 effects between the plausible and implausible test sentences clearly appeared after the participants read the target known words in both the HSS and LSS learning sentences. As those meanings were incongruous with the implausible test sentences, this result can be interpreted as the indices of the semantic anomaly effects as measured by the N400 (e.g., Kutas & Federmeier, 2011). Based on this prerequisite to interpret the N400 effects in the target unknown words, the further results showed that the N400 were modulated in the plausible test sentences only after the participants read the target unknown words in the HSS learning sentences. This means that the word meanings encoded in memory from the HSS learning sentences were no longer unknown when the words were processed again in the new sentences. On the other hand, the N400 effects did not appear in the plausible test sentences after the participants processed the target unknown words in the LSS learning sentences, suggesting that (a) any inferences were not generated so that the subsequent encoding processes did not occur and (b) some inferences were at least generated, but they were incongruous with the message of the plausible sentences. Together, the findings indicate that the exact meanings generated by lexical inferences were so rapidly encoded in the learners' memory that the learners could apply the knowledge representation to semantic processing in another context. This perspective is also supported by the differences in the plausibility judgment accuracy provided in the plausible test sentences between the HSS and LSS conditions.

For a further discussion of RQ3, interestingly, the effects of L2 reading proficiency on the plausibility judgment differed between the N400 effects and the off-line judgment accuracy. Whereas the N400 effects did not become a predictor of the participants' L2 reading proficiency, the judgment accuracy did. This suggests that the skilled participants were better able to derive enough meaning of the target unknown words from the HSS learning sentences to explicitly decide if it was plausible or not in the test sentences than the less-skilled participants were; however, the implicit use of the word knowledge represented in memory did not differ according to their L2 reading proficiency. Although it is beyond the scope of this experiment to determine if the linguistic knowledge acquired implicitly is procedural or declarative, the word information encoded by the skilled L2 learners was so robust that they could consciously made accurate judgments on its contextual usage. In contrast, the information derived from the same words by the less-skilled learners might be unconscious, resulting in the difference between the unconscious plausibility judgment as reflected by the N400 effects and the conscious plausibility judgment as reflected by its accuracy.

Although previous ERP research on L1 vocabulary acquisition used a deliberate, word-list type of learning to test the encoding processes, the present experiment added the findings that the word representation constructed by inference can be encoded in memory. The overall findings are consistent with Borovsky et al. (2010), which examined L1 readers' word encoding processes during sentence comprehension. The test timing was significantly different between the present experiment (i.e., immediate) and that of Elgort et al. (2015; one day after); nevertheless, both results provided similar evidence that the word knowledge representation encoded from reading was available to another situation in text processing.

More importantly, the most salient difference between Experiment 1 and previous studies is the method to extract and represent the contextual usage meaning of unknown words. Based on the usage-based model, the present experiment employed the LSA theory because of the assumption that speakers induce the rules of word usage from the frequent information of how a particular word co-occurs with other words (Crossley et al. 2008, 2014; Ellis, 2002; Ellis & Ferreira-Junior, 2009; Kidd et al., 2010; Tomasello, 2003). The outcomes discussed above supported this because the initial stage of incidental L2 word learning was promoted as predicted by the LSA values. Similarly, other studies showed that the guessability defined as the contextual constraint and contextual cues improved success in L2 lexical inference attempts (e.g., Elgort et al., 2015; Hamada, 2012; Huckin & Bloch, 1993; Li, 1988; Ma et al., 2015; Ushiro et al., 2013; Webb, 2008). At this time, although it is impossible to determine which is a stronger predictor of incidental L2 vocabulary learning, the latter is disputable in terms of whether it relates to vocabulary growth (e.g., Li, 1988; Ma et al., 2015) or not (e.g., Mondria & Wit-de Boer, 1991; Webb, 2008). In contrast, if the LSA theory is closely linked to incidental L2 vocabulary gains, the word information derived and encoded in learners' memory should be retained as vocabulary knowledge, which will be addressed in Study 3.

In summary, similar to previous think-aloud research in L2 (e.g., Paribakht & Wesche, 1999), Experiment 1 showed the first step of incidental L2 vocabulary learning, that is, the generation and encoding of lexical inference during reading comprehension (e.g., Bordag et al., 2015). However, some inevitable methodological problems of ERPs prevent the current findings from providing a fuller picture of these two processes. Because the procedure of the stimuli presentation was the RSVP in this experiment to inhibit the artifacts caused by eye movements, the participants were forced to engage in an unnatural reading activity. In relation to this, unknown words are not always elaborated by a prior context in a natural discourse, as mentioned in Section 2.3.2. Therefore, the next experiments aimed at solving these two issues.

3.2 Experiment 2A: Forward and Backward Lexical Inference and Encoding

3.2.1 Design, Hypothesis, and Research Question

Experiment 2A used the semantic relatedness judgment test described in Section 2.5.2, instead of ERPs, to examine the semantic similarity effects between a word and its prior sentence and its subsequent sentence on the generation and encoding of lexical inferences. The purpose of Experiment 2A was to reveal whether or not the meanings of unknown words are generated and encoded while participants were trying to process those words that were semantically elaborated by subsequent contexts.

Two levels of word-context semantic similarity (HSS and LSS), as in the example below, were compared in the same manner as in Experiment 1:

(4a) *He tried to put the pieces of the broken plate back together with* **marf**. (HSS)

(4b) *She walked across the large room to Mike's messy desk and returned his* **marf**. (LSS)

Sentences (4a) and (4b) both contained a nonword (in bold; *glue*). This target word followed either HSS or LSS contexts (italics) that determined whether inferences about what kind of thing the target word refers to were generated in participants' memory (see Experiment 1). In Experiment 2A, participants read learning sentences and then made yes-no judgments regarding the semantic relatedness between a target (e.g., *marf*) and a probe (e.g., *glue*). Given that forward contextual elaboration facilitated generating lexical inferences, even if an explicit word-meaning guessing task was not given to learners, it is hypothesized that the two-word relatedness judgments will be faster and more accurate after reading Sentence (4a) than Sentence (4b). It should be noted that the judgments for Sentence (4b) indicate lucky guesses to some extent, because the LSS contexts were designed not to constrain the possible meaning of target words, so participants could not generate and encode their specific concepts. Even if participants made a correct judgment, it would indicate that they changed their prior answer (e.g., *something on the desk*) to be more specific (i.e., *glue*), by checking the meaning of a probe with their understanding of the sentence. In this case, RTs should be longer compared with a case in which a specific meaning is initially represented in memory.

Experiment 2A was also conducted to examine backward contextual elaboration effects. Similar to the forward condition, Sentences (5a) and (5b) both shared the same nonword (in bold; *instrument*), but this target word was (not) semantically relevant with subsequent contextual information (italics):

(5a) Joe picked up the **asdor** and *began to play a melody*. (HSS)

(5b) Joe picked up the **asdor** and *walked home*. (LSS)

When participants encountered the target word, its meaning should not be represented because they had not met the contextual information semantically relevant to it yet. If, as a sentence unfolded, they integrated the contextual information with an unknown word's meaning, the results of the semantic relatedness judgment test can be explained in the same manner as forward contextual elaboration does. However, insignificant differences in accuracy and RTs between the HSS and LSS conditions would suggest that the L2 learners were not sensitive to backward contextual elaboration and did not generate lexical inferences. Thus, the hypothesis (H) and RQ addressed in Experiment 2A are summarized as follows:

H1: Forward contextual elaboration facilitates processing of unknown words by lexical inferences.

RQ4: Does backward contextual elaboration facilitate the processing of unknown words by lexical inferences?

3.2.2 Method

3.2.2.1 Participants

Twenty Japanese EFL learners participated in this RT experiment (10 females and 10 males; average age = 19.2, range = 18–21). None had participated in the prior study. They were undergraduate students at the University of Tsukuba, majoring in social studies, international relations, education, or engineering. At the time of this experiment, all had studied English for at least six years in Japan. Their CEFR level was assumed to be from A2 to B2 based on their self-report. They gave informed consent before the experiment and gained ¥1,000 for their participation.

3.2.2.2 Materials

Learning sentences. Two levels of word-context semantic similarity (HSS and LSS) and two types of elaborated direction (forward and backward) were crossed in a factorial design (see Table 3.8). There were two types of elaborative information for each learning sentence: (a) forward elaboration, in which a prior context was semantically related to the inferable meaning of the target words; and (b) backward elaboration, in which a subsequent context were semantically related to the same meaning of the target words. All the learning sentences are presented in Appendix 5

Table 3.8

Sample Sets of Stimuli Used in Experiment 2A

Learning sentences	Probes
Forward elaboration	
HSS: Always knock before you open my *door*/*cauge*.	ドア
LSS: The girl moved slowly toward the *door*/*cauge*.	[door]
Backward elaboration	
HSS: Joe picked up the *instrument*/*asdor* and began to play a melody.	がっき
LSS: Joe picked up the *instrument*/*asdor* and began to walk home.	[instrument]

Note. The target words were either a real English word (e.g., *door*/*instrument*) or an English-like pseudoword (e.g., *cauge*/*asdor*). The probes were presented in Japanese.

For the backward condition, 12 learning sentence pairs were adapted from Chaffin et al. (2001) because these sentences showed a large difference in the strength of semantic similarity between the HSS and LSS conditions (see Table 3.9). It should be noted that

LSA was run for evaluating the semantic similarity between the target words (e.g., *instrument*) and the following contextual information (e.g., *and began to play a melody*). Low-frequency basewords in the sentences (4,000-word level and over) were substituted by easier synonyms based on JACET (2003). For the forward condition, the same number of learning sentences was selected from Experiment 1, which exhibited a salient sign of meaning generation and encoding found by a visual inspection of ERPs. To confirm that the LSA value differed between the HSS and LSS conditions but not between the forward and backward conditions, a two-way ANOVA was performed with Context (HSS and LSS) and Direction (Forward and Backward) as within-participants variables. These requirements were satisfied with (a) a significant main effect of Context, $F(1, 11) = 94.40$, $p < .001$, $\eta^2 = .22$, (b) insignificant main effect of Direction, $F(1, 11) = 0.05$, $p = .826$, $\eta^2 < .01$, and (c) insignificant two-factor interaction, $F(1, 11) = 0.44$, $p = .520$, $\eta^2 < .01$. The mean number of words per condition was almost the same (see also Table 3.9).

Table 3.9

Characteristics of Experimental and Filler Sentences and Probes

		Learning sentences		Probes
Condition	*k*	Sentence length	LSA	Word length
HSS				
Forward	6	10.7	.37	2.85
Backward	6	11.3	.39	2.85
LSS				
Forward	6	11.2	.23	2.83
Backward	6	11.0	.23	2.86
Filler	24	11.2		2.72

Note. Sentence length means the average number of words included in each sentence.

In addition, 24 filler sentences were prepared. Twelve of these were adapted from Experiment 1 as the forward condition, and the other fillers from Chaffin et al. (2001) as the backward condition. There were no differences from the experimental sentences except that the target words were high-frequency real English words. The target words for the forward condition were always at the end of each sentence. Those for the backward condition were the same as in Chaffin et al.

Probe words. Probe words corresponding to each learning sentence pair were created for the semantic relatedness judgment test. The original meanings of the target nonwords were used basically, but some target words were replaced based on the result of a pilot test (40 Japanese EFL learners at the same university were asked to fill in the first word that came to mind when reading the experimental sentences). For example, Chaffin et al. (2001) originally used the target word *asdor* as *guitar* in the context "Joe picked up the *asdor* and began to strum a tune." However, after substituting easier words for low frequency words (e.g., *strum a tune → play a melody*), few participants answered that the meaning of *asdor* was *guitar*. In these cases, the hypernyms were used as the probe words (e.g., *guitar → instrument*). Each probe word was translated into Japanese to avoid effects of English word familiarity on RTs. Every set of 24 fillers also had similar probe words, but the target-probe pairs of the fillers were semantically unrelated to each other. Therefore, the participants had to respond "yes" to the experimental target-probe pairs, and "no" to the filler target-probe pairs because "if related items generate data for answering a research question, unrelated items are fillers" (Jiang, 2012, p. 141). The word-length of the probes were the almost same among the conditions, Fs < 1.

3.2.2.3 Procedure

Participants were tested individually in a single experimental session. After a participant completed the same reading proficiency test within a 30-minute time limit as in Experiment 1, they were instructed on how to perform the semantic relatedness judgment test. To familiarize participants with the test procedure, they completed four practice sets before the main test. SuperLab 4.5 for Windows (Cedrus, the U.S.) and response pad (RB-730 model, Cedrus, the U.S.) was used to provide all of the practices, instructions, experimental sets, and record responses and RTs. Figure 3.12 illustrates the sequence of each trial.

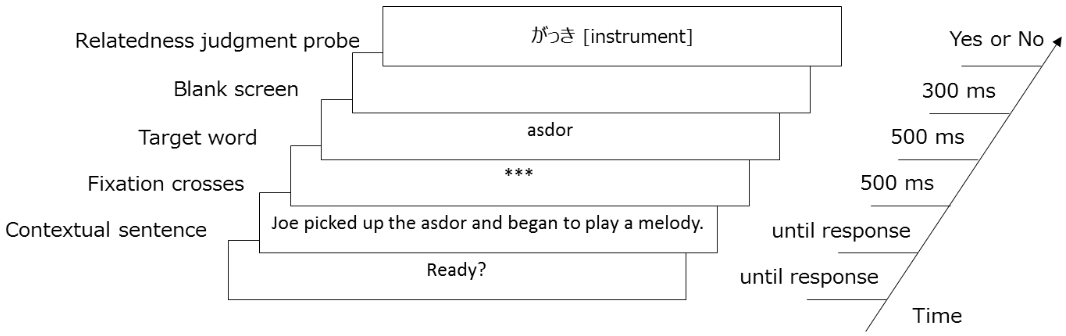


Figure 3.12. The sequence of events in each trial during the on-line semantic relatedness judgment test in Experiment 2A.

In the semantic relatedness judgment test, 12 forward and 12 backward conditions were presented in a random order on a computer screen. A set of 24 fillers was also randomly inserted. Participants were asked to read the sentences for comprehension at their own pace by pressing a button on the response pad, but were not instructed to infer the meanings of the target words. When they finished each sentence, a row of fixation crosses appeared in the center of the screen for 500 msec to ready them for the onset of the target word. Then, crosses were replaced by the prime word (i.e., the target nonword) for 500 msec. After an additional 300 msec blank screen, a corresponding probe word was flashed in Japanese. The task was to judge whether the two presented words were semantically related as swiftly and accurately as possible by pressing the yes-no buttons. Participants spent approximately 20 minutes completing all of the trials.

3.2.2.4 Data Analysis

To examine the effects of forward and backward contextual elaboration, two-way ANOVAs were conducted on judgment accuracy and RTs, with Context (HSS and LSS) and Direction (Forward and Backward) as within-participants variables. Prior to analyzing the RT data, when the participants incorrectly responded to the experimental target-probe pairs in each trial, it was excluded from the data analysis. Then, trials where RTs were ±2.5 *SD*s beyond the mean for each participant were substituted with the scores of the *M* ± 2.5 *SD*s as outliers (approximately 1% of all observations).

3.2.3 Results

Mean correct response rates on filler trials reached 97%, supporting the claim that participants performed the semantic relatedness judgment test appropriately. Table 3.10 presents the descriptive statistics for the test performance.

Table 3.10

*Mean Accuracy Rates and RTs for the Semantic Relatedness Judgment Test (*N *= 20)*

	Forward elaboration			Backward elaboration		
Condition	*M*	95% CI	*SD*	*M*	95% CI	*SD*
Accuracy rates						
HSS sentence	.81	[.73, .89]	.17	.80	[.73, .87]	.15
LSS sentence	.58	[.48, .68]	.22	.43	[.33, .53]	.21
RTs in msec						
HSS sentence	867	[733, 1,001]	286	863	[726, 1,001]	293
LSS sentence	1,141	[969, 1,313]	367	1,149	[965, 1,332]	392

Accuracy. Figure 3.13 indicates that judgment accuracy was higher in the HSS than the LSS condition, regardless of the direction of contextual elaboration. The two-way ANOVA supported this observation; although the main effect of Context was significant, $F(1, 19) = 36.38$, $p < .001$, $\eta^2 = .44$, there was no significant main effect of Direction, $F(1, 19) = 4.25$, $p = .053$, $\eta^2 = .03$, and these factors did not interact, $F(1, 19) = 3.57$, $p = .074$, $\eta^2 = .02$.

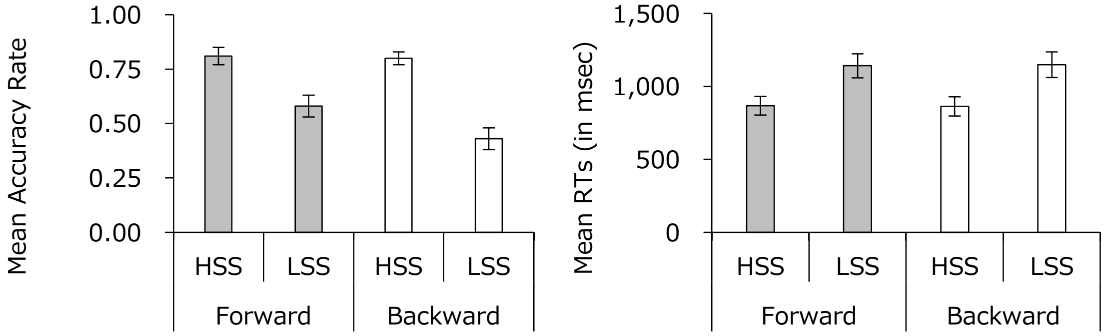


Figure 3.13. Mean accuracy rates and RTs with ±*SEM* bars in Experiment 2A.

Reaction times. To examine whether the RT data were consistent with the accuracy data, a similar two-way ANOVA was performed. As shown in Figure 3.13, RTs were shorter in the HSS than the LSS condition, $F(1, 19) = 32.90$, $p < .001$, $\eta^2 = .51$. However, the Direction effect did not appear, $F(1, 19) = 0.01$, $p = .944$, $\eta^2 < .01$, nor interact with the Context effect, $F(1, 19) = 0.05$, $p = .826$, $\eta^2 < .01$.

3.2.4 Discussion

One of the important findings of Experiment 2A is that the highly elaborative contexts led to more accurate responses and promoted faster RTs. Although the HSS and LSS learning sentences shared common target words, participants could respond to the probes more accurately and swiftly in the HSS than LSS condition. In the LSS condition, contextual information exhibited relatively lower semantic similarity to the meaning of the target words, and this manipulation ensured that specific lexical inferences were rarely made while reading. Therefore, the different effects between the HSS and LSS conditions suggest that participants could generate the meaning of unknown words based on the highly elaborative information. This result supported the study hypothesis and was consistent with the findings of Experiment 1 and L1 reading studies (e.g., Borovsky et al., 2010; DeLong et al., 2005; Otten & van Berkum, 2008; Szewczyk & Schriefers, 2013), indicating that Japanese EFL learners processed unknown words by context-based word anticipation (and integration) in the forward elaboration condition.

However, direction did not affect semantic relatedness judgment performance; accuracy and RTs in the forward and backward elaboration conditions did not reliably differ. Similar to the effects of forward contextual elaboration, the HSS learning sentences in the backward condition promoted higher accuracy and faster RTs compared to the LSS learning sentences. These results suggest that participants could infer the meaning of unknown words based on backward contextual elaboration. This is consistent with the findings that contextual elaboration facilitates backward lexical inferences (Chaffin et al., 2001), and with some think-aloud results showing that L2 learners are able to wait for additional information to identify the meaning of unknown words until they meet highly elaborative contexts (Huckin & Bloch, 1993). As many researchers have stated (e.g., Bensoussan & Laufer, 1984; Li, 1988; Mondria & Wit-de Boer, 1991; Nassaji, 2006; Nation, 2013; Webb, 2008), semantically strong relationship between unknown words and contextual information is essential for making lexical inferences. This study especially indicates that the effects of word-context semantic similarity computed by LSA on Japanese EFL learners' lexical inferences are not specific to a think-aloud task.

3.3 Experiment 2B: Follow-Up Study of Experiment 2A

3.3.1 Design, Hypothesis, and Research Question

Although the findings of Experiment 2A suggest that participants were sensitive to backward elaborative contexts, it did not reveal why they could make backward lexical inferences when the target words were semantically related to the subsequent contexts. Whereas forward contextual elaboration facilitated processing of unknown words by context-based word anticipation, the backward condition should tap readers for strategic rereading for context-word integration (Cain et al., 2004; Chaffin et al., 2001; Daneman & Green, 1986). Given that the full sentences always appeared on the computer screen in Experiment 2A, the effects of forward and backward conditions cannot be distinguished. Therefore, Experiment 2B was required to remove the effects of strategic rereading, in which participants read learning sentences that were segmented into chunk units. The same hypothesis and RQ as in Experiment 2A were examined here.

3.3.2 Method

3.3.2.1 Participants

Twenty-eight Japanese EFL learners participated in the follow-up experiment (15 females and 13 females; average age = 20.1, range = 18–24). None had participated in the prior study. They were undergraduate and graduate students at the University of Tsukuba, majoring in psychology, biology, engineering, or medicine. At the time of this experiment, all had studied English for at least six years in Japan. Their CEFR level was assumed to be from A2 to B2 based on their self-report. They gave informed consent before the experiment and gained ¥1,000 for their participation.

The participants' English reading proficiency was estimated using the same Eiken test as in Experiment 1 in order to ensure that the proficiency level was homogeneous between Experiments 2A and 2B. The test scores were not statistically different, $t(46)$ =

0.30, p = .767, d = 0.09, Cronbach's α = .83 (Experiment 2A: M = 11.80, 95% CI [9.73, 13.87], SD = 4.42; Experiment 2B: M = 11.43, 95% CI [9.82, 13.03], SD = 4.13).

3.3.2.2 Materials, Procedure, and Data Analysis

Twelve short story pairs were newly constructed (three-sentence length; see Table 3.11). The first sentences of the stories, which were chunk-parsed, shared the same target unknown words as Experiment 2A.

Table 3.11

Sample Sets of Experimental Short Passages Used in Experiment 2B

Forward condition
First sentence
HSS: She tried / to put / the pieces of the broken plate back together / with *marf*. /
LSS: She walked across the large room / to Mike's dirty desk / and returned / his *marf*. /
Second and third sentences
After that, she ran to school because her school started in twenty minutes. Today, she had an important test so she wanted to get there on time.
Backward condition
First sentence
HSS: Joe picked up / the *asdor* / and began to / play a melody. /
LSS: Joe picked up / the *asdor* / and began to / walk home. /
Second and third sentences
Then, he noticed that he had just got an email from his boss. The news about his promotion made him very happy.

Note. Slashes represent pause-chunks. The probes were the same as in Experiment 2A.

This chunk segmentation was performed by two independent raters based on Hijikata's (2012, p. 38, Appendix 6) criteria, and we obtained high inter-rater agreements (97%). All disagreements were resolved through discussion. Each first sentence pair (i.e., HSS and LSS) was followed by a common second and third sentence to leave equal space between the target words and the semantic relatedness judgment test in both the forward and backward conditions (forward: 27.00 words; backward: 26.83 words in average), $t(11) = 0.43$, $p = .674$, $d = 0.20$.

Great care was taken to ensure that the second and third sentences did not semantically elaborate the meaning of target words. These sentences were created in cooperation with a native speaker of English such that the following two criteria were satisfied: (a) coherence, which indicates that the three sentences form a coherent story; and (b) contextual cues, which indicate that the second and third sentences do not provide helpful cues for inferring the meaning of target words. Rating data were collected from 23 Japanese EFL learners (none of them had participated in any of the main experiments) to verify the validity of the second and third sentences. First, a questionnaire asked them to evaluate the degree of coherence of each story. Participants read both 12 HSS and 12 LSS version stories, and then, answered whether each story was congruent or not using a 7-point Likert scale ranging from 1 *not at all coherent* to 7 *very coherent* (HSS: $M = 5.91$, $SD = 0.52$; LSS: $M = 5.84$, $SD = 0.42$). The high average rating showed that the stories were coherent and there was no significant difference between the inference and control versions, $t(11) = 0.43$, $p = .673$, $d = 0.15$. Next, participants were presented 12 target words and corresponding second and third sentences, and asked to judge whether or not the sentences allowed identification of the meaning of the presented words using a 7-point Likert scale ranging from 1 *not at all available* to 7 *very available*. The low average rating indicated that the second and third sentences did not allow the meaning generation to be made and did not differ between the forward condition ($M = 2.17$, $SD = 0.83$) and backward condition (*M* = 2.00, *SD* = 0.74), $t(11) = 0.48$, $p = .638$, $d = 0.22$. These manipulations ensure that the second and third sentences did not direct participants' attention to target words excessively.

The only difference in the procedure compared to Experiment 2A was that the chunk-parsed first sentences appeared one at a time, and the second and third sentences appeared sentence by sentence on the computer screen (see Figure 3.14). Participants were instructed to read each story for comprehension at their own pace, pressing a button on the response pad to indicate that they were ready for the next chunk or sentence. The same measures were taken as in Experiment 2A, and were analyzed in the same way. The outlier treatment procedure resulted in the substitution of about 1% of all observations.

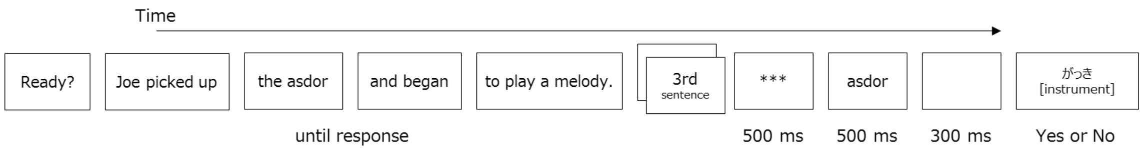


Figure 3.14. The sequence of events in each trial during the on-line semantic relatedness judgment test in Experiment 2B. The second and third sentences were presented sentence by sentence.

3.3.3 Results

Mean correct response rates of the filler items were 98%, indicating that participants completed the semantic relatedness judgment test properly. Table 3.12 shows the descriptive statistics for the test performance.

Accuracy. Figure 3.15 shows that accuracy was clearly higher in the HSS than LSS sentences. This observation was supported by a two-way ANOVA; there was a significant main effect of Context, $F(1, 27) = 118.48$, $p < .001$, $\eta^2 = .51$. Although the main effect of Direction did not reach significance, $F(1, 27) = 3.24$, $p = .083$, $\eta^2 = .01$, this effect on the accuracy interacted with the Context effect, $F(1, 27) = 9.57$, $p = .005$, $\eta^2 = .04$. Post hoc comparisons showed that the accuracy in the HSS condition was significantly higher in the forward elaboration than in the backward condition ($p < .001$, $d = 1.04$, $M_{\text{diff}} = .13$, 95% CI [.07, .19]). In the LSS condition, there was no difference between the forward and backward condition ($p = .412$, $d = 0.26$, $M_{\text{diff}} = –.04$, 95% CI [–.13, .05]).

Table 3.12

*Mean Accuracy Rates and RTs for the Semantic Relatedness Judgment Test (*N *= 28)*

	Forward elaboration			Backward elaboration		
Condition	*M*	95% CI	*SD*	*M*	95% CI	*SD*
			Accuracy rates			
HSS sentence	.81	[.77, .86]	.12	.68	[.63, .73]	.13
LSS sentence	.43	[.37, .49]	.15	.47	[.40, .53]	.16
			RTs in msec			
HSS sentence	831	[760, 903]	185	996	[914, 1,079]	213
LSS sentence	1,146	[1,071, 1,221]	194	1,258	[1,145, 1,371]	291

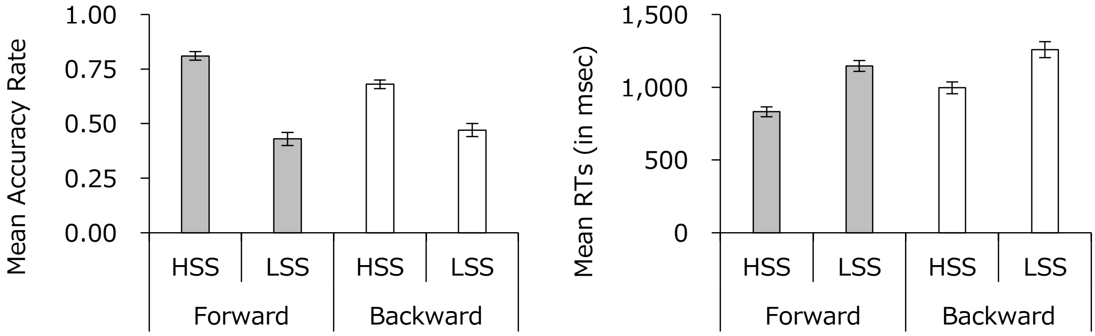


Figure 3.15. Mean accuracy rates and RTs with ±*SEM* bars in Experiment 2B.

Reaction times. Figure 3.15 indicates different trends in RTs data from Experiment 2B. A two-way ANOVA revealed significant main effects of Context, $F(1, 27) = 44.43$,

$p < .001$, $\eta^2 = .28$, and Direction, $F(1, 27) = 9.52$, $p = .005$, $\eta^2 = .06$. That is, RTs were shorter in the HSS than LSS condition. More importantly, RTs were shorter in the forward than backward condition. The interaction of Context × Direction was not significant, $F(1, 27) = 0.42$, $p = .523$, $\eta^2 < .01$.

3.3.4 Discussion

Similar to Experiment 2A, the LSA similarity had an influence on word meaning generation and encoding in both the forward and backward conditions. Again, accuracy for the HSS learning sentences was higher and RTs were faster compared to the LSS learning sentences. These results provide evidence that the meanings of the target words were encoded in a mental representation after processing the HSS learning sentences. Moreover, the Direction effect also appeared, showing that there were differences in the meaning generation process between forward and backward conditions. When participants encountered unknown words whose meanings were related to subsequent contextual information, judgment accuracy significantly decreased, and accordingly, RTs increased. This suggests that the meanings encoded by backward lexical inferences were generated more weakly than those encoded by forward lexical inferences. Therefore, it is possible that participants found backward lexical inferences more difficult than forward lexical inferences, and this supports the assumption that backward lexical inferences might require more complicated cognitive processes.

The judgment accuracy and RTs did not seem to worsen in spite of the insertion of the second and third sentences compared to Experiment 2A. This can be attributed to the nature of the semantic relatedness judgment test. In this test, regardless of the distance between the target word and probe word in a learning sentence, the target word was flashed again just before the probe word was presented. As the flashed target word reactivated the meaning of the probe word encoded in a mental representation, it could allow participants to retrieve it from their memory. For example, the meaning of *asdor*—inferred from *Joe picked up the asdor and began to play a melody*—could be reactivated by the target word (i.e., *asdor*) even if the memory of its meaning would weaken while reading the second and third sentences. Therefore, it is possible that the Direction effect led to different results between Experiments 2A and 2B. Although the participants could infer the meaning of unknown words in the backward inference condition, the backward lexical inferences were more difficult than the forward lexical inferences.

The overall findings support the current hypothesis and are consistent with previous L1 (e.g., Borovsky et al., 2010; Otten & van Berkum, 2008, 2009; Szewczyk & Schriefers, 2013; van Berkum et al., 2005) and bilingual (Martin et al., 2013; van Assche et al., 2011) studies although they focused on the contextual constraint effects instead of using the word-context semantic similarity provided by LSA. Early think-aloud research suggested that lexical inferences are complicated processes that require an informed guess for learners to engage their linguistic and non-linguistic knowledge to derive the meaning of unknown words (e.g., de Bot et al., 1997; Huckin & Bloch, 1993; Nassaji, 2006; Paribakht & Wesche, 1999; Wesche & Paribakht, 2010). Given that the ability to infer word meaning differs according to the quality of contextual information (Webb, 2008), lexical inferences will sometimes be difficult for L2 learners. Contrary to this view, the findings demonstrate that when a prior context semantically related to the possible meaning of upcoming words, Japanese EFL learners generate dynamic inferences such as word prediction, resulting in representing word knowledge that is relevant to contextual information.

In relation to the RQ4, generating the meaning of unknown words via backward lexical inference might involve three stages: (a) keeping an encountered unknown word in working memory (Cain et al., 2004; Daneman & Green, 1986), (b) finding a relevant highly elaborative context (Daneman & Green, 1986; Huckin & Bloch, 1993), and (c) integrating the inferred meaning with contextual information (Chaffin et al., 2001). The results of Experiment 2B, whether the participants strategically established a connection between an unknown word and contextual message, are relevant to stages (b) and (c). If the participants executed such a time-consuming integration process to infer the well-matched meaning of unknown words, accuracy and reaction times should have suffered when rereading intra-sentence was inhibited. As mentioned above, the judgment accuracy and RT was lower and slower in the backward compared to forward condition in Experiment 2B, supporting the idea that identifying the meaning of unknown words using backward lexical inferences requires effortful processes.

However, it should be noted that the comparison between the inference and control sentences revealed that in the backward condition, the meaning of the target words was also represented in the learners' mental representations in Experiment 2B. Although Huckin and Bloch (1993) demonstrated that L2 learners were able to wait for additional informative contexts to identify the possible meanings of unknown words, it is unclear whether special goals, such as completing a think-aloud task, elicited such a lexical inference behavior. The findings from Experiments 2A and 2B support the early think-aloud studies, suggesting that L2 learners are able to process unknown words by lexical inferences when those meanings are semantically elaborated by immediately-following contexts.

3.4 Summary of Study 1

Experiment 1 suggests that Japanese EFL learners activated the meaning of novel words via lexical inference that was sensitive to the semantic similarity between words and contexts computed by LSA (RQ1). Furthermore, the newly activated meaning was encoded in memory enough to immediately respond to the plausibility of verb-argument semantic links (RQ2). The ERP data revealed that the generation and encoding of lexical inference were not constrained by learners' L2 reading proficiency (RQ3). These findings were replicated by the forward inference condition in Experiments 2A and 2B using the semantic relatedness judgment test (H1). Other findings support the idea that Japanese EFL learners can establish the semantic connection between target words and their subsequent contextual information, although identifying the meaning of unknown words using backward lexical inferences requires effortful processes (RQ4).

Overall findings show that the word-context semantic similarity computed by LSA regulates the generation and encoding of lexical inference; however, the scope of its effect is still unclear in terms of semantic and lexical representations encoded in memory. In the experiments in Study 1, it is possible that participants did not infer the accurate meaning of critical words because of the limitation of the experimental procedures. For example, let us consider the following sentence pair with a pseudoword *marf*, used in Study 1.

(6a) She tried to put the pieces of the broken plate back together with *marf*.

(6b) She walked across the large room to Mike's dirty desk and returned his *marf*.

In both sentences, participants were implicitly required to infer from the contexts that the meaning of *marf* was "a glue." However, even though they could not accurately identify its meaning, it might not be directly reflected in the ERPs and RTs data. In other words, general inference such that the meaning of "a glue" is *a kind of tool* might influence the ERP waveforms and RTs. Given this methodological limitation, it is necessary to examine how specific meanings of unknown words are activated according to the word-context semantic similarity.

According to Chaffin (1997), the inferred meaning of unknown words typically includes information about either synonyms or hypernyms of those words. For example, the lexical inferences activated from sentence (6a) might be "a glue" as a synonym of *marf* or "tool" as a hypernym of *glue*. Following a hierarchical lexical network between hyponyms (e.g., *glue*) and hypernyms (e.g., *tool*), the former has more semantically specific characteristic than the latter (Miller et al., 1990; Rosch et al., 1976). Therefore, Study 2 considered the generation of synonymous meanings as *specific* lexical inferences, whereas that of hypernymic meanings is defined as *general* lexical inferences. In addition, L2 reading proficiency effect was included as a factor affecting the specificity of generated inferences, as predicted in Section 2.3.2.

Study 2 also addressed examining the lexical memory representations in order to reveal the relative difficulty in establishing form-meaning connections while reading in L2. Although the findings of Study 1 suggest that certain forms of semantic knowledge was constructed while reading in L2, it is still unclear why the semantic knowledge cannot be retrieved as measured by posttests such as recall and multiple-choice tests. As reviewed in Section 2.3.1, the theories of memory representation construction discourse processing suggest that the relative accessibility to surface representations in text memory will be decrease when the text meaning is constructed (e.g., Gerrig et al., 2009). Besides, a further experiment investigated the effects of task instruction on the construction of semantic and lexical memory representations of unknown words and the establishment of their connections.

Chapter 4

Study 2: Memory Representations of Words Established by Lexical Inference

4.1 Experiment 3A: Semantic Representations of Words Encoded in Memory

4.1.1 Design and Research Questions

The experiment presented here used the semantic relatedness judgment test in the same way as in Experiments 2A and 2B in order to test how specific inferences Japanese EFL learners can encode in a mental representation after processing unknown words. Given that the learners' L2 reading proficiency is an influential factor in narrowing down the possible inferable meaning, Experiment 3A was designed to address the interaction between contextual quality and L2 reading proficiency with two RQs summarized as follows:

RQ5: Does the strength of word-context semantic similarity affect the specificity of lexical inference encoded in memory?

RQ6: Are higher L2 reading proficiency levels required in order for EFL learners to establish a semantically specific memory of words?

As in LSS learning sentences, the prior experiments implied that Japanese EFL university participants could not generate and encode the specific lexical inference. However, if they are able to infer at least the general meaning of an unknown word from the context, the two-word relatedness judgments such as *marf-tool* in Example (6b) will be facilitated than those such as *marf-glue*. Furthermore, if the university participants indeed generate and encode the specific meaning of unknown words from HSS learning sentences, the prime-probe semantic relatedness judgments such as *marf-glue* in Example (6a) will be faster and more accurate than those such as *marf-tool* and *marf-glue* in Example (6b).

4.1.2 Method

4.1.2.1 Participants

Fifty-two Japanese EFL learners participated in the experimental session (22 females and 30 males; average age = 19.2, range = 18–21). None had participated in the previous experiments. They were undergraduate students at the University of Tsukuba, majoring in social studies, education, comparative cultures, international relations, engineering, chemistry, medicine, or nursing. At the time of the experiment, all had studied English for at least six years. Participants gave informed consent before the experiment and gained ¥1,500 for their participation.

The participants' English reading proficiency was estimated with the same test used in the prior experiments. A total of 27 participants were determined to have higher L2 reading proficiency based on a median split for their performance on the test (Cronbach's $\alpha = .82$). These 27 participants, regarded as the Upper group, showed substantially better performance on the test ($M = 17.41$, 95% CI [16.48, 18.33], $SD = 2.34$) than the other 25 participants categorized as the Lower group ($M = 8.76$, 95% CI [7.54, 9.98], $SD = 2.95$), $t(50) = 11.76$, $p < .001$, $d = 3.26$, $M_{\text{diff}} = 8.65$ (95% CI [7.17, 10.12]).

4.1.2.2 Materials

Probe words. Two pilot studies were conducted to confirm the validity of probe words used for a semantic relatedness judgment test. Because the meanings of unknown words inferred from contexts will be synonyms or hypernyms (Chaffin, 1997), two levels of the specificity of probes (hereafter P_{specific} and P_{general}) were created (see Table 4.1).

A total of 39 high-frequency target words as P_{specific} were selected from past studies (Griffin & Bock, 1998; van Assche et al., 2011). The hypernyms corresponding to each target word (e.g., *shark* → *fish*) were created as P_{general} based on WordNet's definition, "hyponym is a kind of hypernym" (Miller et al., 1990, p. 8). The task in the first pilot study asked 28 Japanese EFL university participants to make yes-no judgments on the hyponymy-hypernymy relationship between probe words, for example, "Is a *shark* a kind of *fish*?" The result showed that more than 80% of the respondents regarded 29 out of all the target words as having the hyponymy-hypernymy relationship (M = 94%, range: 82%–100%). The other target words were judged not to have such a connotational relationship (M = 48%, range: 14%–79%) and were excluded from the stimuli.

Table 4.1

Sample Sets of Stimuli Used in Experiment 3A

Conditions	Learning sentences	Probes
HSS/$P_{specific}$	The surfers were attacked by a dangerous *sind* in the sea.	サメ [shark]
HSS/$P_{general}$	The surfers were attacked by a dangerous *sind* in the sea.	さかな [fish]
LSS/$P_{specific}$	The group was surprised by a large *sind* in the sea.	サメ [shark]
LSS/$P_{general}$	The group was surprised by a large *sind* in the sea.	さかな [fish]
Filler	His coat was open because it was missing a *button*.	ライト [light]

Note. Target words are italicized.

In the second pilot study, 29 university participants took part in a lexical decision test to confirm no significant differences of probe recognition speed between $P_{specific}$ and $P_{general}$ without the corresponding prime stimuli. Each probe pair was translated into Japanese katakana or hiragana to avoid the effects of English word familiarity on probe recognition (average word length = 3.04, range: 2–4 letters). A total of 116 lexical decision items were paired with each of the 58 experimental sets and the same number of pronounceable Japanese nonwords. The participants had to determine if a probe presented on a computer screen was an existing word or a nonword as quickly as possible. A two-tailed paired t test found no significant differences in probe recognition between the probe pairs ($P_{specific}$: *M* = 536, 95% CI [504, 568], *SD* = 84; $P_{general}$: *M* = 550, 95% CI [514, 586], *SD* = 94), *t*(27) = 1.58, *p* = .126, *d* = 0.16, M_{diff} = 14 (95% CI [–4, 33]).

Learning sentences. The reading materials were 29 learning sentence pairs used in previous research (Griffin & Bock, 1998; van Assche et al., 2011), corresponding to the selected target probe words. However, nine out of 29 sentence pairs were removed as they included low-frequency words (level 5 and over) according to JACET (2003). Using LSA, I confirmed that the sentences categorized as HSS learning sentences had relatively high semantic similarity to the target words than the others did. There was a significant difference in the strength of word-context semantic similarity between the HSS (*M* = .38, 95% CI [.34, .43], *SD* = .10) and LSS learning sentences (*M* = .30, 95% CI [.25, .35], *SD* = .11), *t*(19) = 5.48, *p* < .001, *d* = 0.76, M_{diff} = .08 (95% CI [.05, .11]).

Each set of pseudowords (e.g., *sind*) had two types of probe words (i.e., $P_{specific}$ = サメ [same], $P_{general}$ = さかな [sakana]) in the semantic relatedness judgment test to test the encoding of the synonymous and superordinate meanings extracted from each learning sentence. Every set of 20 fillers also had similar probe words, but the target-probe pairs of the fillers were unrelated to each other (e.g., *button* → ライト [light]). The materials are all presented in Appendix 7.

4.1.2.3 Procedure

Participants worked individually in a single session that lasted about 50 minutes. First, the participants were notified of the general purpose of the study. After 30 minutes were given to take the English reading proficiency test, they were instructed on how to perform the semantic relatedness judgment test. The participants were presented with six practice sets to familiarize them with the procedure before the experimental session. SuperLab 4.5 for Windows and response pad (RB-730 model, Cedrus, the U.S.) was used to record the responses and RTs of the semantic relatedness judgment test.

In the semantic relatedness judgment test, the 10 HSS and 10 LSS learning sentences and the 20 fillers were presented in a random order to each participant. The participants read each learning sentence in the center of a computer screen at their own pace by pressing a button on a response pad. When they finished each sentence, a row of central fixation crosses appeared in the center of the screen for 500 msec; then, the crosses were replaced by the target prime word for 500 msec. After a 300-msec presentation of a blank sheet, a corresponding probe word was displayed. The participants were asked to judge whether the prime-probe pairs were semantically related by pressing the appropriate keys as swiftly and accurately as possible. Figure 4.1 shows the sequence of each trial.

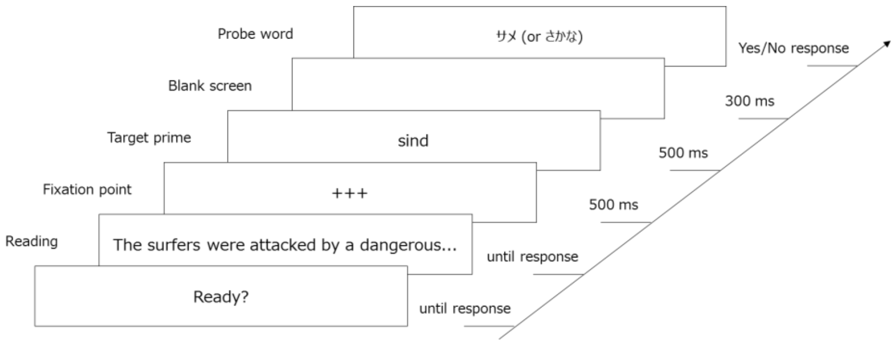


Figure 4.1. The sequence of events in each trial during the on-line semantic relatedness judgment test in Experiment 3A.

4.1.2.4 Data Analysis

To answer the RQs, the correct response rates and RTs of the semantic relatedness judgment test were analyzed in the same way as in Experiments 2A and 2B. RTs included in the analysis were from only correct responses per participant. Regarding outliers, RTs longer or shorter than *M*s ± 2.5 *SD*s per participant were replaced by scores of *M*s ± 2.5 *SD*s, respectively. This resulted in the substitution of 5% of all observations in each sentence condition. Additionally, the data from seven participants were excluded due to lack of the RTs data because their correct response rates were zero in any of the trials.

The between-participant variable was Proficiency (Upper and Lower), and the within-participant variables were Context (HSS and LSS) and Specificity ($P_{specific}$ and $P_{general}$).

4.1.3 Results

Accuracy. Table 4.2 shows mean accuracy rates of the test per condition. A first analysis used three-factor mixed ANOVA for the mean accuracy rates to investigate the interaction of Context × Proficiency on lexical inference specificity, encoded in sentence memory. The ANOVA results showed the significant main effects of Context, $F(1, 50) = 19.23$, $p < .001$, $\eta^2 = .04$, Specificity, $F(1, 50) = 41.39$, $p < .001$, $\eta^2 = .06$, and Proficiency, $F(1, 50) = 5.88$, $p = .019$, $\eta^2 = .05$. More importantly, the interaction of Context × Specificity was significant, $F(1, 50) = 6.98$, $p = .011$, $\eta^2 = .02$ (see also Table 4.3).

Figure 4.2 visualizes the different accuracy rates of the test per proficiency group. The first concern was to test the specificity of the meanings encoded from the LSS learning sentences. Post hoc analyses of the interaction between Context and Specificity showed that in the LSS condition, the mean accuracy rates were higher when $P_{general}$ was presented compared to $P_{specific}$, $F(1, 50) = 31.66$, $p < .001$, $\eta^2 = .39$, $M_{diff} = .22$ (95% CI [.14, .30]). As predicted, although the participants had relative difficulty inferring the specific meanings from the LSS contexts, they could generate and encode at least the general lexical inference for reading comprehension.

The HSS learning sentences elicited more correct responses than the LSS learning sentences did in the presentation of $P_{specific}$, $F(1, 50) = 24.36$, $p < .001$, $\eta^2 = .32$, $M_{diff} = .20$ (95% CI [.12, .28]). However, the mean accuracy rates did not differ between $P_{specific}$ and $P_{general}$ when they read the HSS learning sentences, $F(1, 50) = 1.67$, $p = .203$, $\eta^2 = .03$, $M_{diff} = .05$ (95% CI [–.12, .03]).

Table 4.2

Mean Accuracy Rates With 95% CIs and SD*s of the Semantic Relatedness Judgment Test*

		Specific inference						General inference					
		HSS learning sentence			LSS learning sentence			HSS learning sentence			LSS learning sentence		
Proficiency	*n*	*M*	95% CI	*SD*	*M*	95% CI	*SD*	*M*	95% CI	*SD*	*M*	95% CI	*SD*
Upper	27	.74	[.65, .83]	.23	.47	[.39, .56]	.21	.73	[.65, .80]	.19	.72	[.62, .81]	.24
Lower	25	.53	[.42, .63]	.26	.40	[.28, .52]	.30	.64	[.53, .75]	.26	.59	[.47, .71]	.29
Total	52	.64	[.56, .71]	.27	.44	[.37, .51]	.26	.68	[.62, .75]	.23	.66	[.58, .73]	.27

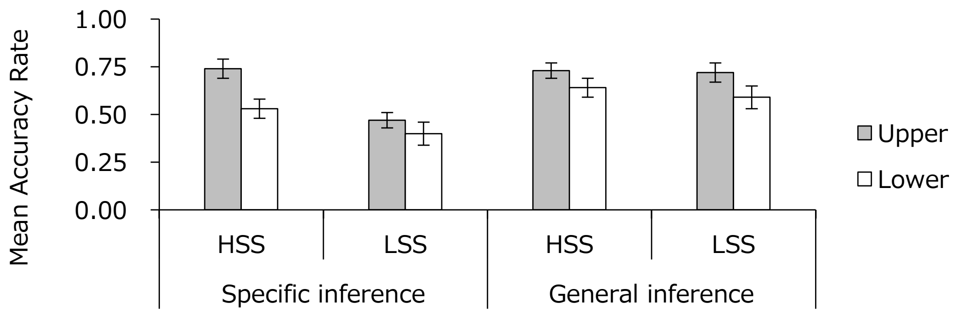


Figure 4.2. Mean accuracy rates with ±*SEM* bars of the semantic relatedness judgment test between Upper (*n* = 27) and Lower (*n* = 25) groups.

Table 4.3

ANOVA Results for Main Effects and Interaction Effects of Proficiency, Context, and Specificity on the Judgment Accuracy in the Test

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
			Between participants			
Proficiency (P)	1	0.81	0.81	5.88	.019	.06
Error (P)	50	6.88	0.14			
			Within participants			
Context (C)	1	0.66	0.66	19.23	< .001	.04
C × P	1	0.03	0.03	0.91	.344	< .01
Error (C)	50	1.71	0.03			
Specificity (S)	1	0.92	0.92	41.39	< .001	.06
S × P	1	0.02	0.02	0.80	.374	< .01
Error (S)	50	1.12	0.02			
C × S	1	0.37	0.37	6.98	.011	.02
C × S × P	1	0.10	0.10	1.95	.169	.01
Error (C × S)	50	2.68	0.05			

Reaction times. Table 4.4 presents the descriptive statistics for the RTs obtained from the semantic relatedness judgment test. The three-factor mixed ANOVA for the RTs showed significant main effects of Context, $F(1, 43) = 22.27$, $p < .001$, $\eta^2 = .04$, and Specificity, $F(1, 43) = 15.80$, $p < .001$, $\eta^2 = .04$. Although the main effect of Proficiency was not significant, $F(1, 43) = 1.17$, $p = .285$, $\eta^2 = .02$, the RTs were affected by a significant three-way interaction, $F(1, 43) = 4.49$, $p = .040$, $\eta^2 = .01$. Table 4.5 shows the overall results of the ANOVA.

Table 4.4

Mean RTs With 95% CIs and SDs *of the Semantic Relatedness Judgment Test*

		Specific inference						General inference					
		HSS learning sentence			LSS learning sentence			HSS learning sentence			LSS learning sentence		
Proficiency	*n*	*M*	95% CI	*SD*	*M*	95% CI	*SD*	*M*	95% CI	*SD*	*M*	95% CI	*SD*
Upper	26	1,008	[865, 1,151]	354	1,562	[1,259, 1,864]	749	1,103	[944, 1,262]	394	1,197	[957, 1,437]	594
Lower	19	1,463	[1,150, 1,776]	649	1,605	[1,291, 1,919]	652	1,096	[924, 1,267]	355	1,287	[1,026, 1,547]	541
Total	45	1,200	[1,037, 1,363]	543	1,580	[1,369, 1,791]	702	1,100	[988, 1,212]	374	1,235	[1,064, 1,406]	568

Note. The RTs data from seven participants were excluded (Upper: *n* = 1, Lower: *n* = 6) as described in Section 4.1.2.4.

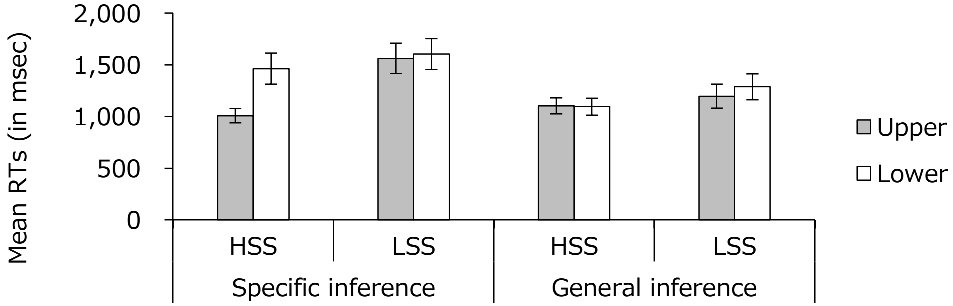


Figure 4.3. Mean RTs with ±*SEM* bars of the semantic relatedness judgment test between Upper (*n* = 26) and Lower (*n* = 19) groups.

Table 4.5

ANOVA Results for Main Effects and Interaction Effects of Proficiency, Context, and Specificity on RTs in the Test

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
Between participants						
Proficiency (P)	1	924754.82	924754.82	1.17	.285	.02
Error (P)	43	33957853.58	789717.53			
Within participants						
Context (C)	1	2639958.55	2639958.55	22.27	< .001	.04
C × P	1	272432.94	272432.94	2.30	.137	< .01
Error (C)	43	5097776.30	118552.94			
Specificity (S)	1	2504887.26	2504887.26	15.80	< .001	.04
S × P	1	477386.25	477386.25	3.01	.090	.01
Error (S)	43	6819011.89	158581.67			
C × S	1	462702.77	462702.77	2.93	.094	.01
C × S × P	1	708645.61	708645.61	4.49	.040	.01
Error (C × S)	43	6781679.52	157713.48			

Figure 4.3 graphically presents the differences of the RTs in each proficiency group. A first post hoc analysis examined the encoding of lexical inferences in the LSS learning sentences with a focus on each level of L2 reading proficiency. The mean RTs for $P_{general}$ were shorter than those for $P_{specific}$ regardless of the participants' L2 reading proficiency, Upper: $F(1, 43) = 6.15$, $p = .017$, $\eta^2 = .12$; Lower: $F(1, 43) = 4.10$, $p = .049$, $\eta^2 = .08$. Second, the proficiency effect was appeared in the HSS learning sentences. Whereas there was no difference in the mean RTs between $P_{specific}$ (1,008 msec) and $P_{general}$ (1,103 msec) in the Upper group, $F(1, 43) = 0.00$, $p = .975$, $\eta^2 < .01$, the mean RTs of the Lower group

were longer when $P_{specific}$ was presented (1,463 msec) compared to $P_{general}$ (1,096 msec), $F(1, 43) = 12.86$, $p = .001$, $\eta^2 = .23$.

4.1.4 Discussion

In relation to RQ5, the mean accuracy rates of the semantic relatedness judgment test provided evidence that the specificity of lexical inference changed according to the word-context semantic similarity quantified by LSA. First, in the LSS learning sentences, the mean accuracy rates for the presentation of $P_{general}$ (e.g., *sind-fish*) were significantly higher. This means that the LSS learning sentences drew out a general lexical inference, which can broadly match a given contextual sentence in meaning. Previous think-aloud studies have also shown consistent results that readers make such general or vague lexical inferences during both L1 (e.g., Fukkink et al., 2001) and L2 reading (e.g., Hamada, 2011; Huckin & Bloch, 1993).

Higher accuracy were obtained when $P_{specific}$ (e.g., *sind-shark*) was presented in the HSS learning sentences than in the LSS ones. Although this result partially supports the assumption that the participants generated and represented the specific meaning of the target words in their sentence memory, it should be noted that the mean accuracy rates for $P_{general}$ in the HSS contexts were also high to the same degree of those for $P_{specific}$. Accordingly, it is necessary to consider the RTs data together, because an alternative interpretation still remains as follows: Participants initially encoded a general inference in their mind (e.g., *fish*), but when $P_{specific}$ (e.g., *shark*) was flashed on the computer screen, they could change their general inferences for specific ones by checking the meaning of $P_{specific}$ with their encoded general inference based on the learning sentence's propositions. If so, the RTs for the presentation of $P_{specific}$ will be longer compared with the case that they initially generate and encode a specific lexical inference. This issue is discussed further in the analysis of the RTs.

The other finding of the ANOVA was the effects of L2 reading proficiency on lexical inference specificity (a partial answer to RQ6). The result indicates that the skilled participants responded to either probe type more accurately than the less-skilled ones did. In other words, the HSS learning sentences should promote the generation and encoding of specific meanings of unknown words; nevertheless, lower L2 reading proficiency will inhibit this benefit for lexical inference. Similarly, in the LSS learning sentences, the limitations of the learners' L2 reading proficiency decrease the mean accuracy rates for the presentation of $P_{general}$. These findings are supported by many previous conclusions that learners' reading proficiency affects lexical inference success (Fukkink et al., 2001; Haastrup, 1991; Nassaji, 2006; Wesche & Paribakht, 2010). Furthermore, Experiment 3A suggests that L2 reading proficiency affects the specificity of the meanings encoded in sentence memory.

As a further discussion of RQs 5 and 6, the RTs data from the semantic relatedness judgment test indicated that the encoding level of the inferred meaning was affected by the interaction between contextual quality and L2 reading proficiency. First, the higher L2 proficiency learners could respond to both $P_{specific}$ and $P_{general}$ at almost the same speed when the meaning of the target words was semantically related to the sentence meaning. As stated above, it would take a longer time to respond to $P_{specific}$ after the participants strongly encoded the general meaning of the target words in their sentence memory; therefore, the possibility that the learners with higher L2 reading proficiency changed their general inferences for more specific ones by checking the flashed $P_{specific}$ can be ruled out. Additionally, the insignificance of the difference in RTs between the probe types suggests that the participants simultaneously encoded the general inferences as a possible meaning of the target words. That is, good learners might initially encode the specific meaning of unknown words, keeping the general inferences in their mind, while they are reading the HSS contexts.

On the other hand, even though the sentence meaning was semantically related to the inferable meaning of the target words, the less-skilled learners' RTs were longer for $P_{specific}$ than for $P_{general}$. This finding has two likely interpretations; one is that such learners inferred the general meaning of the target words from the HSS learning sentences at first, and then, the flashed $P_{specific}$ changed their inferences into specific, leading to relatively high accuracy rates for $P_{specific}$. The other account is that those learners could activate the specific meanings of the target words but needed more time to derive such meanings using word-context semantic similarity.

As in the LSS learning sentences, the mean RTs for $P_{general}$ were shorter than for $P_{specific}$ regardless of the learners' L2 reading proficiency. This demonstrates that they indeed encoded the general meaning of unknown words in sentence memory after they comprehended the LSS learning sentences. Whereas most past studies have concentrated on only the accuracy of lexical inferences (e.g., de Bot et al., 1997; Fukkink et al., 2001; Huckin & Bloch, 1993; Nassaji, 2003, 2006; Wesche & Paribakht, 2010), an important suggestion of the current experiment is that the L2 learners' inference specificity can vary from specific to general lexical inferences on-line for reading comprehension.

4.2 Experiment 3B: Follow-Up Study of Experiment 3A

4.2.1 Design and Hypotheses

Experiment 3B employed an off-line lexical inference test to obtain a similar result to Experiment 3A. The off-line test allows to analyze the inferences encoded in participants' memory both quantitatively and qualitatively. It will confirm that the nature of the on-line semantic relatedness judgment test did not provide a distorted observation of the outcomes of Experiment 3A.

In Experiment 3A, skilled L2 readers derived the specific meanings of unknown words after understanding the learning sentences contrary to less-skilled L2 readers.

However, the results also indicated that less-skilled L2 readers could encode the specific meanings of unknown words from the HSS learning sentences but needed enough time to strategically narrow down the inferable meanings. If so, the off-line lexical inference test provides the evidence that those learners can infer and encode the specific meanings of unknown words from the HSS learning sentences because they are given sufficient time to integrate the target word meanings into mental representations in this test. On the other hand, as the results of Experiment 3A showed, the LSS learning sentences should not be able to elicit the specific inference even if the test-on-time is sufficiently provided. Together the off-line lexical inference test further examined the word-context semantic similarity computed by LSA on the specificity of encoded inferences from the viewpoint of the strategic processing, which was designed to address two hypotheses:

H2: The higher word-context semantic similarity elicits the specific meanings of target words by lexical inference.

H3: The lower word-context semantic similarity elicits the general meanings of target words by lexical inference.

4.2.2 Method

4.2.2.1 Participants

Forty Japanese EFL learners participated in the experimental session (21 females and 19 males; average age = 19.5, range = 18–23). None had participated in the prior experiments. They were undergraduate students at the University of Tsukuba, majoring in humanities, social sciences, engineering, biology, medicine, and nursing. At the time of this experiment, all had studied English for at least six years in Japan. Participants gave informed consent before the experiment and gained ¥1,000 for their participation.

The English reading proficiency test was assigned to classify the participants into two proficiency groups in the same manner as in the previous experiments. A two-tailed paired *t* test confirmed that half of the participants, regarded as the Upper group (M = 17.30, 95% CI [15.89, 18.71], SD = 3.01), showed better performance on the test than the other 20 participants categorized as the Lower group (M = 10.55, 95% CI [9.54, 11.56], SD = 2.16), $t(38) = 8.14$, $p < .001$, $d = 2.58$, $M_{\text{diff}} = 6.75$ (95% CI [5.07, 8.43]).

4.2.2.2 Materials

The target words and learning sentences were the same as those used in Experiment 3A. There were 20 English-like nonwords as the target words in order to ensure that the participants had no prior knowledge of the words presented.

These target words were underlined and presented in the corresponding learning sentences with the following small modifications. Instead of using a computer, two types of booklets were prepared to provide the learning sentences and an answer sheet. Each booklet included the 10 HSS and 10 LSS learning sentences, which were counterbalanced between the booklets. Additionally, confidence ratings for attempts for lexical inference by a 5-point Likert scale were inserted to examine the effects of context quality on learners' self-confidence in inferences. According to Rebuschat (2013), the analysis of the confidence ratings indicates that participants are aware of having acquired knowledge. As reviewed in Section 2.5.1, it is possible that during a think-aloud task, participants would avoid reporting their specific comprehension of unknown words. Experiment 4B, therefore, aims at complementing the findings reported in previous think-aloud studies (e.g., Huckin & Bloch, 1993; Paribakht & Wesche, 1999).

4.2.2.3 Procedure

The participants were tested individually or in two to three members in a single session that lasted for 40 minutes. After the explanation of the general purpose of the study, they were asked to complete the English reading proficiency test within 30 minutes. One of the two booklets was randomly assigned to each participant, and then they were instructed on how to perform the lexical inference test.

In the test, the participants were asked to write the meanings of the target words in Japanese that first came to mind when reading the learning sentences at their own pace, which took about 10 minutes. At the same time, they rated their confidence level of their attempts for each lexical inference (range = 1: *not sure at all* to 5: *definitely sure*).

4.2.2.4 Scoring and Data Analysis

Responses were classified based on their semantic relation to the original meaning of the target words using the relation categories adapted from Chaffin (1997), as shown in Table 4.6. There were two semantic relations between target and generated words as *definitional* relations: synonym and category. The meanings generated were classified as *synonyms* if the Japanese words could refer to the same thing that the original meanings of the target words did. The criteria for *category* responses were that one word can be more semantically general than the other; such relationship between the two words is defined as the frame, "A *specific* is a kind of a *general*," as in Experiment 3A (e.g., *coffee* → *something to drink*). Following Hamada (2011), additional criteria for *associates* were added as *event-based* relations and responses that were contextually appropriate were regarded as associates (e.g., responses such as *snack*, *bread*, and *egg* could be contextually fit in *He wants to stop for moment because he wants to buy a pack of this* _______ *in the shop*). Finally, responses that did not satisfy any of these criteria were defined as *no identifiable relation* (e.g., *coffee* → *toy*). Two raters individually rated all the responses, resulting in a 94% inter-rater agreement. All disagreements were resolved by discussion.

Proportions were calculated for each participant per proficiency group by dividing the frequency of each response type by the overall frequency of produced responses. A three-factor mixed ANOVA was run for the mean production rates; it included Proficiency (Upper and Lower) as a between-participants variable, and Response (Synonym, Category, Associate, and Non-relation) and Context (HSS and LSS) as within-participants variables.

Table 4.6

Taxonomy of Semantic Relations Used to Classify Meanings Generated in the Test

Relations	Criteria
Definitional relations	
Synonym	Responses are the same as target words.
Category	Responses/target words are kinds of target words/responses.
Event-based relations	
Associate	Responses are contextually appropriate.
Non-semantic relations	
No identifiable relation	Responses do not satisfy any of the above criteria.

4.2.3 Results

Lexical inference test. Table 4.7 shows the mean ratio of each type of responses produced in the lexical inference test. Importantly, the ANOVA result showed that the interaction of Response × Context affected the outcomes of lexical inference, $F(3, 114) = 60.75$, $p < .001$, $\eta^2 = .21$. A main effect of Proficiency was not significant, contrary to Experiment 3A, $F(1, 38) = 3.11$, $p = .086$, $\eta^2 < .01$, and did not interact with other factors.

Figure 4.4 visualized the differences in each response between the HSS and LSS learning sentences. Post hoc comparisons showed that the participants produced more responses synonymously related to the target word meanings from the HSS learning sentences than LSS ones, $F(1, 39) = 324.79$, $p < .001$, $\eta^2 = .89$. On the other hand, the LSS learning sentences elicited more category, $F(1, 39) = 7.43$, $p = .010$, $\eta^2 = .16$, and associate responses, $F(1, 39) = 200.31$, $p < .001$, $\eta^2 = .84$, than the HSS learning sentences did. A difference between the HSS and LSS learning sentences for Non-relation was not found, $F(1, 39) = 0.73$, $p = .398$, $\eta^2 = .02$. Table 4.8 presents the other results of ANOVA.

Table 4.7

Means With 95% CIs and SDs *of Responses Produced in the Test*

	Upper (*n* = 20)			Lower (*n* = 20)			Total (*N* = 40)		
Relation	*M*	95% CI	*SD*	*M*	95% CI	*SD*	*M*	95% CI	*SD*
				HSS learning sentence					
Synonym	.57	[.50, .64]	.15	.58	[.52, .64]	.12	.57	[.53, .62]	.14
Category	.09	[.04, .14]	.10	.07	[.04, .10]	.07	.08	[.05, .11]	.08
Associate	.23	[.17, .28]	.12	.20	[.15, .25]	.12	.21	[.18, .25]	.12
Non-relation	.12	[.07, .17]	.11	.15	[.09, .21]	.12	.13	[.10, .17]	.12
				LSS learning sentence					
Synonym	.21	[.13, .30]	.18	.16	[.10, .21]	.12	.19	[.14, .23]	.15
Category	.16	[.09, .22]	.14	.16	[.10, .21]	.11	.16	[.12, .20]	.13
Associate	.51	[.43, .59]	.17	.55	[.48, .61]	.14	.53	[.48, .58]	.16
Non-relation	.12	[.08, .16]	.08	.14	[.10, .18]	.08	.13	[.10, .16]	.08

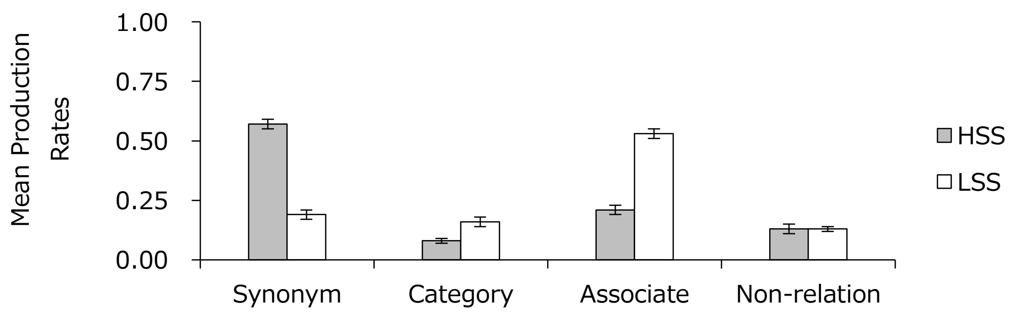


Figure 4.4. Mean production rates of each response with ±*SEM* bars in the lexical inference test between the HSS and LSS learning sentences (*N* = 40).

Table 4.8

ANOVA Results for Main Effects and Interaction Effects of Proficiency, Response Type, and Context on Lexical Inference Outcomes

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
			Between participants			
Proficiency (P)	1	0.00	0.00	3.11	.086	< .01
Error (P)	38	0.00	0.00			
			Within participants			
Response (R)	3	1.94	0.65	37.89	< .001	.13
R × P	3	0.03	0.01	0.60	.615	< .01
Error (R)	114	1.94	0.02			
Context (C)	1	5.03	5.03	252.18	< .001	.34
C × P	1	0.01	0.01	0.24	.629	< .01
Error (C)	38	0.76	0.02			
R × C	3	3.19	1.06	60.75	< .001	.21
R × C × P	3	0.04	0.01	0.69	.559	< .01
Error (R × C)	114	2.00	0.02			

Confidence rating for lexical inference. Table 4.9 and Figure 4.5 show mean confidence levels per condition. To examine that the HSS and LSS learning sentences affected the confidence of lexical inference differently, a main effect of Context was tested and found significant, $F(1, 38) = 99.49$, $p < .001$, $\eta^2 = .39$. It was due to higher confidence ratings for the HSS learning sentences than for the LSS ones. This context effect did not differ between the Upper and Lower groups because there were no significant interaction of the two factors, $F(1, 38) = 1.21$, $p = .277$, $\eta^2 = .01$, and a main effect of Proficiency, $F(1, 38) = 1.91$, $p = .175$, $\eta^2 = .02$. The overall results of the ANOVA are reported in Table 4.10.

Table 4.9

Mean Rating Scores With 95% CIs and SD*s for the Lexical Inference Confidence*

Proficiency	*n*	HSS learning sentence			LSS Learning sentence		
		M	95% CI	*SD*	*M*	95% CI	*SD*
Upper	20	3.50	[3.28, 3.72]	0.47	2.71	[2.43, 3.00]	0.61
Lower	20	3.39	[3.15, 3.63]	0.51	2.41	[2.12, 2.69]	0.61
Total	40	3.44	[3.29, 3.60]	0.49	2.56	[2.36, 2.76]	0.62

Note. A possible score range is 1.00 to 5.00.

Table 4.10

ANOVA Results for Main Effects and Interaction Effects of Proficiency and Context on the Confidence Level in Lexical Inference

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
Between participants						
Proficiency (P)	1	0.87	0.87	1.91	.175	.02
Error (P)	38	17.41	0.46			
Within participants						
Context (C)	1	15.58	15.58	99.49	< .001	.39
C × P	1	0.19	0.19	1.21	.277	.01
Error (C)	38	5.95	0.16			

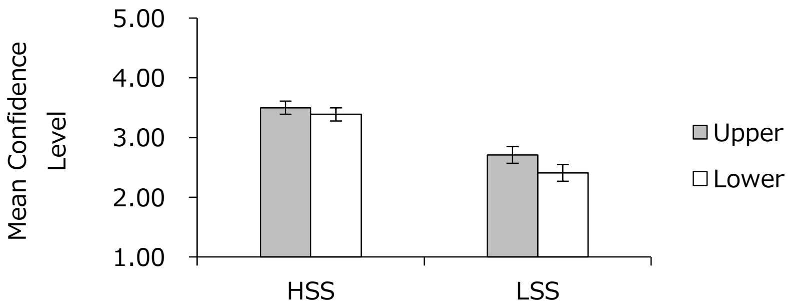


Figure 4.5. The mean rating scores of confidence with ±*SEM* bars in the lexical inference test per condition.

4.2.4 Discussion

In relation to H2, responses to unknown words presented with the HSS learning sentences were more likely to be synonyms of the intended meanings of those words compared with the other response types. This result strengthens the conclusion of Experiment 3A, suggesting that the participants in Experiment 3A could generate and encode the specific meaning of unknown words in memory while trying to comprehend the HSS learning sentences.

Although Chaffin (1997) demonstrated that the outcomes of meaning generation often converged on either synonyms or hypernyms of target unknown words in L1 reading, the participants in Experiment 3B produced more associated responses than categorical responses. A possible explanation is that the fineness of constructed mental representation of a word was significantly different between L1 and L2 reading. When readers encounter an unknown word in a text, they try to determine what kind of concept the word represents. At this time, establishing the semantic field to which the word belongs occurs, which allows the word meaning to be associated with other word meanings in the same field in a mental lexicon (Chaffin, 1997; Chaffin et al., 2001). On the other hand, even though the HSS learning sentences are related to the inferable category to which unknown words belong, L2 learners are more likely to build not the definitional relations, but the event-based relations between unknown words and known concepts.

These results support some investigations that examined the outcomes of lexical inference with relatively less-skilled readers such as L1 children and L2 learners. Fukkink et al. (2001) demonstrated that less-skilled readers could not extract the decontextualized meaning of unknown words from contexts and made some vague lexical inferences that widely matched the contextual information in meaning. Hamada (2011) also showed that the degree of lexical inference success affected the establishing semantic relations among words. Taken together, the current findings demonstrate that the difference of the mental representation of a word distinguishes the specificity of lexical inferences, resulting from L2 reading proficiency.

In the LSS learning sentences, word memory encoded were more likely to converge on associated responses instead of synonymous responses. The other response types were also fewer than associated responses. Thus, unexpectedly, the participants did not produce more categorical responses, such as probes used in Experiment 3A, related to the intended meaning of unknown words from the LSS learning sentences. This result is inconsistent with the findings of Experiment 3A in terms that the participants inferred the hypernymic meaning (i.e., category response) of unknown words from the LSS learning sentences. In Experiment 3A, in order to respond to the flashed $P_{general}$ (e.g., *sind-fish*) as accurately and swiftly as possible, the participants must encode the same concepts as the probes or the categorical meaning at least. However, in Experiment 3B, the total production rates of synonymous and categorical responses produced from the LSS learning sentences were 34%, which is still less than the associated responses (53%). This result is the same as in the HSS learning sentences, suggesting that the participants produced widely interpreted meanings from the event-based relations between the possible meaning of an unknown word and sentence representation. Whereas Chaffin (1997) used only sentences that fit meaningfully with the original meanings of nonwords, Experiment 3B used the contexts with LSS to the target word and its category. This suggests that participants had no choice but to use the events established from the LSS learning sentences to generate the meaning of unknown words.

Although H3 was rejected, some previous research on vocabulary learning from reading supports the results of Experiment 3B. For example, Bolger et al. (2008) showed that L1 readers first acquire event-based, contextualized meaning of words but not their decontextualized definitional meaning from contexts. That is, the partial knowledge of a word acquired from a single context is based on the event expressed in the context. Also, L2 studies explained that lexical development is a lengthy, incremental process, normally requiring multiple exposures to new words in various contexts (Chen & Truscott, 2010; Elgort et al., 2015; Rott, 1999, 2007; Webb, 2007b, 2008; Waring & Takaki, 2003). Thus, accounts for the lexical development suggests that the event-based meanings encoded in memory by lexical inferences will grow to be a partial to full knowledge of a word.

The insignificant proficiency effects is a clear difference between Experiments 3A and 3B. In the off-line lexical inference test, the Lower group could generate and encode the specific meaning of unknown words from the HSS learning sentences. In contrast, Experiment 3A showed that in the semantic relatedness judgment test, the Lower group could not encode the specific lexical inferences during on-line sentence processing. These results offer a reasonable interpretation that the less-skilled L2 readers need enough time to identify the intended meanings of unknown words strategically. In other words, those learners first generate vague lexical inferences and gradually narrow down them into specific ones. On the other hand, the lexical inference outcomes did not converge on the specific concepts of unknown words in the LSS learning sentences even when both the Upper and Lower groups attempted lexical inference strategically. This result was highly expected and consistent with past findings that L2 learners fail in generating definitional meanings of unknown words from low-guessability contexts (e.g., Bensoussan & Laufer, 1984; Huckin & Bloch, 1993; Laufer, 1997; Li, 1988; Mondria & Wit-de Boer, 1991).

Although the associated responses semantically matched the propositions of the learning sentences, that is, those responses were not necessarily incorrect word meanings, the LSS learning sentences vitiated the participants' confidence levels. Hulstijn (1993) also demonstrated that L2 learners were sensitive to guessability of unknown words from context. In his study, L2 learners were more likely to consult a dictionary when they were worried if they could generate the correct meaning of unknown words even though the word meaning generated were correct. Furthermore, when they had confidence in their outcomes of lexical inference, they reduced the relative frequency of dictionary use. Thus, these findings show that L2 learners feel more uncertain about the event-based meanings than about the definitional meanings encoded in memory by lexical inference.

The confidence rating data was associated with the strength of word-context semantic similarity of LSA. Although the current experiments used only dichotomous contexts including an unknown word, the statistical analyses suggest that the confidence level of meaning generation increases as the word-context semantic similarity strengthens. First, when contextual information was highly semantically related to the inferable meanings of the target words, even less-skilled L2 readers performed the lexical inference test with plenty of confidence. Given that concrete mental presentations of texts increase the guessability of an unknown word (Hasegawa, 2012; Morimoto, 2006), the availability of word-context semantic similarity promotes to generate and encode the specific concept of target words with confidence.

Although both Experiments 3A and 3B provided evidence of the generation and encoding of lexical inference during L2 reading comprehension, it is still unclear why the established meaning memory could not be retrievable, as measured by a meaning recall test (e.g., Mondria & Wit-de Boer, 1991; Webb, 2008). As mentioned in Sections 2.3.1 and 2.3.2, one of the possible problems is the difficulty in incidental establishment of form-meaning connections as a consequence of discourse processing. To test this assumption, further experiments were conducted to investigate the activation of lexical representation of unknown words after their semantic representation was built in memory.

4.3 Experiment 4A: Lexical Representations of Words Encoded in Memory

4.3.1 Design and Research Question

The present RT experiment tested whether the establishment of semantic representations of unknown words by lexical inference changed the relative accessibility to their word-form memory. As suggested in Section 2.3.1, I predicted that the successful inferences change the structure of text memory in terms of semantic and lexical representations of words. If so, participants would be quicker in recognizing that the form of target words had appeared in the story when they could not establish a semantic representation of the word by lexical inference than when they could do so. Accordingly, Experiment 4A used a probe recognition test to address the following RQ:

RQ7: Does successful lexical inference change the relative accessibility of the word-form memory of unknown words?

4.3.2 Method

4.3.2.1 Participants

Twenty Japanese EFL learners participated in the experiment (12 females and eight males; average age = 19.2, range = 18–24). They were undergraduate or graduate students at the University of Tsukuba, majoring in humanities, social studies, education, engineering, biology, or medicine. At the time of the experiment, all had studied English for at least six years in Japan. Their estimated proficiency in English was at the B1 or B2 levels of the CEFR based on their self-report. Participants were paid a sum of ¥1,000 for their participation after they gave written informed consent to join this study, approved by the research ethics committee of the University of Tsukuba.

4.3.2.2 Materials

Learning sentences and probes. The same learning sentences and target probe words were used as in Experiment 2A. Twelve target unknown words embedded in short stories were divided into two lists with each target word only presented once per participant. Two types of first sentences always shared a target word (HSS and LSS). The subsequent two sentences were common for each first sentence. It should be noted that the second and third sentences did not provide any contextual cues for generating lexical inference or direct participants' attention to target words excessively. Every set of 12 fillers also had similar probe words, but the probes selected were known words in the first sentences in order to counterbalance the number of "Yes" and "No" responses in a probe recognition test. The word length and the number of syllables of the probes were the same as the experimental sets. Table 4.11 shows a sample set of stimuli.

Table 4.11

Sample Sets of Stimuli Used in Experiment 4A

First sentence
The train always arrives on time at the *xxx* in Tokyo. (HSS)
My friend sometimes arrives too late at the *xxx* in Tokyo. (LSS)
Second and third sentences
However, today, a car accident threw traffic into great confusion. For this reason, I broke my promise to play with my friend, and I had to apologize to her.

Note. The target nonword *xxx* refers to *station*.

Multiple-choice test. To verify whether participants could infer the meaning of target unknown words while reading each story, they received an unexpected multiple-choice test after all tests were completed. This posttest indirectly assessed the accuracy of lexical inferences between the HSS and LSS learning sentences because in a probe recognition test, any interventions (e.g., a questionnaire, interview, and think-aloud task) were not available for insertion. In this test, 12 words were presented with equivalents and three distracters in Japanese. The distracters consisted of the same parts of speech as the target words (noun), which appeared in filler stories. Participants were encouraged to mark the possible meaning of the target words from four given options even though they were not sure of form-meaning connections. When they failed to identify the unknown word meanings, the accuracy in the posttest suffered in an observable manner.

Originally, I predicted that form-meaning connections were not established in the HSS condition, and one might think that the accuracy in the multiple-choice test showed participants' ability to connect the target forms with inferred meanings (e.g., Nation, 2013; Nation & Webb, 2011; Schmitt, 2010). However, even though participants do not establish a form-meaning connection, it is possible that they can choose a correct option from their text memory including the inferred meanings of the target words. In contrast, an L2-L1 translation test should assess the degree of form-meaning connections because learners are required to retrieve an inferred meaning corresponding to a test item.

4.3.2.3 Procedure

The participants worked individually in a single session that lasted approximately for 50 minutes. After 30 minutes were provided for taking the English reading proficiency test, they were instructed on how to work on the probe recognition test. Six practice sets were provided to the participants in order to familiarize them with the procedure. SuperLab 4.5 for Windows (Cedrus, the U.S.) and response pad (RB-730 model, Cedrus, the U.S.) were used to provide all of the practices, instructions, experimental sets presentation, and data collection (i.e., responses and RTs).

Figure 4.6 illustrates the sequence of events in each trial of the probe recognition test. Six HSS sets, six LSS sets, and 12 filler sets were presented in a random order to each participant. The participants read each story in the center of a computer screen, sentence by sentence at their own pace by pressing a button on a response pad. When they finished each story, a row of central fixation crosses appeared in the center of the screen for 500 msec; then, the crosses were replaced by the target word form. The participants were required to judge whether or not the flashed recognition probes had appeared in the sentences that they had just read by pressing yes or no keys as quickly and accurately as possible. They were not explicitly instructed to generate the meanings of target unknown words. After the completion of all the probe recognition trials, the participants answered the multiple-choice test, which took about five minutes.

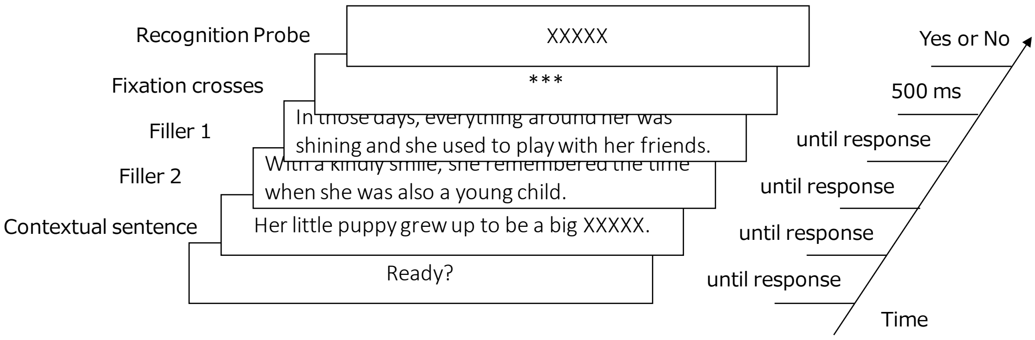


Figure 4.6. The sequence of events in each trial during the on-line probe recognition test.

4.3.2.4 Data Analysis

The average score of the multiple-choice test was compared between the HSS and LSS conditions by a two-tailed paired *t* test. RTs included in the analysis consisted only of the correct responses per participant. Regarding outliers, RTs longer or shorter than *M*s ± 2.5 *SD*s per participant were replaced by scores of *M*s ± 2.5 *SD*s, respectively. This resulted in the substitution of less than 1% of all observations in each condition. The mean correct responses and RTs were compared using a two-tailed paired *t* test in order to examine the context quality effects on the probe recognition performances.

4.3.3 Results

Analysis of the multiple-choice test. The performance of the multiple-choice test was better in the HSS condition (M = .63, 95% CI [.53, .74], SD = .22) than in the LSS condition (M = .24, 95% CI [.15, .33], SD = .19), $t(19) = 5.10$, $p < .001$, $d = 1.90$, M_{diff} = .39 (95% CI [.23, .55]).

Analysis of the probe recognition test. Table 4.12 shows the descriptive statistics for the probe recognition performances. Mean accuracy rates of the filler items were high enough (M = .95, 95% CI [.91, .98], SD = .07), indicating that participants completed the probe recognition test properly. Mean accuracy rates of the target items were also high and did not differ between the HSS and LSS conditions, $t(19) = 0.15$, $p = .883$, $d = 0.06$, M_{diff} = .01 (95% CI [–.13, .11]). These ensure that the RTs data obtained were highly satisfactory for analysis. The result of the *t* test showed that the recognition speed of the unknown words in the HSS condition were slower than that in the LSS condition, $t(19)$ = 5.66, $p < .001$, $d = 0.57$, M_{diff} = 143 (95% CI [90, 196]).

Table 4.12

Mean Accuracy and RTs With 95% CIs and SD*s of the Probe Recognition Test (*N *= 20)*

	Accuracy			RTs (in msec)		
Conditions	*M*	95% CI	*SD*	*M*	95% CI	*SD*
HSS sentence	.85	[.76, .94]	.19	1,095	[969, 1,221]	270
LSS sentence	.86	[.79, .93]	.16	952	[846, 1,059]	228

4.3.4 Discussion

The results of the multiple-choice test showed that the participants were more able to construct the meaning-based mental representation of the target words in the HSS condition than in the LSS condition. As predicted, this is comparable with the results of the prior experiments despite the different method involved in assessing the generation encoding of lexical inference. Because the presupposition behind the logic of Experiment 4A was satisfied, the RTs data obtained from the probe recognition test should be affected by the different memory representations of each story between the HSS and LSS conditions.

For an answer to RQ7, the recognition speed was slower when the participants read the target unknown words in the HSS learning sentences than in the LSS ones. This result supports the assumption that the word-form memory of unknown words will become less accessible when the meanings of unknown words are represented in memory by successful lexical inference. In contrast, in the absence of reliable contextual cues for lexical inference such as the word-context semantic similarity, the participants were more swiftly able to recognize that the letter strings had appeared in the stories. Although the nature of the probes used in previous research (e.g., character names: Gerrig et al., 2009; Love et al., 2010; antecedents: Dopkins & Nordlie, 2011; Gernsbacher, 1990) was different from Experiment 4A, the finding is consistent with these studies. According to them, because both protagonists' names and antecedents play a central role in narrative comprehension, the unresolved text elements are activated in the mental representations in order to be available to subsequent text processing such as the integration of additional information with the unresolved element. This perspective can be applicable to unknown words because they create holes in L2 learners' comprehension of a text if they are not appropriately interpreted (Bensoussan & Laufer, 1984; Paribakht & Wesche, 1999; Pulido, 2007). Actually, the unresolved meaning of unknown words often elicit additional text processing such as strategic rereading (Cain et al., 2004; Chaffin et al., 2001; Daneman & Green, 1986; Huckin & Bloch, 1993) and dictionary use (Hulstijn, 1993), resulting in allocating more focal attention (Godfroid et al., 2013; Pellicer-Sánchez, 2015). These suggest that the words unresolved in discourse become salient in text memory until their appropriate meanings are generated or provided.

The decrease in the accessibility of word-form information suggests that the verbatim text memory was transformed into a semantic representation of the text as described by van Dijk and Kintsch's (1983) multilevel representation model. However, in incidental word learning from reading, such asymmetric activation and encoding of a word form and its meaning in a mental representation would be unfavorable because the establishment of form-meaning connections requires their simultaneous activation (e.g., Ellis, 1994; Leung & Williams, 2011; van Patten et al., 2004). Accordingly, in Experiment 4B, the effects of task implementation on the probe recognition performance were tested because the goal-oriented text processing provided by a particular task should enable the participants to allocate their attention to both aspects of word forms and meanings (Horiba & Fukaya, 2006, 2012, 2015).

4.4 Experiment 4B: Task-Induced Lexical Processing and Representations

4.4.1 Design, Hypothesis, and Research Question

The first purpose of this experiment was to replicate the results of Experiment 4A. Following Horiba and Fukaya (2006, 2012, 2015), a recall task was implemented to manipulate participants' attention to a word form and its meaning. In an L1 recall task, to communicate text contents to others in their L1, the participants had to translate English inputs into Japanese while reading, thus contributing to the construction of semantic representations of the target unknown words (i.e., the meaning-focused group). The results in this condition are seen to replicate the results of Experiment 4A, i.e., that the participants would be slower in recognizing target words in the HSS learning sentences than in the LSS ones. In contrast, an L2 recall task required the participants to encode the text elements in their memory in L2 while building the mental representations of a text (i.e., the form-focused group), which may result in the retention of lexical as well as semantic representations of the target unknown words. In other words, it can be expected that the form-focused group will make faster recognition of the probes than the meaning-focused group. More importantly, the difference in the accessibility of target word forms between the HSS and LSS conditions would disappear. Thus, the hypothesis and an RQ addressed here are summarized as follows:

H4: Successful lexical inference reduces the relative accessibility of the word-form memory of unknown words.

RQ8: Does task-induced lexical processing change the relative accessibility of the word-form memory of unknown words?

4.4.2 Method

4.4.2.1 Participants

Forty Japanese EFL learners participated in the experiment (23 females and 17 males; average age = 18.9, range = 18–23). None had participated in the prior experiments. They were undergraduate or graduate students at the University of Tsukuba, majoring in humanities, social studies, psychology, international relations, engineering, biology, or medicine. Their CEFR level was the same as Experiment 4A, at B1 or B2 as based on their self-report. All participants were paid ¥1,000 for their participation after they gave written informed consent to join this study.

The participants were randomly assigned into the meaning-focused group or the form-focused group. Since reading ability has been a primary factor in successful input processing by lexical inference, the individual differences between the two groups were controlled. Participants' English reading proficiency was estimated using the same Eiken test as in the prior experiments, which did not differ between the meaning-focused group (M = 15.45, 95% CI [13.34, 17.56], SD = 4.50) and the form-focused group (M = 15.10, 95% CI [12.83, 17.37], SD = 4.85), $t(38) = 0.24$, $p = .814$, $d = 0.08$, $M_{diff} = 0.35$ (95% CI [–2.65, 3.35]). Therefore, the two groups were seen as statistically homogeneous.

4.4.2.2 Materials, Procedure, and Data Analysis

The stimuli were the same as in Experiment 4A. In the experimental session, the meaning-focused group and the form-focused group were asked to understand each story in preparation for the successive L1 and L2 recall tasks, respectively. The recall task randomly appeared four times on a computer screen due to the time constraint of the experiment and in an attempt to avoid imposing an unreasonably high cognitive load to complete the experimental tasks. Additionally, the recall task was implemented immediately after the participants read four out of 12 disguised sets, to avoid directing participants' attention toward the target words in the experimental sets. Apart from this, no other changes in procedure from Experiment 4A were made.

The same measures were taken as in Experiment 4A, and the outliers were dealt with in the same way. A two-factor mixed ANOVA was used to verify that the participants could infer the meaning of the target words more correctly in the HSS condition than in the LSS condition; the within-participants variable was Context (HSS and LSS) and the between-participants variable was Task (Meaning-focused and Form-focused). Using the same factorial design, two 2 × 2 mixed ANOVAs were performed for the mean accuracy rates and RTs of the probe recognition test to test (a) whether there were no differences in recognition accuracy among the conditions, and (b) the interaction effects of Context × Task on the relative accessibility to word-form memory.

4.4.3 Results

Analysis of the multiple-choice test. Table 4.13 shows the descriptive statistics of the multiple-choice test in Experiment 4B. Whereas a significant main effect of Context was found, $F(1, 38) = 85.17, p < .001, \eta^2 = .48$, the effect of Task was not significant, $F(1, 38) = 3.73, p = .061, \eta^2 = .03$, and also did not interact with the Context factor, $F(1, 38) = 0.44, p = .514, \eta^2 < .01$. These results showed that the participants in each group could identify the target word meanings in the HSS condition, as compared to the LSS condition.

Table 4.13

Mean Accuracy Rates With 95% CIs and SD*s of the Multiple-Choice Test*

		HSS learning sentence			LSS learning sentence		
Tasks	*n*	*M*	95% CI	*SD*	*M*	95% CI	*SD*
Meaning-focused	20	.68	[.60, .76]	.17	.31	[.22, .40]	.19
Form-focused	20	.58	[.49, .66]	.19	.25	[.17, .33]	.18

Analysis of the probe recognition test. Table 4.14 shows the descriptive statistics of the probe recognition test in Experiment 4B. No significant main effects of Context, $F(1, 38) = 1.19, p = .282, \eta^2 = .01$ or Task, $F(1, 38) = 0.29, p = .594, \eta^2 < .01$ were found. These two factors did not interact, $F(1, 38) = 0.53, p = .471, \eta^2 < .01$. Since the recognition accuracy was high enough and did not find itself distorted among the conditions, the next section will report the results of RTs.

As the between-participants variable of Task was included in a factorial design in Experiment 4B, the results can be interpreted in the following two ways: the recognition speed was affected by (a) the Context and Task factors as expected, and (b) the individual differences in reaction speed between the two task groups. In order to contradict the latter interpretation, first, I examined whether the RTs to filler trials did not differ between the task conditions. Recognition accuracy in the filler trials were high enough (*M* = .95, 95% CI [.92, .97], *SD* = .07), indicating that participants completed the probe recognition test properly. A two-tailed paired *t* test showed that no significant difference existed in RTs between the meaning-focused group (*M* = 1,078, 95% CI [962, 1,195], *SD* = 249) and the form-focused group (*M* = 1,076, 95% CI [991, 1,161], *SD* = 181), $t(38) = 0.03$, $p = .973$, $d = 0.01$, $M_{\text{diff}} = 2$ (95% CI [–137, 142]). These results ensure that the individual differences between the task conditions did not distort the results of the target trials.

Table 4.14

Mean Accuracy Rates and RTs With 95% CIs and SD*s of the Probe Recognition Test*

	Accuracy			RTs (in msec)		
Conditions	*M*	95% CI	*SD*	*M*	95% CI	*SD*
Meaning-focused (*n* = 20)						
HSS sentence	.89	[.84, .94]	.14	1,216	[1,103, 1,328]	240
LSS sentence	.93	[.89, .98]	.10	1,063	[951, 1,175]	239
Form-focused (*n* = 20)						
HSS sentence	.89	[.83, .96]	.11	928	[866, 990]	132
LSS sentence	.90	[.84, .96]	.14	901	[828, 973]	155

Figure 4.7 displays the Context × Task effects on the RTs. As Table 4.15 shows, the ANOVA found the significant main effects of Context, $F(1, 38) = 16.41$, $p < .001$, $\eta^2 = .04$, and Task, $F(1, 38) = 14.81$, $p < .001$, $\eta^2 = .24$. More importantly, a significant interaction of Context × Task was found, $F(1, 38) = 7.91$, $p = .008$, $\eta^2 = .02$. A post hoc comparison revealed that in the meaning-focused group, the RTs to the target trials were longer in the HSS condition than in the LSS condition, $F(1, 38) = 23.55$, $p < .001$, $\eta^2 = .38$. The RTs of the form-focused group did not differ between the HSS and LSS conditions, $F(1, 38)$

= 0.77, p = .386, η^2 = .01. Regardless of the Context effect, the meaning-focused group recognized the target words more slowly than the form-focused group, as seen through HSS condition: $F(1, 38) = 22.00$, $p < .001$, η^2 = .02, and LSS condition: $F(1, 38) = 6.49$, p = .015, η^2 = .01.

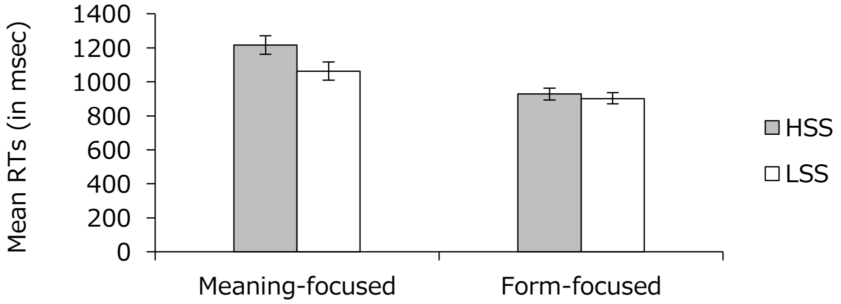


Figure 4.7. Mean RTs with ±*SEM* bars of the probe recognition test influenced by the interaction of Context × Task (*n* = 20 in each task condition).

Table 4.15

ANOVA Results for Main Effects and Interaction Effects of Task and Context on the Probe Recognition Speed

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
Between participants						
Task	1	1013223.88	1013223.88	14.81	< .001	.24
Error (Task)	38	2600365.06	68430.66			
Within participants						
Context	1	162414.56	162414.56	16.41	.001	.04
Task × Context	1	78262.35	78262.35	7.91	.008	.02
Error (Context)	38	376049.92	9896.05			

4.4.4 Discussion

Experiment 4B produced results consistent with the study hypothesis. The pattern of the probe recognition performance of the meaning-focused group was similar to Experiment 4A. Since the L1 recall task tapped the participants for constructing a message-based memory of both the stories and the unknown words, the result implies that successful lexical inference decreased the relative accessibility of the memory of the word forms. Importantly, the form-focused group were able more quickly to respond to the words that had appeared in the discourse as compared to the meaning-focused group, even though their inference performances (the multiple-choice test scores) were comparable with the meaning-focused group. Moreover, their recognition speeds did not differ between successful and unsuccessful lexical inference. Together, these results indicate that the task for allocating the learners' attention to L2 lexical properties improved the robust representation of both semantic and lexical features of the unknown words.

Despite the methodological differences in assessing the knowledge representation established from reading, the finding is consistent with Horiba and Fukaya (2012, 2015) that showed the superiority of the form-focused task in the recall of L1 meaning from L2 word form. Since the L2-to-L1 recall test of vocabulary meaning requires L2 learners to connect the lexical knowledge between form and meaning (e.g., Nation, 2013; Schmitt, 2010), the results of their research can be explained from the perspective of the increase in the accessibility of the lexical representation. Experiment 4B showed that this relative change of accessibility was elicited by task demands. As predicted in accounts of task-based language learning (e.g., Robinson, 2011), the different demands from the meaning-focused and the form-focused tasks differently focused the learners' attention on meaning and forms, respectively. Several findings in the literature suggest that if only the meaning-focused tasks are used, aspects of vocabulary knowledge acquired will be limited (e.g., Barcroft, 2009; de la Fuente, 2006; Nation, 2013), particularly in the construction of form-meaning connections (e.g., Ellis, 1994; Nation, 2013). In contrast, they found the positive effects of the form-focused tasks on the consolidation of the connections. In the context of incidental vocabulary learning from reading, although comprehension does not necessarily lead to vocabulary growth (e.g., Mondria & Wit-de Boer, 1991), the findings provide the advantage of focus-on-form tasks in terms of the multilevel representation theory (e.g., Fletcher & Chrysler, 1990; van Dijk & Kintsch, 1983).

4.5 Summary of Study 2

Study 2 conducted four experiments to examine the establishment of semantic and lexical representations of unknown words generated by lexical inference. First, both RTs and descriptive data indicate that some semantic forms of word memory were built by Japanese EFL readers, but that the specificity of the word meanings inferred in the time course of L2 reading varied according to the word-context semantic similarity computed by LSA (RQ5, H2, and H3). Besides, the lexical inference specificity interacted with the learners' L2 reading proficiency (RQ6). Second, the findings of Experiments 4A and 4B support the assumption that task-induced lexical processing promoted incidental L2 word learning in terms of the construction of form-meaning connections (RQ7, RQ8, and H4). Attention allocation to word forms by tasks resulted in establishing more sophisticated lexical representations without inhibiting the ability to derive new word meanings as exhibited in the probe recognition and multiple-choice tests.

The experiments reported so far have not yet directly addressed the issue on the knowledge gains about how novel word meanings are used in any context. Moreover, rather than compiling the laboratory experiments examining the usage-based model of L2 vocabulary learning, Study 3, which was classroom research, would have informative pedagogical implications by replicating the prior results in the settings of Japanese EFL classrooms. In particular, the next experiments were designed to investigate the effects of contextualized vocabulary learning on the knowledge gains in word meanings and usage with a VKS test. As one of the various goals of vocabulary acquisition, the knowledge obtained through learning should be stable and robust enough to be overt in both receptive and productive situations (Nation, 2013; Webb, 2005). Although the plausibility judgment test used in Experiment 1 tapped the participants for adapting word usage knowledge derived from contextual information (Borovsky et al., 2010), this was not found to be a productive situation.

Considering the issues involved in the difficulty of testability in examining covert word knowledge incidentally gained from reading, Study 3 started with the investigation of usage-based context effects on deliberate vocabulary learning. As the previous experiments demonstrated the influences of word-context semantic similarity on lexical inference generation and encoding, Experiment 5A was designed to test the effects on learning. Furthermore, the vocabulary learning strategies used by individual learners were considered as the influential factors because the outcomes of contextualized vocabulary learning will depend on whether the participants focus their attention on a given context (see Section 2.4). For example, the word-analysis strategy will reduce the benefits from context (e.g., Barcroft, 2002, 2009), whereas it is possible that the good use of context promotes gains in semantic and syntactic knowledge.

In Experiment 5B, participants engaged in incidental L2 vocabulary learning, in which the task demands on the use of context are seen to increase. As learners do not process input with the conscious intention of learning it incidentally (Hulstijn, 2005), it is possible that the lack of attention to input reduces the knowledge gains that should be improved by HSS learning sentences. On the other hand, another possibility is that HSS learning sentences improve incidental L2 word learning because, as mentioned above, this learning mode requires involving contextual information in lexical processing, more than the intentional learning mode does (e.g., McKeown, 1985; Webb, 2007a; Wesche &

Paribakht, 2010). Namely, the context effects on word learning might increase more in the incidental mode than in the intentional one. Furthermore, the assumption of the language acquisition model suggested by LSA is essentially an incidental mode (e.g., Bolger et al., 2008). To elucidate these contrastive views, it is important to determine if LSA predicts the performance of incidental L2 vocabulary learning.

Chapter 5

Study 3: Incidental Learning of L2 Word Meanings and Usage

5.1 Experiment 5A: Using LSA to Predict Contextualized Vocabulary Learning

5.1.1 Design and Research Questions

Final experiments were classroom-based studies, in which I aimed at replicating the findings of the previous laboratory experiments. In Experiment 5A, to test whether LSA similarity has influences on contextualized L2 vocabulary learning, its interaction with vocabulary learning strategies were included in a factorial design. The analysis focused on knowledge gains in L2 word meanings and usage by comparing the effects of semantic similarities among the types of vocabulary learning strategies. Thus, the experiment was designed to address the following three RQs:

RQ9: Do the semantic similarities between target words and learning sentences promote gains in knowledge of word meanings?

RQ10: Do the semantic similarities between target words and learning sentences promote gains in knowledge of word usage?

RQ11: Do the effects of word-context semantic similarities on contextualized vocabulary learning interact with vocabulary learning strategies?

5.1.2 Method

5.1.2.1 Participants

One hundred and ten Japanese EFL learners participated in this classroom-based session for credit in a regular English class. Five students' data were removed from the analyses because they were absent from some sessions (42 females and 63 males; average age = 18.1, range 18–19). They were first-year undergraduates at Teikyo University of

Science, majoring in physiotherapy or nursing. At the time of the experiment, all had studied English for at least six years in Japan. They gave informed consent about how the personal data collected would be used.

A reading subsection (k = 50) of a TOEIC Bridge® practice test (Educational Testing Service, 2007) and the 1,000 to 5,000 word level of version 3 of the Mochizuki vocabulary size test (Aizawa & Mochizuki, 2010) were used to estimate their English proficiency. Given that the results of the TOEIC Bridge® (M = 21.67, 95% CI [20.22, 23.12], SD = 7.43, Cronbach's α = .84) and vocabulary size test (M = 2,693, 95% CI [2,548, 2,839], SD = 744, Cronbach's α = .95), it was assumed that they were beginner-level EFL learners (A1 to A2 level in CEFR).

Using a vocabulary learning strategies questionnaire, the participants were grouped with a cluster analysis; it applied the Ward method with squired Euclidean distance technique in the same way as Mizumoto and Takeuchi (2009). Prior to the cluster analysis, the raw scores were synthesized into four factor scores as the questionnaire items were composed of four factors of vocabulary learning strategies, that is, association, imagery, structure, and context (Mizumoto, 2006). Based on an outputted dendrogram, a cut-off point was set at 5.00, and the participants could be divided into three groups. This decision ensured that each group was significantly different in their vocabulary learning strategies, as illustrated in Figure 5.1 and Table 5.1.

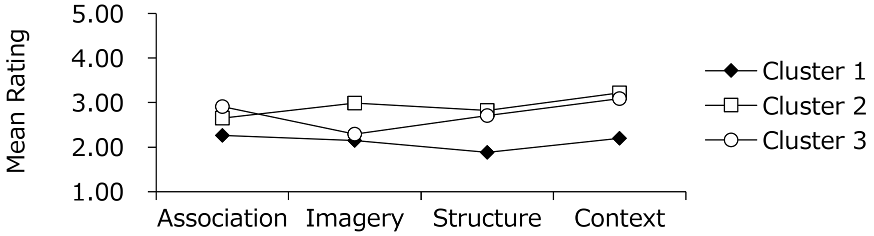


Figure 5.1. Cluster profiles of vocabulary learning strategies among the three strategy groups (Cluster 1, n = 34; Cluster 2, n = 35; Cluster 3, n = 36).

Table 5.1

Description of Each Cluster

Strategy	Items	Cluster 1 (*n* = 34)	Cluster 2 (*n* = 35)	Cluster 3 (*n* = 36)	Tukey HSD Comparison
Association	1-4	2.26 (0.89)	2.65 (0.78)	2.91 (0.97)	C3 > C1
Imagery	5-8	2.15 (1.03)	2.99 (0.75)	2.29 (0.57)	C2 > C1, C3
Structure	9-11	1.88 (1.01)	2.82 (0.83)	2.71 (0.97)	C2, C3 > C1
Context	12-14	2.20 (0.91)	3.22 (0.74)	3.09 (0.95)	C2, C3 > C1

Note. *SD*s are in parentheses. Each difference in the multiple comparisons was significant at the level of .05. All the questionnaire items are presented in Appendix 10.

One-way ANOVAs performed on the mean scores of four strategies showed the significant differences in the strategy use among three clusters: association, $F(2, 102) = 4.28$, $p = .017$, $\eta^2 = .08$; imagery, $F(2, 102) = 10.61$, $p < .001$, $\eta^2 = .18$; structure, $F(2, 102) = 10.23$, $p < .001$, $\eta^2 = .18$; context, $F(2, 102) = 14.08$, $p < .001$, $\eta^2 = .23$. As shown in Table 5.1, post hoc comparisons with Tukey HSD indicated the different patterns of the participants' strategy use as follows:

- The participants categorized as Cluster 1 could be considered the less frequent strategy users because their overall use of strategies was lower than the other groups. Particularly, the learners in Cluster 1 least relied on using contexts in memorizing new words among the three strategy groups.
- Whereas the learners in Cluster 2 made frequent use of word imagery, those in Cluster 3 were likely to report the use of the association strategy; however, there was no significant difference in the use of contexts between Clusters 2 and 3.
- It should be noted that the participants did not rely on a certain strategy as a whole

because the mean scores were around 3.00 or below, indicating that the participants could not say for sure which strategies they had used.

Nevertheless, the results of the questionnaire suggested that the three groups differed according to their strategy use in vocabulary learning.

5.1.2.2 Materials

In the same manner as in the prior experiments, Experiment 5A prepared two types of learning sentences including each target word: HSS vs. LSS contexts provided by LSA. Table 5.2 summarizes the characteristics of the contexts and target words.

Table 5.2

Linguistic and Psycholinguistic Characteristics of Learning Sentences and Target Words

	HSS (k = 10)		LSS (k = 10)				
Characteristics	*M*	*SD*	*M*	*SD*	*t*(18)	*p*	*d*
Learning sentences							
LSA	.45	.12	.14	.07	7.00	< .001	3.16
Words	8.90	2.28	9.80	2.78	0.79	.439	0.36
Target words							
Letters	4.70	1.16	5.90	1.83	1.87	.078	0.78
Syllables	1.30	0.48	1.80	0.79	1.71	.105	0.77
Frequency	244.30	256.16	219.90	221.95	0.23	.822	0.10
Familiarity	5.42	0.55	5.38	0.71	0.14	.887	0.06

Note. The value of frequency means the number of times where the target words appear in the NTT database corpus (N = 341,771). The range of word familiarity is 1.00 (*very low*) to 7.00 (*very high*).

Target words. A total of 20 target words were selected from Experiment 3A and Webb (2007a), which are all provided in Appendix 8. Half of them were embedded into HSS learning sentences (six nouns and four verbs), and the others into LSS learning sentences (seven nouns and three verbs). Based on Nation and Webb (2011), I controlled as many factors of the target words critical for learning their meaning and use as possible between the HSS and LSS conditions (i.e., word length such as letters and syllables, word frequency, and word familiarity, reported in Table 5.2). In rating word frequency and familiarity, Japanese word properties (Amano & Kondo, 1999, 2000) were referred because word learning depends on whether the meaning of unknown words is well lexicalized in learners' L1 (Wesche & Paribakht, 2010). All target nouns except *headline* were concrete, but its concept was assumed to be familiar to the participants. The target verbs had no novel or special meanings.

Learning sentences. The learning sentences corresponding to the target words were also adapted from Experiment 3A and Webb (2007a). Because the participants' average vocabulary size was 2,754 words, each sentence was modified to consist of high-frequency basewords (levels 1 to 2) based on JACET (2003). Whereas prior studies used discourse to examine incidental word learning (e.g., Hulstijn et al., 1996), Experiment 5A adopted the sentence format because of the participants' low English reading proficiency. Sentential information is a primary source of lexical inference, and the use of discourse knowledge requires higher L2 proficiency (Ushiro et al., 2013; Wesche & Paribakht, 2010). Thus, the sentence was used in this experiment to ensure that the participants could make better use of contextual information. Sentence length did not differ between HSS and LSS contexts (see Table 5.2).

The strength of semantic similarities between target words and context meanings was computed by LSA, based on the semantic space of "General reading up to first year college." A total of 10 contexts were determined to have higher LSA similarities based on the median split. The HSS learning sentences had substantially higher LSA similarities than the LSS learning sentences (see Table 5.2); the outcomes of contextualized learning were affected by the LSA similarities if any significant differences were found between the HSS and LSS conditions.

Vocabulary Knowledge Scale test. The VKS test was employed to evaluate the participants' gains in word knowledge (see Appendix 9). The VKS has been designed to measure knowledge of words, ranging from total unfamiliarity with a target word to the ability to use a target word with semantic and syntactic accuracy in a sentence (e.g., Wesche & Paribakht, 1996, 2010). Because the purpose of this study was to find how the LSA values between words and learning sentences affected the gains in each aspect of word knowledge (i.e., knowledge of semantics and usage), the VKS was needed to isolate each type of knowledge in the assessment.

Vocabulary learning strategies questionnaire. As described in Section 5.1.2.1, a questionnaire was used to arrange each participant into the identical vocabulary learning strategy groups. This questionnaire was originally developed by Gu and Johnson (1996) to measure learners' vocabulary learning behaviors while they are engaged in memorizing new vocabulary. The current experiment used 14 items for evaluating their word encoding strategies (see Appendix 10). These items could function as a psychometrically valid scale, and they were used to subcategorize the learners' encoding strategies into association, imagery, structure, and context (Mizumoto, 2006). The questionnaire used a 5-point Likert scale ranging from 1 (*not at all true to me*) to 5 (*very true to me*). Each item description was translated into Japanese by the author, so that the participants could take the questionnaire in their native language.

5.1.2.3 Procedure

The experiment was conducted in two sessions within 90-minute class periods;

each session was composed of a learning trial and a recall trial. This experiment used the 20 target words but they were presented in tens in consideration of a learning burden; thus, a set of learning and recall trial was repeated twice.

In the learning trial, first, the participants were given 10 target words with the pre VKS test in order to familiarize them with how to work on the test by replying to whether or not they previously had any knowledge of each target word. This procedure was also intended to exclude the target words that the participants reported they already knew. Next, they were given a target word list (see Figure 5.2) and asked to learn as many target words as possible, which took seven minutes. At that time, the participants were not provided any instructions about how to memorize the words. After the learning trial, 60 minutes were for a regular English class, which was completely unrelated to the experimental session, in order to avoid the effects of short-term memory on word learning. Then, in the recall trial, the participants received another sheet for the post VKS test and were required to recall the target words they had memorized. They were given as much time as they needed to complete the test.

覚える単語	意味	例文
(1) mako	「アオザメ」	The surfers were attacked by a dangerous **mako** in the sea.
(2) abhor	「ひどく嫌い」	We really **abhor** his English class.
(3) cattle	「家畜用の牛」	The farmer milked the **cattle**.

Figure 5.2. Part of a word list presented in the vocabulary memorization activity. A fuller format and instruction are presented in Appendix 11.

One week later, the second learning and recall trial was conducted in the same manner as the first session. The presentation order of the target words was counterbalanced using the Latin square method (two sessions × two word lists). The participants had taken three profiling tests (i.e., TOEIC Bridge®, Mochizuki vocabulary size test, and vocabulary learning strategies questionnaire) in former class periods.

5.1.2.4 Scoring and Data Analysis

Prior to rating the VKS test, every target word that any participant reported to have known was excluded from the analysis. Particularly, as almost all the participants knew the target word *vehicle*, it was removed from the whole analyses. The scoring was made in accordance with Wesche and Paribakht (1996). When the participants reported that they knew a target word meaning, one credit was given if they properly produced its definition (i.e., VKS3). When the participants wrote any sentences using a target word, I determined if its usage was appropriate (i.e., VKS4). Because the participants' English proficiency was at the beginner level, a credit was given when the target word usage was appropriate even though the whole sentence was not grammatical (e.g., "This puppy is *a* very cute"). Thus, the criterion was whether the target word was used correctly in terms of its part of speech, inflection, semantics, and syntax. Any spelling mistakes were also disregarded. Finally, when a participant wrote a meaning or sentence with a target word but could not complete it accurately, it was treated as if they gained no knowledge of that word's meaning or usage (i.e., VKS2). Cronbach's α was .81 for assessing the gains in word meaning and .92 for word usage.

To examine the knowledge gains in word meaning and usage, VKS scores were interpreted as follows: (a) VKS1-2, in which the participants could not recall the meaning of a target word or use it, was regarded as no gains in word meaning or usage knowledge; (b) VKS3-4 means that they succeeded in integrating the semantic knowledge of a target word into their mental lexicon; therefore, their mean rate was regarded as meaning knowledge gain; and similarly, (c) the mean rate of VKS4 was seen as usage knowledge gain. Two separated two-factor mixed ANOVAs were performed for the mean score rates;

Context (HSS and LSS) was a within-participant variable and Strategy (Clusters 1, 2, and 3) was a between-participants variable.

5.1.3 Results

5.1.3.1 Knowledge Gains in Word Meanings

Table 5.3 shows the descriptive statistics of the score rates for the post VKS test in terms of word meaning knowledge. The 2 × 3 mixed ANOVA showed a significant main effect of Context, $F(1, 102) = 53.81$, $p < .001$, $\eta^2 = .07$, but not Strategy, $F(2, 102) = 0.42$, $p = .660$, $\eta^2 = .01$. Importantly, a significant interaction of Context × Strategy was found, $F(2, 102) = 3.67$, $p = .029$, $\eta^2 = .01$. Table 5.4 presents the overall results of the ANOVA.

Table 5.3

Mean Rates With 95% CIs and SD*s of the VKS Scores of Word Meaning*

		HSS			LSS			Mean difference	
Strategy	*n*	*M*	95% CI	*SD*	*M*	95% CI	*SD*	HSS–LSS	95% CI
Cluster 1	34	.57	[.51, .64]	.19	.38	[.32, .45]	.18	.19	[.13, .25]
Cluster 2	35	.56	[.49, .64]	.23	.47	[.38, .55]	.25	.10	[.03, .16]
Cluster 3	36	.57	[.48, .66]	.27	.48	[.39, .57]	.27	.09	[.03, .14]
Total	105	.57	[.52, .61]	.23	.44	[.40, .49]	.24	.12	[.09, .16]

Figure 5.3 graphically displays the interaction between Context and Strategy on the gains in knowledge of word meanings. First, a significant simple main effect of Context was found in all three clusters (all *p*s < .01), indicating that the HSS condition promoted the learning of word meaning knowledge better than the LSS condition did, regardless of the types of vocabulary learning strategies.

Table 5.4

ANOVA Results for Main Effects and Interaction Effects of Strategy and Context on Word Meaning Learning

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
Between participants						
Strategy	2	0.08	0.04	0.42	.660	.01
Error (Strategy)	102	9.71	0.10			
Within participants						
Context	1	0.83	0.83	53.81	< .001	.07
Strategy × Context	2	0.11	0.06	3.67	.029	.01
Error (Context)	102	1.57	0.02			

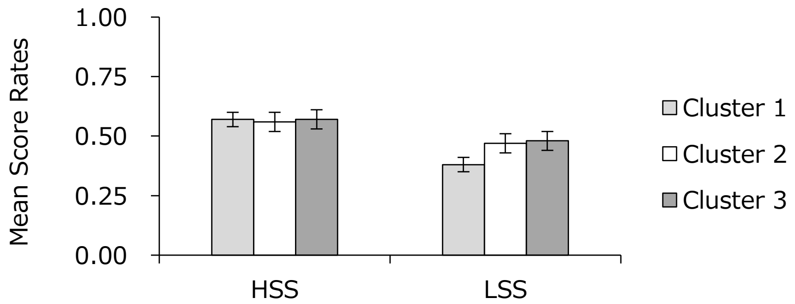


Figure 5.3. Mean score rates with ±*SEM* bars of the gains in word meaning knowledge among three strategy groups (Cluster 1, n = 34; Cluster 2, n = 35; Cluster 3, n = 36).

5.1.3.2 Knowledge Gains in Word Usage

Table 5.5 presents the descriptive statistics of the post VKS scores in terms of word usage knowledge. The 2 × 3 mixed ANOVA showed that there was a significant main effect of Context, $F(1, 102) = 17.00$, $p < .001$, $\eta^2 = .01$, but not Strategy, $F(2, 102) = 1.28$, $p = .282$, $\eta^2 = .02$. The interaction between the two factors was not significant, $F(2, 102) = 1.12$, $p = .329$, $\eta^2 < .01$. Table 5.6 summarizes the results of the ANOVA.

Table 5.5

Mean Rates With 95% CIs and SD*s of the VKS Scores of Word Usage*

		HSS			LSS			Mean difference	
Strategy	*n*	*M*	95% CI	*SD*	*M*	95% CI	*SD*	HSS–LSS	95% CI
Cluster 1	34	.09	[.04, .14]	.14	.06	[.02, .10]	.12	.03	[–.01, .06]
Cluster 2	35	.17	[.08, .25]	.24	.12	[.05, .20]	.22	.04	[.00, .09]
Cluster 3	36	.17	[.08, .25]	.25	.10	[.03, .17]	.19	.07	[.03, .10]
Total	105	.14	[.10, .18]	.22	.10	[.06, .13]	.18	.05	[.02, .07]

Table 5.6

ANOVA Results for Main Effects and Interaction Effects of Strategy and Context on Word Usage Learning

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
			Between participants			
Strategy	2	0.19	0.10	1.28	.282	.02
Error (Strategy)	102	7.59	0.07			
			Within participants			
Context	1	0.11	0.11	16.99	< .001	.01
Strategy × Context	2	0.02	0.01	1.12	.329	< .01
Error (Context)	102	0.63	0.01			

Figure 5.4 illustrates the Context effects on the learning of word usage in contexts. The statistical analysis yielded the following findings: (a) The HSS condition helped the participants obtain knowledge of word usage better than the LSS condition did, and (b) the effects of semantic similarities on learning word usage did not differ according to the learners' vocabulary learning strategies.

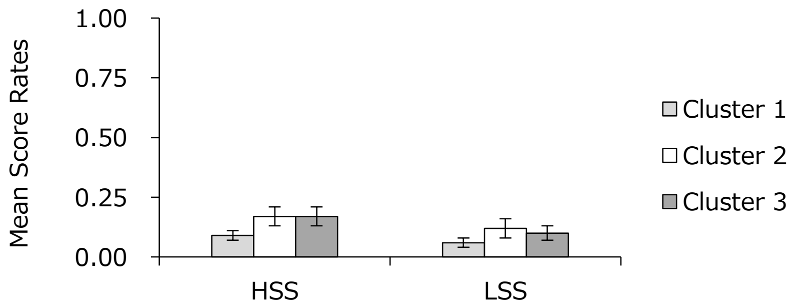


Figure 5.4. Mean score rates with ±*SEM* bars of the gains in word usage knowledge among three strategy groups (Cluster 1, *n* = 34; Cluster 2, *n* = 35; Cluster 3, *n* = 36).

5.1.4 Discussion

Regarding RQ9, Experiment 5A showed that the knowledge of word meanings was acquired when the words were naturally woven into context. This finding supported the facilitation effects of contexts on the gains in knowledge of word meanings (e.g., Bolger et al., 2008; Hasegawa, 2013; Webb, 2008), contrasting with the null results of some past studies (Griffin, 1992; Laufer & Shmueli, 1997; Prince, 1996; Webb, 2007a). Because this experiment expanded on earlier investigations by controlling context quality through the application of LSA, it provided evidence as to why Webb (2007a) could not reveal any differences between contextualized learning and paired-associate learning. According to LSA, of the 12 learning sentences used in his study, six contexts had higher semantic similarities, but the others had lower semantic similarities (see Appendix 8). Thus, the significant difference of contextual quality caused an invalidated result about the effects of context on vocabulary learning.

As predicted by the basic principles of LSA (e.g., Landauer et al., 1998), the high semantic similarities between learning items and corresponding contexts promoted word meaning learning. Similar results have been reported in other studies. For example, Bolger et al. (2008) found that the degree of semantic constraints for individual contexts plays a substantial role in learning various types of lexical knowledge. Hasegawa (2013) also demonstrated that mental imagery between words and contexts is conductive to learning new word meanings. Although it is beyond the scope of this experiment to address the distinction between these two perspectives and semantic similarities, it can be concluded that LSA is available to predict the outcomes of contextualized vocabulary learning in terms of word meanings.

The Context × Strategy interaction showed that the negative effects of LSS contexts were different according to the participants' vocabulary learning strategies (a partial answer to RQ11). Although the participants in Cluster 1 were categorized as the less frequent users of context, their learning outcomes of word meaning knowledge did not differ compared to the other groups when the HSS learning sentences were presented. In contrast, the LSS learning sentences reduced the context effects on those participants' word meaning learning more than on the learners' in Clusters 2 and 3. In addition to the result that the LSS learning sentences were consistently less effective in word meaning learning, this indicates that such hindrance effect differs in the types of strategy use.

In the learning trial, the participants were not explicitly instructed to associate the target words with the given sentences. Nevertheless, as Clusters 2 and 3 were the strategy users of context relatively, it appears that the participants attempted to associate the target words with the sentences even though they had lower semantic similarities. On the other hand, the findings demonstrate that the Cluster 1 participants failed to link the target words to each LSS learning sentence because they were usually the poorer strategy users of context. These findings supported evidence that the learning outcomes differ according to the strategy types used for learning words (e.g., Gu & Johnson, 1996; Hamada, 2011; Mizumoto & Takeuchi, 2009). Specifically, Experiment 5A expands previous studies by demonstrating that the learners who make less use of contextual information suffer from futile learning sentences in terms of LSA similarities.

In relation to RQ10, considering that the participants in Experiment 5A were at a beginner-level in English language learning, it is reasonable that the mean rates of the gains in knowledge of word usage were relatively low and the data distribution was larger. Nevertheless, the results demonstrated a clear trend that the LSA similarities affected the learning of word usage from contexts. As in the gains in word meaning knowledge, it indicates that the word-context semantic similarities contribute to the acquisition of word usage knowledge. However, it should be noted that the mean difference was relatively small (M_{diff} = .05, 95% CI [.02, .07]), and the semantic similarity effect was smaller on word usage knowledge (η^2 = .01) than on word meanings (η^2 = .07). In other words, LSA similarity could explain the learning outcomes of word usage from contexts, but its effectiveness was small.

Given that the data of VKS4 were not robust among three strategy groups, any analysis focusing on strategy use in contextualized learning of word usage must be relatively tentative. Specifically, the large *SD*s and the range of 95% CIs of means suggest a floor effect. In fact, the result showed no significant differences of the learning outcomes among Clusters 1 to 3, although a difference seemed to appear between Cluster 1 and Clusters 2–3, as reported in Figure 5.4. The LSA similarity reflected the learning of word usage, but the insignificant interaction with vocabulary learning strategies differed from the results of learning word meanings.

These findings can be explainable as follows: Contextual diversity was required even when the HSS condition possessed rich contextual information; and the beginner-level participants could not abstract the lexical knowledge of morphology, syntax, and usage from a single exposure to a word. For example, Bolger et al. (2008) demonstrated that a single context with a word definition is not superior to context variations. As many researchers have suggested (e.g., Bolger et al., 2008; Fukkink et al., 2001; Grabe, 2009; Jiang, 2000; Nation, 2013; Nation & Webb, 2011; Webb, 2008), partial knowledge of a word develops into full knowledge incrementally through multiple exposures to each word. Despite the word meaning knowledge as vocabulary breadth, a single exposure to target items did not lead to the large gains in word usage knowledge as vocabulary depth, even though the contexts provided typical usage of words. Therefore, it is not surprising that the participants did not obtain the knowledge of word usage compared with that of word meanings. Nevertheless, this explanation does not affect in any way the central claim that the HSS learning sentences contribute to gains in knowledge of word usage. As incremental vocabulary learning advances slowly (e.g., Nation, 2013), many studies have suggested that context effects will gradually increase with the number of word encounters (e.g., Bolger et al., 2008; Grabe, 2009; Jiang, 2000; Nation, 2013; Rott, 1999, 2007; Swanborn & de Glopper, 1999; Webb, 2007b, 2008).

In this experiment, the effectiveness of contextualized vocabulary learning was not directly compared with that of paired-associate learning. Nevertheless, a set of findings suggest that the inefficiency of contextualized vocabulary learning (e.g., Griffith, 1992; Prince, 1996; Webb, 2007a) may be a bold claim. Because evidence from past studies has proposed that context has both effective and ineffective aspects (Grabe, 2009; Huckin & Bloch, 1993; Laufer & Shmueli, 1997; Nation, 2013; Webb, 2008), the application of LSA should be required to manipulate the context effects from the viewpoint of the semantic similarities.

5.2 Experiment 5B: Using LSA to Predict Incidental Vocabulary Learning

5.2.1 Design, Hypothesis, and Research Questions

In Experiment 5B, I built a research design to compare the effects of semantic similarity (HSS and LSS) on incidental knowledge gains in word meaning and usage when learners engaged in lexical inference, MCG, and dictionary use tasks. It also aimed at examining whether the performance of deriving word information from context differs according to LSA similarities because success in lexical inference is the prerequisite of incidental word learning. To assess the acquisition of two types of word knowledge, this experiment used an unannounced VKS test, following the definition of incidental learning as the learning mode “in which participants are not forewarned of an upcoming retention test for a particular type of information” (Hulstijn, 2005, p. 132). A hypothesis and two RQs were addressed here:

H5: Does the higher semantic similarity between a target word and context computed by LSA improve the learners’ deriving unknown word meaning from context?

RQ12: Does the higher semantic similarity between a target word and context computed by LSA improve the incidental learning of word meaning?

RQ13: Does the higher semantic similarity between a target word and context computed by LSA improve the incidental learning of word usage?

5.2.2 Method

5.2.2.1 Participants

One hundred and fifty-three Japanese EFL learners participated in the classroom-based session for credit in a regular English class. They were all but one first-year undergraduates (104 females and 49 males; average age = 18.3, range = 18–20), majoring in child education, life sciences, physiotherapy, or nursing. They were not the participants in Experiment 5A. At the time of the experiment, all had studied English for at least six years in Japan. The same tests as in Experiment 5A were used to estimate their English reading proficiency. Given the results of the TOEIC Bridge® (*M* = 28.58, 95% CI [27.54, 29.61], *SD* = 6.45, Cronbach’s α = .84) and vocabulary size test (*M* = 2,754, 95% CI [2,654, 2,867], *SD* = 665, Cronbach’s α = .95), it was assumed that they were beginner-level EFL learners (A1 to A2 level in CEFR). Participants gave informed consent about how the personal data collected would be used.

5.2.2.2 Materials

The same target words and learning sentences were used as in Experiment 5A. For the MCG task, the gloss options appeared in the margin of each sentence. Participants had three choices: the definition of the target words, the meaning of words phonologically similar to the target words, and the meaning of contextually unfit words. For example, in "The farmer milked the cattle," two distractors of "cattle" were *yakan* [kettle] and *bokusou* [grass], respectively. Note that the results obtained from this task were not directly relevant to the current research goals. Both MCG and dictionary use tasks were conducted to inform the participants of the correct meaning of all the target words because this study intended the LSS learning sentences to cause disadvantage in terms of lexical processing in the inference task. If only the inference task was used, different outcomes in incidental word learning between the HSS and LSS conditions could be simply viewed as the degree of success in deriving target word meanings, but not as effects of word-context semantic similarity.

5.2.2.3 Procedure

The experiment was conducted in two sessions within a 90-minute class period: (a) the incidental-learning phase (meaning identification tasks including lexical inference, MCG, and dictionary use) and (b) the testing phase.

In the lexical inference task, the 20 learning sentences were presented in a booklet, with the target words underlined. The participants were asked to derive their meanings from the contexts and wrote down their answers in Japanese (20 minutes). At that time, they circled any target words that they had already known prior to this task. The presentation order of the target words differed among participants using random function. In the next task, MCG, they were given the same learning sentences and target words with three gloss options, and asked to read the contexts again and select the best-fit meaning of each target word (15 minutes). After that, they consulted a dictionary to verify whether the meanings of the target words they selected were correct. Finally, the author confirmed that the participants appropriately completed these meaning identification tasks, and retrieved the worksheets (10 minutes). It should be noted that the participants' intention in these tasks was to practice lexical processing strategies, not to learn the words intentionally. Appendix 12 presents the details of task directions.

The participants received another worksheet for the VKS test, in which the target words were presented in an alphabetical order, and they were instructed to recall the target words. They were given as much time as they needed to complete the test (25 minutes). During the incidental-learning phase, the author did not warn them of the VKS test in advance. The participants had taken two profiling tests (i.e., TOEIC Bridge® and Mochizuki vocabulary size tests) in former class periods.

5.2.2.4 Scoring and Data Analysis

Prior to assessing the task performances, every target word that any participant reported to have known was excluded from the analysis. Strict scoring was employed to evaluate the performance of the inference and MCG tasks; one credit was given when the inferred and selected meaning was correct. The internal consistency of each task was enough (Cronbach's $\alpha = .85$ for the inference; .76 for the MCG). A two-way ANOVA was performed for the mean score rate; it included Context (HSS and LSS) and Task (Inference and MCG) as within-participants variables.

The VKS test was scored in the same manner as in Experiment 5A. Cronbach's α was .82 for assessing the gains in word meaning and .85 for word usage. A two-way ANOVA was performed for the mean score rate; Context (HSS and LSS) and Knowledge (Meaning and Usage) were within-participants variables.

5.2.3 Results

5.2.3.1 Meaning Identification Tasks

Table 5.7 shows the descriptive statistics of learner performances for the inference and MCG tasks. As Table 5.8 displays, the most important result of the two-way ANOVA was a significant interaction of Context × Task, $F(1, 152) = 156.03$, $p < .001$, $\eta^2 = .05$. Therefore, a subsequent analysis was implemented to examine the Context effects on each task performance (see Figure 5.5).

Table 5.7

Mean Rates With 95% CIs and SDs *of the Inference and MCG Task Scores (*N *= 153)*

	HSS			LSS			Mean difference	
Task	*M*	95% CI	*SD*	*M*	95% CI	*SD*	HSS–LSS	95% CI
Inference	.49	[.45, .54]	.27	.16	[.13, .19]	.19	.33	[.29, .37]
MCG	.82	[.79, .85]	.19	.79	[.76, .82]	.20	.03	[–.02, .06]

Table 5.8

ANOVA Results for Main Effects and Interaction Effects of Context and Task on Meaning Identification

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
Context	1	483.56	483.56	147.02	< .001	.07
Error (Context)	152	499.94	3.29			
Task	1	3502.12	3502.12	943.19	< .001	.49
Error (Task)	152	564.38	3.71			
Context × Task	1	354.83	354.83	156.03	< .001	.05
Error (Context × Task)	152	345.67	2.27			

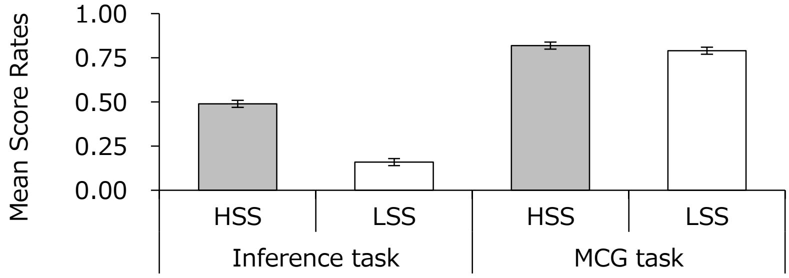


Figure 5.5. Mean score rates with ±*SEM* bars of the inference and MCG tasks' scores between HSS and LSS conditions (N = 153).

Lexical inference task. The Context factor substantially affected the inference performance, $F(1, 152) = 232.85$, $p < .001$, $\eta^2 = .61$. It showed that the participants could more accurately infer the meaning of the target words from the HSS condition than the LSS condition, and the mean difference was large ($M_{diff} = .33$, 95% CI [.29, .37]).

Multiple-choice glosses task. In the MCG task, the mean score rates substantially increased, compared to the inference task, by .33 (95% CI [.29, .37]) in the HSS condition, $F(1, 152) = 267.27$, $p < .001$, $\eta^2 = .64$, and by .63 (95% CI [.59, .67]) in the LSS condition, $F(1, 152) = 1034.18$, $p < .001$, $\eta^2 = .87$. Importantly, the mean score rates did not differ between the HSS and LSS conditions, $F(1, 152) = 1.85$, $p = .176$, $\eta^2 = .01$, $M_{diff} = .03$ (95% CI [–.02, .06]).

5.2.3.2 Incidental Word Learning

Table 5.9 provides the descriptive statistics for the VKS test. Our interest was in the effect of word-context similarity on the incidental gains in word meaning and usage knowledge, and the Context effect was large, $F(1, 152) = 316.08$, $p < .001$, $\eta^2 = .17$. Moreover, the knowledge gains depended on its type, $F(1, 152) = 190.93$, $p < .001$, $\eta^2 = .18$, and the two-factor interaction was also significant, $F(1, 152) = 9.50$, $p = .002$, η^2

< .01. To answer RQs 12 and 13, this study reported Context effects on incidental learning of word meaning and usage respectively. Table 5.10 shows the overall results.

Incidental gains in word meaning. Figure 5.6 displays the Context effects on the incidental gains in word meaning knowledge. A post hoc analysis showed that word meaning learning was promoted better in the HSS condition than the LSS one, $F(1, 152) = 232.85$, $p < .001$, $\eta^2 = .61$, and this Context effect was large, $M_{\text{diff}} = .33$ (95% CI [.29, .37]).

Table 5.9

Mean Rates With 95% CIs and SD*s of the VKS Score Rates (*N *= 153)*

	HSS			LSS			Mean difference	
Knowledge	*M*	95% CI	*SD*	*M*	95% CI	*SD*	HSS–LSS	95% CI
Meaning	.83	[.80, .86]	.18	.50	[.45, .56]	.33	.33	[.29, .37]
Usage	.50	[.44, .55]	.34	.24	[.19, .28]	.28	.26	[.22, .30]

Table 5.10

ANOVA Results for Main Effects and Interaction Effects of Context and Knowledge on Incidental Vocabulary Learning

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
Context (C)	1	13.24	13.24	316.08	< .001	.17
Error (C)	152	6.37	0.04			
Knowledge (K)	1	13.83	13.83	190.93	< .001	.18
Error (K)	152	11.01	0.07			
C × K	1	0.19	0.19	9.50	.002	< .01
Error (C × K)	152	3.05	0.02			

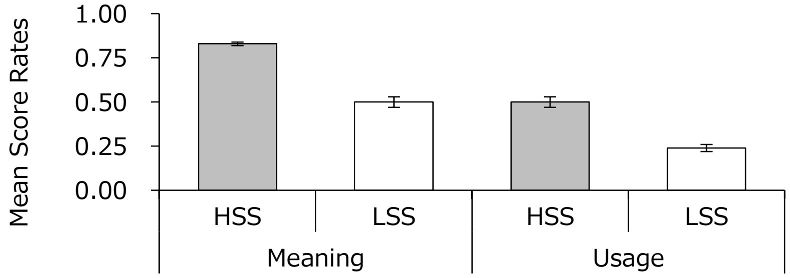


Figure 5.6. Mean score rates with ±*SEM* bars of the knowledge gains in word meaning and usage between HSS and LSS conditions ($N = 153$).

Incidental gains in word usage. The Context effects on the incidental gains in word usage knowledge can also be visualized in Figure 5.6. A post analysis demonstrated that incidental learning of word usage was promoted better in the HSS condition than in the LSS one, $F(1, 152) = 194.97$, $p < .001$, $\eta^2 = .56$, and the Context effect was large, $M_{\text{diff}} = .26$ (95% CI [.22, .30]). The other result was that the participants had more difficulty learning word usage than word meanings, regardless of the context conditions: HSS, $F(1, 152) = 171.19$, $p < .001$, $\eta^2 = .53$, $M_{\text{diff}} = .33$ (95% CI [.29, .37]); LSS, $F(1, 152) = 128.07$, $p < .001$, $\eta^2 = .46$, $M_{\text{diff}} = .26$ (95% CI [.22, .30]).

5.2.4 Discussion

In relation to H5, the experiment produced results consistent with the usage-based model. When the target words were presented in the HSS learning sentences, the performances of L2 lexical inference were greatly promoted. Considering that the HSS learning sentences evaluated by LSA reflected the high typicality of word meaning in a context (Inohara & Kusumi, 2011, 2012; Landauer & Dumais, 1997; Landauer et al., 1998), this finding indicates that L2 learners extracted word meaning from contextual word-usage information. For example, the target word *lick* was used in the HSS condition like "the dog jumped up and did something to her owner's face." In this case, it is reasonable to suppose that the participants found the most plausible meaning of the target word by thinking about how it was used to convey a coherent message; they relied on their prior experiences of word usage in contexts similar to this one.

The effects of contextual features on meaning identification during reading had been established by prior studies. Bolger et al. (2008) and Elgort et al. (2015) showed that high contextual constraint, as measured by a cloze test, promoted the activation of lexical inference. As LSA values are likely to be correlated with the cloze probability provided by lower-proficiency L2 learners (Nahatame, 2012), these two factors may be inseparably related in a natural text. Although it is difficult to distinguish their validity in terms of improving the performance of identifying the meaning of unknown words from context, it can be concluded that LSA is available to predict L2 learners' performance in lexical inference.

Prior research often suggests that MCG led to better performance in meaning identification than lexical inference (Hulstijn, 1992; Hulstijn & Laufer, 2001; Rott, 2005). The results of Experiment 5B showed that, consistent with these past studies, meaning identification was improved by the MCG task. Interestingly, the effects of word-context semantic similarity disappeared; however, there are no studies that examined context effects on MCG performance, to my knowledge. In this experiment, the participants were asked to select the most plausible gloss, distinguishing it from two distractors. Because the distractors were completely unrelated to the contextual propositions regardless of the difference in the LSA similarities (see Appendix 12), the insignificant differences in the performance of MCG can be attributed to the fact that the distractors were very easy to reject in either type of context. As explained above, the participants engaged in the MCG and dictionary use tasks to verify the correct meanings of the target words. Obtaining the high and similar scores in the MCG task between HSS and LSS conditions supported the claim that the participants could process all the target words regardless of the context conditions. After the MCG task, they used a dictionary and checked the target word meanings. These ensure that VKS results do not simply reflect the degree of success in lexical inference but the incidental gains in word knowledge affected by the Context factor.

The results of incidental learning support the assumption of RQ12, demonstrating that the participants retained the word meanings represented in memory from the HSS learning sentences better than those from the LSS learning sentences. Although they knew all the target word meanings regardless of the context conditions thanks to the MCG and dictionary use tasks, a remarkable difference in the incidental gains in word meaning knowledge was found between the HSS and LSS conditions. Similar effects of word-context semantic similarity have been obtained in the intentional learning mode reported in Experiment 5A. Thus, this finding is consistent with prior L2 vocabulary learning research at the point that a context factor will appear regardless of the learning modes (Hasegawa, 2012, 2013; Jiang, 2000; McKeown, 1985).

More importantly, the LSA values proved to predict the outcomes of incidental learning of L2 word meaning. This indicates that L2 learners obtain this type of knowledge by experiencing word usage in context, as predicted by the assumption of LSA (e.g., Landauer et al., 1998). Although the participants had an opportunity to derive the target word meanings even from the LSS learning sentences in the MCG task, the results show that the HSS learning sentences helped the participants establish a robust memory representation of the target words. The large difference in word meaning learning can be attributed to making lexical inference based on contextual information. Whereas in the intentional learning using definitions, in which lexical inference is not necessary, the Context effects were medium (η^2 = .07 in Experiment 5A), they turned out to be very large in Experiment 5B. Thus, the findings of this study demonstrated that constructed memory of a word learned from a meaningful context is incidentally integrated into the mental lexicon, as suggested by prior research (Bolger et al., 2008; Bordag et al., 2015; Elgort et al., 2015; Hamada, 2011; Hulstijn, 1992; Wesche & Paribakht, 2010).

When the possible meaning of unknown words is semantically constrained by context, incidental learning of word meaning also occurs (Bolger et al., 2008). However, there is a substantial difference between word-context semantic similarity and contextual constraint in terms of how the context features can be quantified. The strength of contextual constraint, evaluated by a cloze test, will differ according to any individual differences among learners (e.g., L2 proficiency), and this technique vitiates ecological validity (Nation & Webb, 2011). In contrast, LSA represents real-world data based on a corpus linguistic method (Landauer & Dumais, 1997; Landauer et al., 1998). Although this study did not establish which factors are more relevant, the use of natural language data like a large-scale corpus should be important in discussing the process of L2 vocabulary learning based on the usage-based model and empirical data.

Acquisition of word knowledge is an incremental process; partial knowledge of a word grows into full knowledge through multiple exposures (e.g., Jiang, 2000). This perspective was replicated by the significant difference in the knowledge gains between word meaning and usage in this study. This result is consistent with past studies, which showed that knowledge gains in word usage were followed by those in word meaning (Bolger et al., 2008; McKeown, 1985; Webb, 2005; Wesche & Paribakht, 2010).

The more valuable finding is that the Context effects were very large in incidental learning of word usage (RQ13). The grand average of the VKS4 rates (7.4 out of 20 target words; see Table 5.9) demonstrated that the use of various lexical processing strategies, including evaluation and elaboration, led to a certain degree of knowledge gain in word usage. Furthermore, the HSS learning sentences still promoted incidental learning much better than the LSS learning sentences. Whereas learning word meanings depends on the quality of context in the intentional learning mode (Hasegawa, 2013), it has small effects on learning word usage in Experiment 5A. However, in incidental vocabulary learning, in which learners have to make better use of context, the finding showed that the quality of context greatly affected the incidental gains in word usage knowledge. Such large effects of context are consistent with the assumption of the usage-based model (e.g., Ellis, 2002). Taken together, the results suggest that when L2 learners engage in cognitive processes to identify the meaning of new words from the information about how to use them in context, this information source (i.e., word usage) can also be integrated into L2 learners' mental lexicon simultaneously.

5.3 Summary of Study 3

Experiment 5A showed that contextualized vocabulary learning is affected by the semantic similarity between word and context quantified by LSA. The overall findings have important implications for the role of contexts in learning new lexical knowledge. Specifically, it demonstrates that HSS learning sentences promoted the deliberate learning of word meanings and usage regardless of vocabulary learning strategies (RQs 9 and 10). In addition, LSS learning sentences reduced its context effectiveness; especially, the inhibition effects appeared for less frequent strategy users of context when they memorize new word meanings (RQ11). Similarly, Experiment 5B demonstrated the applicability of LSA to improve the effectiveness of incidental L2 vocabulary learning. The overall findings showed that the LSA similarity should be considered an essential factor to examine the quality of contexts for purposes of incidental word learning. The answers to the hypothesis and two RQs supported some aspects of the usage-based model of language learning because the participants could derive lexical properties from HSS learning sentences better than from LSS learning sentences (H5) and they consequently gained knowledge of word meaning and usage based on the higher semantic similarity between a target word and context (RQs 12 and 13).

Chapter 6

General Discussion

6.1 Overview of Findings

In this dissertation, I examined the cognitive processes and products of incidental L2 vocabulary learning through reading activities. Based on the induction processes in language learning suggested by the usage-based model and the noticing hypothesis, the present research hypothesized that LSA similarity predicts the outcome of incidental L2 vocabulary learning. Specifically, the three empirical studies investigated how the word-context semantic similarity contributes to the generation and encoding of lexical inference, the form of memory representations of words, and their retention and retrieval. In Chapter 6, the findings extracted from each study are discussed by comparing previous views and offering some new insights into the incidental L2 vocabulary learning from reading. First, however, this section summarizes the results obtained from the individual experiments.

In Experiments 1 to 2, both ERP and RT data demonstrated that Japanese university students were so sensitive to the LSA similarity that they could generate and encode the meaning of unknown words through forward lexical inference (RQ1, RQ2, and H1). Despite the relative difficulty involved in generating backward lexical inference, further RT data demonstrated that learners could notice the semantic links between target words and their following context (RQ4). Furthermore, ERPs found that the effects of English reading proficiency appeared differently between the generation and encoding of lexical inference (RQ3).

Furthermore, Experiment 3 revealed the interaction effects of LSA similarity × L2 reading proficiency on the specificity of semantic memory of words encoded by inference (RQ5, RQ6, H2, and H3). Although HSS context promoted the generation of lexical inference, the meaning encoded in a mental representation was its superordinate for less-skilled L2 learners, whereas it was a synonym for skilled L2 learners. This encoding pattern was salient when using the on-line semantic relatedness judgment test, but the proficiency effects disappeared in the off-line lexical inference test. Regarding the establishment of form-meaning connections, Experiment 4 implied that the semantic and lexical representation of words could not be simultaneously activated after successful lexical processing (RQ7). However, the further experiment showed that task-induced text processing changed learners' allocation of attention to types of word property while reading, promoting the strong encoding of lexical representation (RQ8 and H4).

Finally, in Experiment 5, the laboratory-based findings acquired thus far were applied to the classroom-based research. The results showed that university students at beginner level gained contextualized knowledge of word meaning and usage from both intentional and incidental word learning (RQ9, RQ10, RQ12, and RQ13). At the same time, evidence showed that the learning outcomes were influenced by their vocabulary learning strategies in terms of the relative use of context while encoding word information (RQ11). Therefore, teaching procedures for learners' allocation of attention to context (e.g., the lexical inference and MCG tasks) promoted the generation and encoding of word information (H5), contributing to incidental gains in word meaning and usage knowledge.

6.2 Initial Stage of Incidental L2 Vocabulary Learning

The first step in incidental vocabulary learning from reading is to establish memory of a word by inferring its meaning from contextual information. The three experiments reported in Study 1 produced results consistent with this assumption. In particular, they demonstrated that the relative strength of LSA similarity determines the generation and encoding of lexical inference while reading in L2. In this section, these processes will be discussed from the following three perspectives: (a) the generation and encoding of inference through word anticipation and informed guessing, (b) the effects of L2 reading proficiency and methodological issues, and (c) LSA theory as a factor predicting the lexical inference processes.

First, evidence for context-based L2 lexical prediction is consistent with the recent ERP data provided by Martin et al. (2013). Although they could not find the N400 effects on unpredictable indefinite articles (e.g., "She has a nice voice and always wanted to be *a*/*an singer*/*artist*," predictable and unpredictable, respectively), those effects appeared on their next nouns (i.e., *artist*) in the same way as in Experiment 1. This provides evidence that the prior context activates a specific concept of upcoming words because of the relative difficulty in integrating that context with the unpredictable nouns as measured by the N400. Furthermore, this finding of Experiment 1 is congruent with Borovsky et al.'s (2010) ERP study, which adopts the same experimental design using unknown words as target nouns.

Further RT data from Experiments 2A and 2B strengthen the validity of Experiment 1. In fact, the semantic relatedness judgment test does not become an index of meaning generation in the time course of reading because it requires participants to complete the test based on their mental representation constructed after text processing. However, the different patterns of RT observed in both experiments between HSS and LSS contexts suggest that Japanese university students could encode a semantic memory of words as a consequence of lexical inference generation. Although I did not directly compare L2 learners to L1 readers, the consistent effects of context found in these experiments suggest that highly supportive context plays a significant role in generating lexical inference by anticipation, as demonstrated by a great number of L1 reading research (Camblin et al., 2007; DeLong et al., 2005; Otten & van Berkum, 2008, 2009; Szewczyk & Schriefers, 2013; van Berkum et al., 2005; Wicha et al., 2004).

In Experiments 2A and 2B, the on-line semantic relatedness judgment test showed the difference between forward lexical inference (i.e., context-based word anticipation) and backward lexical inference (i.e., informed guessing). When a target word has a semantic link to its subsequent context, Chaffin et al. (2001) and Huckin and Bloch (1993) have shown that readers engage in strategic rereading to connect the target word with useful contextual cues. However, such informed guessing includes operating an unknown word in working memory, finding supportive context, and then associating the message with the inferable meaning, so that Cain et al. (2004) and Daneman and Green (1986) suggest that less-skilled readers often fail to identify the correct meaning of unknown words. In fact, the current findings have added evidence that backward lexical inference is more difficult to generate during L2 reading than forward lexical inference. Nevertheless, importantly, the results of both Experiments 2A and 2B supported the generation of backward lexical inference when the target word meaning was highly semantically related to a context.

These findings regarding context-based word prediction and informed guessing are incompatible with some earlier L2 studies that concluded that the skills of lexical inference used by learners were ineffective. For example, cognitive models of the lexical inference process have indicated that it is a strategic and problem-solving type of text processing (de Bot et al., 1997; Huckin & Bloch, 1993). However, these models were developed solely from think-aloud protocol data and did not reflect the unconscious passive text processing that is observed in inferential processes. The same can be said of other research based on the relatively low accuracy of lexical inference (e.g., Bensoussan & Laufer, 1984; Hulstijn, 1992, Hulstijn et al., 1996; Nassaji, 2006) because pencil-and-paper posttest data can be indices of explicit and declarative processes and knowledge (e.g., Rebuschat, 2013). Although L2 learners' verbal reports during discourse processing will vary according to a given situation, it often results in a report that includes their lower-level language processing such as word decoding and syntactic analysis (e.g., Horiba, 2000, 2013). Therefore, the verbalizable reports should not necessarily comprise a complete picture of lexical inference.

This methodological issue is related to the null effects of L2 reading proficiency on the generation and encoding of lexical inference observed in Experiment 1. L2 studies have highlighted the importance of L2 proficiency in making lexical inferences ranging from lexical knowledge (Nassaji, 2006; Nation, 2013; Ushiro et al., 2013) to reading comprehension skills (de Bot et al., 1997; Haastrup, 1991; Wesche & Paribakht, 2010). The rationale underlying these studies is the strong correlation between L2 proficiency and success in lexical inference as measured by pencil-and-paper tests and think-aloud protocols. Although these retrospective and introspective procedures suggest that the unsuccessful production of unknown word meanings can be regarded as a failure of lexical inference generation, they do not reflect the word knowledge that is not sufficiently available to verbalize. This supports the results of Experiment 1 that L2 reading proficiency only influenced the off-line plausibility judgment accuracy, but not the N400 effects elicited by the implausible usage of the target unknown words. In other words, these suggest that the processes of less-skilled L2 learners' lexical inference were more likely to be implicit compared to skilled L2 learners.

The more important result is that LSA similarity consistently predicted L2 learners' generation and encoding of lexical inference. The present research employs LSA theory because it is closely related to the usage-based model of language acquisition underlying incidental L2 vocabulary learning (e.g., Crossley et al., 2008; Ellis, 2002; Inohara & Kusumi, 2012). Unlike corpus analyses that have pursued learners' long-term vocabulary development (e.g., Crossley et al., 2008; Kidd et al., 2010), this study hypothesized that L2 learners derive the contextual usage meaning of words from context that is highly related to their prior language experience (i.e., a particular word is more likely to be used in a particular context than in other contexts). LSA inputs a corpus to the psychological knowledge model as language experience (e.g., Landauer & Dumais, 1997), and the L2 learners in Experiments 1 to 2 were as sensitive to this information provided by LSA as to understanding the meanings of new words that are communicated by context. Together, the LSA theory and empirical findings suggest that they were engaged in understanding how the inputs are linked and integrated with their existing knowledge.

6.3 Inference Specificity and the Establishment of Form-Meaning Connections

Beyond demonstrating the generation and encoding of L2 lexical inference, Study 2 investigated the products of lexical inference from the viewpoint of mapping a semantic representation of words to a lexical representation. Study 1 demonstrated that processing a new word in context establishes a memory trace that represents the word's meaning and context; however, the results provided little information about what forms of memory were precisely encoded as a consequence of inference generation. Furthermore, the findings of Study 1 did not respond to the concern of previous studies as to why the word memories established by successful lexical inference are not retained or retrievable.

Regarding the first issue, Experiment 3A demonstrated that the specific memory of word meanings was differently represented because of individual differences in L2 reading proficiency. Unlike Experiments 2A and 2B, this experiment manipulated the specificity of probe words that will reflect inferential information in the semantic relatedness judgment test. Similar to Study 1, LSA similarity was associated with the encoding of inference. When unknown words appeared in HSS learning sentences, the accuracy and RTs of probe-prime relatedness judgments indicated that the skilled L2 learners were more able to represent specific lexical inference in their memory than less-skilled L2 learners were. With some methodological changes, including the use of the off-line lexical inference test, Experiment 3B found that the products of lexical inference differed according to quality of context, but with little effect of L2 reading proficiency on the specificity of inference encoded in memory. This second experiment added the result that whereas the specific meaning of unknown words is derived from HSS learning sentences, the context-dependent meaning is extracted from the same words in LSS sentences.

Experiment 3A demonstrates that the effects of LSA similarity and L2 reading proficiency interact in the following sense: HSS learning sentences encourage L2 learners to generate a certain lexical inference but the skills to use the contextual information differ according to their L2 proficiency, resulting in the representation of different semantic memories of unknown words. Early SLA literature has shown that less-skilled L2 learners more often fail to generate an accurate meaning of unknown words from context and ignore them without trying to make an inference (e.g., Bensoussan & Laufer, 1984; Fraser, 1999; Hulstijn, 1993; Laufer, 1997; Paribakht & Wesche, 1999). However, the findings indicate that the encoding deficit after generating lexical inference should be relevant to those studies' claims because the introspective think-aloud, retrospective interview, and recall methods are usually applicable to revealing what information is encoded and available in working (or long-term) memory (e.g., Ericson & Simon, 1980; Koda, 2005; Magliano et al., 1999). In other words, Experiment 3A suggests that although less-skilled L2 learners can at least generate lexical inference for text comprehension, their inference specificity, which is encoded in memory, determines its verbalizability in the tests.

The results of Experiment 3B support the above interpretation by eliminating time pressure on responding from the lexical inference test. Whereas the less-skilled learners in Experiment 3A could not encode a specific meaning in their memory during sentence processing, the responses produced in the off-line lexical inference test converged upon the intended meanings of the target words. This result indicates two likely interpretations. One is simply that the less-skilled L2 learners in Experiment 3B were more able to infer the specific meanings of unknown words than learners in Experiment 3A were. The other is that there were substantial differences in the meaning specification processes between the on-line and off-line tests. Because the participants' L2 reading proficiencies were not statistically different among all Experiments, the former view must not be the whole story. In contrast, the difference in nature between the two tests is whether each reflects the products as a consequence of strategic lexical inference or not. As the off-line lexical inference explicitly required the participants to produce the inferential information that had come to their mind, it is possible that the task-induced processing allowed the less-skilled L2 learners to use an encoding strategy for the completion of the test.

These two experiments further demonstrated that there were some critical differences not only amongst the experiments, but also between these experiments and previous research. According to Chaffin (1997), when a synonymous meaning of unknown words cannot be generated by L1 inference, its hypernym is activated. However, such definitional relations were not found in Experiment 3B and event-based inferences were more likely to be produced instead. For example, those who could not infer a specific meaning from *The surfers were attacked by a dangerous...* typically did not produce certain marine species but instead came up with *tidal wave*. Whereas Chaffin concluded that unknown words direct a reader's attention to comprehending their definitions through category-based inferences, Experiment 3B shows that L2 learners associate the text propositions constructed by *surfers* and *were attacked* with the unknown word, for example.

Learning words from the event-based message expressed by context is consistent with LSA theory and usage-based incremental vocabulary learning from reading. The acquisition of meaning is incremental with each piece of word knowledge stacked from various contexts (e.g., Ellis, 2002; Landauer & Dumais, 1997). Many researchers argue that such partial knowledge of words comprises contextualized temporary meaning (Balass et al., 2010; Bolger et al., 2008; Fukkink et al., 2001; Perfetti, 2007; Perfetti et al., 2005; Swanborn & de Glopper, 1999). In other words, each word meaning is not arranged into a particular semantic category at once but is gradually decontextualized after a number of exposures in different contexts. Although there are cases of single-shot word learning when readers adequately derive the word information from context (Borovsky et al., 2010, 2012), a reasonable discussion is that the usage-based meanings related to contextual messages are encoded as temporal word knowledge when L2 learners cannot generate more specific meanings.

Concerning the second issue, according to the theory of multilevel mental models constructed through text processing (e.g., van Dijk & Kintsch, 1983), Experiment 4A hypothesized that comprehending an unknown word through successful lexical inference does not promote the establishment of its form-meaning connection that is retrievable in an L2-to-L1 recall test. Constructing a semantic representation of the word results in its less robust lexical representation in terms of the relative accessibility of the word-form memories. Experiment 4B further tested that based on the noticing hypothesis (e.g., Schmidt, 1990), the encoding of word-form information in memory can be supported by task-induced focus-on-form lexical processing.

Whereas knowledge acquisition of concepts and principles described in a text requires inference generation to connect the text representation to readers' background knowledge (e.g., Kintsch, 1994; Horiba & Fukaya, 2015; Ushiro et al., 2015; van den Broek, 2010), it was not necessarily associated with incidental L2 vocabulary learning in terms of the establishment of form-meaning connections. Although the generation and encoding of lexical inference were prerequisites of incidental learning, Experiment 4A showed that some additional encoding processing was necessary because word-form information was less accessible in L2 learners' working memory when its semantic representation was constructed by successful lexical inference. Particularly, in Experiment 4B, the meaning-focused task tapped the L2 learners for excessively engaging in semantic processing of a text, and similarly, successful inference caused lower accessibility of the word-form information of the target words. Together, these results suggest that the generation and encoding of lexical inference did not satisfy the condition of incidental word learning whereby the form and meaning of targets were activated concurrently when L2 readers tried to integrate them into their long-term memories (Leung & Williams, 2011).

These findings add to the research demonstrating that surface text memory including word-form information is not maintained after successful text comprehension in terms of incidental vocabulary learning. Although it is unreasonable to conclude that levels of mental representation are specified uniquely, and text memory should include both surface and propositional representations, the asymmetric encoding of a word form and its meaning will lead to the disadvantaging of the incidental establishment of form-meaning connections. Despite differences in research paradigm and design, Experiments 4A and 4B are consistent with those that found that the available information represented in working memory differs according to the focus of attention (e.g., Dopkins & Nordlie, 2011; Gernsbacher, 1990; Gerrig et al., 2009), and those that distinguished the three levels of mental representation (e.g., Fletcher & Chrysler, 1990; Horiba, 2013; Kintsch, 1988; Kintsch et al., 1990).

The manipulation of L2 learners' attention allocation by task implementation proved to be relevant for the establishment of form-meaning connections. In Experiment 4B, the form-focused task produced different results from the meaning-focused task; specifically, the accessibility of word-form information did not differ between successful and unsuccessful lexical inferences. Generally, unknown words in a text are salient in the text representation because of a word-frequency effect (e.g., Godfroid et al., 2013), and more attention is directed to them than any other text element (e.g., Paribakht & Wesche, 1999). Nevertheless, the word-form information was less accessible when its meaning was unresolved in the meaning-focused task; contrary to this, it was still maintained after the successful construction of a semantic representation of unknown words.

Although Experiment 4B did not reveal whether a word form and its meaning were indeed connected in memory (and Experiments 5A and 5B did examine this), the findings are consistent with the effects of task-induced focus-on-form instructions on SLA. Whereas meaning-focused instructions such as lexical inference and extensive reading did not always improve L2 learners' vocabulary knowledge (e.g., Hulstijn, 1992; Mondria & Wit-de Boer, 1991; Pulido, 2007; Waring & Takaki, 2003), form-focused instructions and tasks contributed to better recall of target meanings (e.g., de la Fuente, 2006; Horiba & Fukaya, 2012, 2015; Paribakht & Wesche, 1997; Rott, 2007). The form-focused tasks could direct the learners' attention to forms while comprehending their meanings, and this attention allocation supported the learning of incidental L2 vocabulary (Bordag et al., 2015; Godfroid et al., 2013; Pellicer-Sánchez, 2015), as predicted by the noticing hypothesis (e.g., Robinson, 1995; Schmidt, 1990, 1994, 1995, 2001). Thus, the results of Experiments 4A and 4B directly connect with the relationship between attention, task, and language learning, suggesting that it is necessary for learners to control their attention in discourse processing to convert L2 lexical input into robust word knowledge.

6.4 LSA Theory and Knowledge Gains in Word Meaning and Usage

In Study 3, the relationships between LSA theories for language learning and discourse processing for word memory construction were examined in the two classroom-based experiments. Although LSA similarity predicted the generation and encoding of lexical inference in Study 1, Experiment 5A was conducted to determine if L2 learners can use the information about how typically target words are used in context in order to integrate it into their mental lexicon. In Experiment 5B, then, the effects of LSA similarity on incidental L2 vocabulary learning were tested with the implementation of the set of focus-on-form tasks, following the findings of Study 2.

First, both experiments produced results consistent with LSA theory, related to the usage-based model whereby learners extract word information from their prior language experiences (e.g., Ellis, 2002; Tomasello, 2000b, 2003). In LSA procedure, the word-context semantic similarity is assessed by the two principles of co-occurrence of words (see Section 2.2.1) and is correlated with human ratings because a large-scale corpus as the whole of human language experience is used in LSA (e.g., Inohara & Kusumi, 2011; Landauer & Dumais, 1997). According to Tomasello (2003), the results that higher recall rates of target word meanings occur in HSS learning sentences suggest that L2 learners integrate word meaning information into their mental lexicon by understanding a contextual message (e.g., in *The dog jumped up and <u>licked</u> his face*, they try to construct the situational image of what the dog is doing) and ascertaining what semantic category is appropriate for the inferred meaning (e.g., according to their prior experience, they try to categorize the word *lick* as a kind of action that a dog typically takes).

In this respect, Nassaji (2006) and Ushiro et al. (2013) show that a word-association strategy allows learners to link an unknown word to other semantically relevant words or concepts. Moreover, other think-aloud studies suggest that the use of contextual cues in a single sentence (e.g., word information, syntax, and sentence meaning) is a primary strategy (e.g., de Bot et al., 1997; Hamada, 2011; Huckin & Bloch, 1993; Nassaji, 2006; Paribakht & Wesche, 1999; Wesche & Paribakht, 2010). Considering that the knowledge acquired from exposure to words is formed as a semantic network (Crossley et al., 2008) and is semantically related to the words to which learners usually expose themselves (Inohara & Kusumi, 2012), the overall findings suggest that L2 learners can use the word-context semantic similarity quantified by LSA to extract and represent word information in their minds, resulting in knowledge gains in word meanings.

The results of Experiments 5A and 5B provide a solution to the contrastive views of the context effects on incidental L2 vocabulary learning. Although L1 research has consistently highlighted the importance of context in vocabulary learning (e.g., Bolger et al., 2008; Cunningham & Stanovich, 1991; McKeown, 1985; Nagy et al., 1987), some L2 studies have reported that there is no difference in the gains in word meaning knowledge between contextualized and decontextualized vocabulary learning (e.g., Griffith, 1992; Laufer & Shmueli, 1997; Prince, 1996; Webb, 2007a). Moreover, further studies have concluded the small effects of deriving word information from context on vocabulary growth (e.g., Bensoussan & Laufer, 1984; Hulstijn, 1992; Hulstijn et al., 1996; Mondria & Wit-de Boer, 1991). However, other L2 studies suggest that lexical inference is one of the most effective strategies in incidental L2 vocabulary learning (e.g., de Bot et al., 1997; Paribakht & Wesche, 1999; Wesche & Paribakht, 2010). According to Bordag et al. (2015), a possible interpretation that explains this discrepancy is the difference in methodology used to visualize the knowledge acquired from reading. The present study further emphasizes the significant connections between text processing and knowledge acquisition, described in LSA theory and relevant models (e.g., the usage-based model), because this perspective has rarely been reflected in the early studies (Inohara & Kusumi, 2012; Horiba & Fukaya, 2012, 2015; Pulido, 2007).

Knowledge gains in word usage also improved in contextualized incidental learning. Because LSA can represent the contextual usage meaning of words, it is possible that contexts with high LSA values promoted L2 learners to derive the information of word usage from them. The effects of lexical inference on the development of productive knowledge were examined in terms of how frequently learners encounter target words in a text because they assumed that a single context has little effect on knowledge gains (e.g., Chen & Truscott, 2010; Rott, 2007; Webb, 2007b). However, the present study shows that the LSA values predict the outcomes of knowledge gains in word usage, suggesting that context effects have a strong influence on the encounter frequency required to gain knowledge of word usage.

It is beyond the scope of Study 3 to test the entire picture of word usage knowledge (e.g., constraints on use), but it does demonstrate that participants used words with higher accuracy of part of speech, inflection, semantics, and syntax when they learned them from HSS learning sentences than from LSS learning sentences. According to Nation (2013), these types of word knowledge can be acquired from implicit instruction such as extensive reading. Study 3 could not directly compare contextualized and decontextualized word learning; however, the findings support the significance of the contextualized learning suggested by Bolger et al. (2008) and McKeown (1985), thus demonstrating that it results in better performance on tasks using words in context.

However, Experiment 5A demonstrates that the advantage of contextualized word learning can deteriorate by using the less frequent strategy of using context. According to Mizumoto and Takeuchi (2009), individual differences in the use of vocabulary learning strategies determine what features of words the L2 learners pay attention to. This may be related to the case of less incidental vocabulary learning because Nassaji (2006) shows that whereas less-skilled L2 learners use word-based strategies (e.g., word analysis) in lexical inference, skilled learners make more frequent and effective use of context-based strategies (e.g., sentence meaning). These can also be explained from the viewpoint of the noticing hypothesis (Robinson, 1995) and the results of Experiment 4, indicating that unless the learners direct their attention toward a context in deriving word information, it results in difficulty in knowledge acquisition from reading.

Regarding the importance of attention in contextualized, incidental L2 vocabulary learning, Experiment 5B demonstrates the effects of the tasks and activities that direct L2 learners' attention to contexts and form-meaning relationships. In the lexical inference activity, as there were no available cues in the target words themselves, the participants were encouraged to use contextual information to generate their meanings. According to Hulstijn et al. (1996) and Rott (2005), the MCG and dictionary use activities[8] were administered for the participants to elaborate the inferred meaning of new words by paying attention to the word itself. Interestingly, these techniques did not promote incidental knowledge gains of all the target words in an equable manner. The results that HSS learning sentences were still superior to LSS learning sentences in incidental L2 vocabulary learning suggest that although L2 learners appropriately allocate their attention to contextual information and form-meaning links, the availability of context defined by LSA plays a central role in vocabulary learning from reading. Thus, related to the usage-based model of language learning, LSA theory can specify the processes of incidental L2 vocabulary learning. The overall findings suggest that knowledge acquisition from context occurs with various language experiences as the construction of word memory progresses through reading comprehension.

[8] Note that the purpose of dictionary use in Experiment 5B was only to inform the participants of the meaning of all the target words as a pedagogical intervention (see Section 5.2.2.2). The relationship between dictionary use and incidental vocabulary gains were not focused here.

Chapter 7

Conclusions

7.1 Summary of This Dissertation

This dissertation have demonstrated the high applicability of LSA to improve the effectiveness of incidental L2 vocabulary learning. The theoretical finding was that LSA theory is closely related to L2 knowledge acquisition, which is associated with theories of discourse processing. From a practical viewpoint, the findings demonstrated that the word-context semantic similarity computed by LSA should be considered an essential factor in examining the quality of contexts for purposes of incidental word learning. Generally speaking, I can conclude in this dissertation that the way in which L2 learners incidentally learned new vocabulary through reading differs according to the interaction effects of contextual quality, learning tasks, and individual differences.

Specifically, the context effect on incidental vocabulary learning is supported by the usage-based model of language learning behind LSA theory. Given that the correct induction of word knowledge from context leads to vocabulary growth, it is important to consider the degree of informativeness of context for improving learning outcomes. This induction process is supported by inference in discourse processing. Furthermore, in the context of vocabulary learning from reading, appropriate allocation of attention to both the form and meaning aspects of words is essential. These cognitive processes can be controlled by task implementation, which enables L2 learners to establish form-meaning connections while reading a text. As has been frequently examined in SLA literature, incidental L2 word learning is further influenced by individual differences in L2 reading proficiency levels and learning strategies. However, this dissertation suggests that the effects of such learner factors can be modulated by the manipulation of the context informativeness and learning tasks. In line with these concluding remarks, I will discuss some pedagogical implications in the next section.

7.2 Pedagogical Implications

When conducting curriculum design for vocabulary learning in a language course, according to Nation (2013, p. 574), it is necessary to reflect the following three principles of vocabulary teaching: content and sequencing, format and presenting, and monitoring and assessment. Although each of these aspects has several constructs, this section will focus on how to achieve the purposes listed below based on the research findings:

(1) *Content and sequencing*: To provide an opportunity to learn the various constructs of vocabulary knowledge.

(2) *Format and presenting*: To ensure that the target words occur in meaning-focused input, language-focused (i.e., form-focused; e.g., R. Ellis, 2008) learning, meaning-focused output, and fluency development activities.

(3) *Monitoring and assessment*: To encourage learners to reflect on their learning.

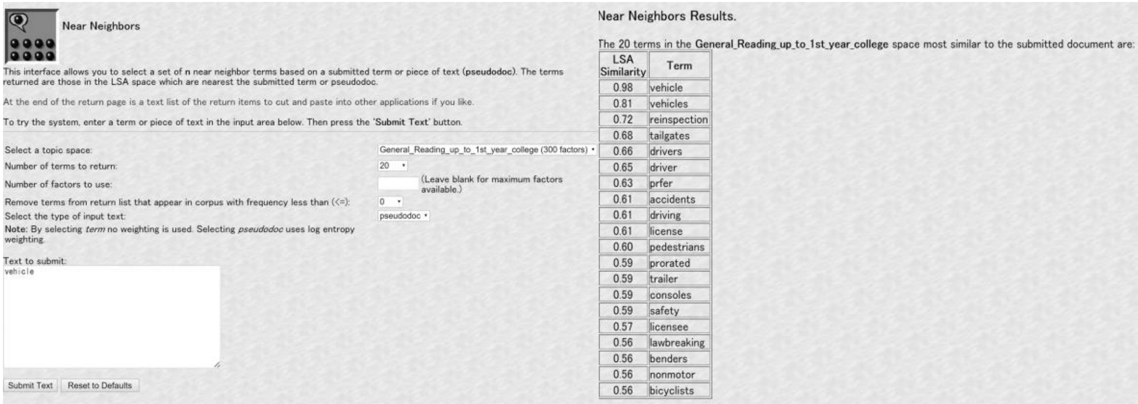
In Section 7.2.1, based on the first principle, I will give an account of the procedure for developing materials for incidental L2 vocabulary learning using LSA. Next, the potential effects of a task-induced focus-on-form approach will be discussed in terms of the second principle. Finally, I will highlight the advantage of the multi-componential and multi-test approach to visualizing the learners' vocabulary knowledge as the third principle.

7.2.1 Using LSA to Develop Teaching Materials

According to Nation (2013), one of the most important decisions regarding content and sequencing is to decide what vocabulary will be used with what ideas (e.g., topics) and uses (e.g., situations) in each lesson or unit. For example, R. Ellis (2008) suggests that if the target tasks are appropriately selected, the suitable representation of vocabulary and grammatical functions will automatically be met. To this end, corpus-based research is sure to provide the procedure underlying the specification of the content and sequencing regarding how vocabulary can be introduced in each task. LSA should play an important role in this decision making because it can assess how typically a target word is used in a particular context and situation.

Specifically, in Experiments 5A and 5B, data on word learning performance were collected from practical activities in a regular English class; the procedure of the material development may be a useful reference to improve contextualized L2 vocabulary learning. Although all the Japanese university students in these experiments engaged in the same lexical processing activities (i.e., memorization, lexical inference, MCG, and dictionary use), the learning outcomes differed according to the quality of context. If an original context does not provide the adequate lexical properties of a target word used in a particular task, it can be modified to better help contextualized learning of the word. Next, LSA can be applied to the modified version of the context to evaluate its informativeness. According to Inohara and Kusumi (2011), the strength of semantic similarity between a target word and its context increases if the context is provided with words that are semantically associated with the target word. Because access to the LSA web site (http://lsa.colorado.edu/) is open, teachers can easily compare the quality of the original contexts with revised ones. Figure 7.1 shows the full procedure when the target word is *vehicle*. Note that the LSA value of the word-context semantic similarity between *This afternoon George taught his son to drive a* and *vehicle* was .22.

[1] To find what kinds of words are typically used with *vehicle*, "the near-neighbors application" can be used, returning a list of words that are most similar to a word entered into "Text to submit" (e.g., *driver*, *accidents*, *license*, *safety*, and *bicyclists*).



[2] After creating a learning sentence with the words listed above (e.g., *Those who drive this vehicle need a special driver's license*), the degree of its informativeness can be evaluated by "the one to many comparison" tool (.22 → .58).

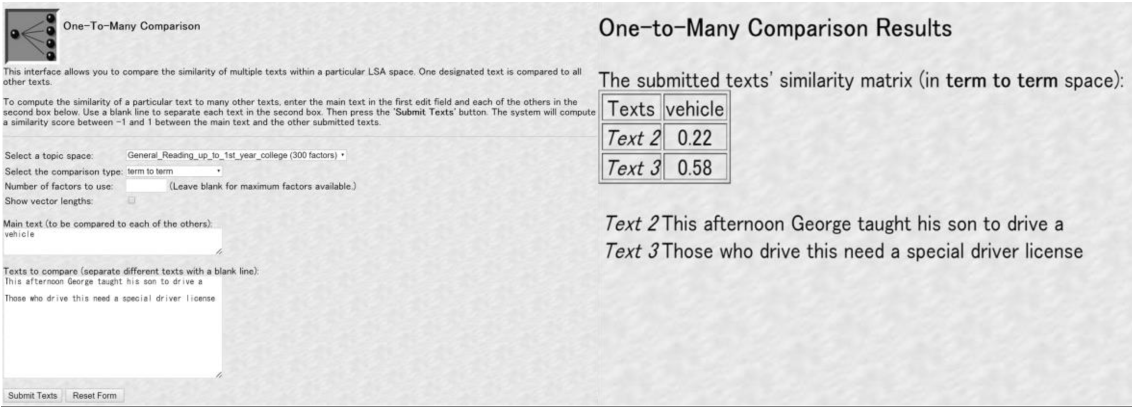


Figure 7.1. An example procedure of using LSA to create the learning sentence with high semantic similarity between a word and a context (see Dennis, 2007 for details).

When I conducted a Google Search for these near neighbors, they were frequently found in the driver's license information served by the Department of Transportation. Therefore, for example, a decision making task, in which students read the information and decide what kind of driver's license is required to drive a particular vehicle, is expected to improve their contextualized knowledge of words from reading. Thus, useful learning sentences require not only that unknown words be comprehensible but also information regarding how the words are typically used in context.

7.2.2 Task-Induced Focus-on-Form Approach

In view of format and presentation, the selection of teaching and learning techniques in a lesson plan is important (Nation, 2013). Note that it is beyond the scope of this dissertation to cover all the important aspects of vocabulary teaching and learning strategies, and the findings of this study are mainly related to meaning-focused and form-focused input processing. The relationship between contextualized vocabulary learning and individual differences in L2 reading proficiencies and vocabulary learning strategies will be discussed here.

As a general implication for reading instructions, it is important to consider the learners' cognitive processes involved in discourse processing (e.g., Grabe, 2009; Horiba, 2000; Koda, 2005; Ushiro, 2010). Discourse processing is closely related to meaning-focused input processing for vocabulary learning from a text because the overall experiments reported in this dissertation demonstrated the complex interaction between context quality and students' English reading proficiency in meaning generation by lexical inference. Regarding the issue of context quality, as mentioned in Section 7.2.1, teachers should evaluate in advance whether the contextual information is semantically related to the inferable meaning of the target words presented for a lexical inference task and understand the strength of semantic similarity effects on lexical inference specificity.

In the use of HSS learning sentences, students are expected to identify the specific meaning of the target words; however, it should be considered that less-skilled L2 learners may have difficulty narrowing down the meaning of unknown words through their own efforts, even if the contextual information is semantically related to their possible meanings. Therefore, in the same way as the given $P_{specific}$ promoted participants' success rates in the semantic relatedness judgment test (Experiment 3A), teachers can help learners by providing multiple marginal glosses or questioning (e.g., *Breaking and entering into a house is against the ____ in every country*. Which of the following goes in the blank: *rule*, *common sense*, or *law*?). The effects of multiple marginal glosses on meaning identification are supported by the results of Experiment 5A. Importantly, according to Hulstijn et al. (1996) and Rott (2005), a dictionary use activity should be preceded by marginal glosses and questioning. Whereas these activities require students to direct their attention to contextual information in order to select a correct answer, if the dictionary use activity is administered in isolation, the students may be satisfied with simply looking up the correct meaning of a target word in a dictionary without integrating it into the context.

LSS learning sentences should be useful for improving students' skills of lexical inference. Experiments 3A and 3B suggest that students may hardly infer the appropriate meaning of target words from only such contextual information, but they are required to generate at least a general or vague lexical inference for reading comprehension. When they come up with some widely interpreted meaning, they must then narrow and verify its specificity with the succeeding contextual information. If the inferred meaning is too vague or unexpected, teachers need to help the students make their inferences more specific by asking for additional explanation of their reasoning.

Task-based focus-on-form instructions should be implemented to promote students to establish the form-meaning connections of target words. This is one of the fundamental goals of vocabulary acquisition; however, Experiments 4A and 4B have demonstrated that the meaning-focused reading does not necessarily contribute to the construction of an intra-word network. Therefore, as reported in Horiba and Fukaya (2012, 2015), a particular task is required for L2 learners to direct their attention not only to word meanings but also their forms in order to improve encoding. Although it is difficult to reproduce a specific communicative situation perfectly, in the context of reading instruction, teachers can present their students with an information transfer task, in which the students must transfer text information that they have understood in some way to complete the task demand (e.g., sending an email regarding hotel bookings to a friend after reading a hotel pamphlet). At the same time, it is important to require the students to communicate this in their L2 because doing so can facilitate the encoding of word forms as well as their meanings (Horiba & Fukaya, 2012, 2015; see also Experiments 4A and 4B).

Finally, the instruction of vocabulary learning strategies is important to derive benefits from HSS learning sentences. Experiment 5A suggests that although the HS learning sentences were consistently effective, regardless of the types of vocabulary learning strategies used, the less frequent users of contexts might have lost these advantages. Because Mizumoto and Takeuchi (2009) have demonstrated that vocabulary learning strategies are trainable skills, teachers should understand how individuals approach the word learning task and what types of strategy are required to integrate the given learning sentences with the lexical knowledge of a target word. To this end, some pedagogical interventions, such as strategy training, are required to direct the students' attention to the effectiveness and informativeness of using context to learn new words (e.g., Fraser, 1999; Gu & Johnson, 1996; Nation, 2013).

Every L2 learner at any proficiency level encounters unknown words in texts; thus, it is important to teach L2 learners how to process unknown words using lexical inference (Fraser, 1999). Although this dissertation indicates that L2 learners generate context-based lexical inference without specific interventions, such as a think-aloud task, concern that their inferred meanings might be incorrect will increase until they receive corrective feedback (Hulstijn, 1993), leading to learning of fewer new words from reading (Laufer, 1997). This study shows that L2 learners are inclined to make lexical inference from highly elaborative contexts. These perspectives suggest that teachers should remember the fundamental principle of SLA suggested by Krashen (1989): A comprehensible input develops learners' language knowledge and skills. Furthermore, cognitive approaches regarding attention allocation by task implementation are expected to compensate incidental L2 vocabulary learning

7.2.3 Multiple-Componential and Multi-Test Approach Assessments

Although both teachers and researchers must monitor learners' progress and the quality of their word knowledge as a consequence of learning, word knowledge includes different developmental stages and constructs, which is remarkable in terms of the knowledge acquired from reading (e.g., Nation, 2013; Webb, 2007a, 2008). Because contextualized knowledge is not conveyed in a single measure, it should be assessed using multiple dependent measures (Bolger et al., 2008). As Pellicer-Sánchez (2015) suggested, the results of this research demonstrate the advantage of a multi-componential and multi-test approach to assessing the stages and constructs of vocabulary knowledge ranging from processes and products.

Because incidental L2 vocabulary learning often starts with the collection of word information from context using inference, first, the initial knowledge encoded in memory, which provides information about L2 learners' inference skills, should be measured. Even in a classroom setting, sensitive measures should be used to rate the inference outcomes produced in the think-aloud and paper-and-pencil tests. For example, many researchers have used a 3-point-scale index such as success, partial success, and failure (e.g., Fukkink et al., 2001; Nassaji, 2006; Wesche & Paribakht, 2010). Successful inferences are defined as those that are semantically, grammatically, and contextually appropriate. Inferences that semantically match contextual information are classified as partially successful, even if they are far from the original meaning of unknown words. In addition to this scoring method, the results of Experiment 3B have highlighted the importance of categorizing the relations between inferences and unknown word meanings (e.g., synonymous, categorical, event-based, or non-semantic relations). This categorization shows what aspects of word and contextual information learners shift their attention to (Chaffin et al., 1997). Fukkink et al. (2001) provided a fuller account of the need for such sensitive measures as follows: (a) readers glean only partial knowledge about an unknown word from a single context, and (b) the knowledge of a word develops incrementally as readers encounter the word in different contexts. In their meta-analysis, Swanborn and de Glopper (1999) showed that multi-scale scoring is more accurate than a dichotomous scoring method when assessing the outcomes of incidental vocabulary learning.

In the context of vocabulary research, data from on-line measures provide fuller information about the covert knowledge that cannot be elicited by the think-aloud method. In Section 2.5, I mentioned several on-line tests that reflect the generation and encoding of lexical inference. When both research design (e.g., stimuli creation) and rationales for interpreting the data obtained from tests are carefully constructed, evidence is provided regarding the relationship between lexical processing and learning products (Pellicer-Sánchez, 2015). In this dissertation, Experiment 1 added the findings that ERPs are applicable to examining and encoding lexical inference generation (Elgort et al., 2015). Although previous studies interpreted that the longer reading times of target unknown words reflect learners' attempts to infer those meanings from context (a self-paced reading method in Bordag et al., 2015; Hamada, 2012; an eye-tracking method in Godfroid et al., 2013; Pellicer-Sánchez, 2015), this index is not necessarily linked with semantic processing, unlike the N400 component. For example, the longer fixations of unknown words can be interpreted as a preliminary step in lexical inference, such as word identification and attention allocation (see Rayner, 2009 for details). Therefore, it is important to distinguish between what the tests can and cannot measure, and to complement the findings with each other using different test approaches for further validation (Jiang, 2012).

Second, multi-componential tests have been used to reveal contextualized word knowledge as a consequence of lexical inference (Bolger et al., 2008; Horiba & Fukaya, 2012, 2015; Webb, 2007a, 2007b, 2008; Wesche & Paribakht, 1996, 2010). To probe the acquisition of word knowledge, at the very least, tests should be divided into the following three constructs: form, meaning, and usage (Nation, 2013). Additionally, the differences in word memory between recognition and recall can be taken into account when tests are designed to reveal the developmental stages of vocabulary growth (Webb, 2005, 2007a, 2007b). L1-to-L2 and L2-to-L1 recall tests have been consistently used in the literature to assess form and meaning knowledge, respectively (e.g., Bolger et al., 2008; Hasegawa, 2012, 2013; Horiba & Fukaya, 2006, 2012, 2015). The same procedure is often applied to the recognition test for the assessment of recognition memory if learners have only a few opportunities to learn words from reading (e.g., Chen & Truscott, 2010; Pellicer-Sánchez, 2015; Webb, 2007b, 2008). These tests can easily be designed and implemented in regular English classes (Aizawa & Mochizuki, 2010) and show the learners' progress in the establishment of form-meaning connections (e.g., Schmitt, 2010).

Regarding knowledge gains in word usage, Experiments 5A and 5B used the VKS test. This test is often criticized in terms of the measurable range of word knowledge because it generally assesses knowledge of only one meaning of the target word and its related use (Schmitt, 2010). Given that the abstract knowledge obtained from multiple exposures to a target word is multidimensional (e.g., Bolger et al., 2008), the VKS test is useful to assess the relatively initial stage of contextualized knowledge. Furthermore, the VKS test is a self-assessment format and test-takers' avoidance strategy will cause a measurement error. It is possible that less certainty of word usage knowledge causes them to avoid attempting the sentence completion category (i.e., VKS4 in Experiments 5A and

5B). Although the degree of self-confidence regarding how to use knowledge in communication is one of the components of linguistic competence (Rebuschat, 2013), teachers must be careful to implement the VKS test for more accurate and validated assessment of word usage knowledge. In this case, the sentence completion test by a forced-choice procedure with multiple test items (i.e., a gap-filling test) is useful for assessing the usage knowledge of a particular word (Bolger et al., 2008).

7.3 Limitations and Suggestions for Future Study

This dissertation has expanded on previous findings of incidental L2 vocabulary learning, providing informative suggestions for L2 lexical inference and successive vocabulary learning. However, the present investigations have not provided an understanding of the entire picture of incidental L2 vocabulary learning.

First, as general limitations that are related to all the experiments, future research should address the following three issues: (a) to reveal how L2 learners incidentally gain knowledge of novel words in discourse beyond a single sentence, (b) to explore the incremental processes of vocabulary knowledge development from the viewpoint of the effects of input frequency, and (c) to distinguish the nature of context quality, particularly, between word-context semantic similarity computed by LSA and contextual constraint provided by a cloze probability. To solve these limitations, specific concerns regarding the individual experiments are described and the required research design is discussed below.

In Experiment 1, the ERP study, the number of stimuli used was relatively lower than typical sentence processing studies (e.g., 20 to 40 stimuli per experimental condition), although the ERP was extracted from the EEG using an averaging procedure (e.g., Morgan-Short & Tanner, 2014). In addition, Experiment 1 (and Experiments 2 to 4) dealt with the processing and products of the concrete nouns only, but concreteness and part of speech also affect both the guessability and learnability of words. However, a recording session for a single ERP experiment usually takes a long time and causes a measurement error due to participants' inability to concentrate. In practice, it is difficult to examine the multiple possible factors affecting the generation and encoding of lexical inference with the ERPs.

Experiments 2A and 2B concluded that the process of lexical inference was not specific to completing a think-aloud task; however, it is possible that the participants in these experiments inferred the target word meanings strategically to perform the semantic relatedness judgment test. Furthermore, the test procedure may develop strategic processing of unknown words because the required response to experimental target-probe pairs was always "yes." Therefore, future research should replicate the findings using different methodologies, such as eye-tracking measures (e.g., Chaffin et al., 2001; Godfroid et al., 2013; Pellicer-Sánchez, 2015), particularly in the backward elaboration condition. A more practical limitation of Experiments 2A and 2B is a sparse collection of judgment accuracy and RT data, although a longer experiment might disimprove the quality of the behavioral data.

Regarding Experiments 3 to 4, research on discourse-based lexical inference should be connected with L2 reading comprehension and successive vocabulary learning. As in many previous studies (e.g., Nassaji, 2003, 2006), the experiments demonstrated that the generated lexical inference was sometimes unexpected even if contexts support meaning specification processes. In this case, L2 learners often distort the original messages of contexts in order to match the misinterpreted word meanings with them (Huckin & Bloch, 1993; Laufer, 1997), causing the incorrect learning of words (Hulstijn, 1992). Therefore, L2 learners are required to modify their incorrect lexical inferences as a text unfolds. In other words, when the meanings extracted from contexts are too vague or inappropriate for a succeeding context, how can the generated inference be revised based on discourse and how can the revised inference be integrated into a mental representation of discourse? Although it seems challenging to test for revision processes of lexical inferences, some approaches, such as think-aloud, eye-tracking, and ERPs methods, can grasp these processes. As mentioned previously, these measures are applicable to understanding discourse-processing and learning mechanisms (e.g., Borovsky et al., 2010; Camblin et al., 2007; Chaffin et al., 2001; Elgort et al., 2015; Just & Carpenter, 1980, 1992; Kutas & Federmeier, 2011; Otten & van Berkum, 2008, 2009; van Berkum et al., 2005), and the integration of each research result will reveal more detail regarding incidental L2 vocabulary learning from discourse.

Although these findings of Experiment 5A provide a new contribution to a better understanding of LSA use to improve contextualized vocabulary learning, the number of learning sentences used in the experiment was small. Given the proficiency level of the participants, it is reasonable that 20 sentences were presented for each target word; however, further research should conduct an item analysis focusing on the variances of LSA values. For example, a regression analysis will reveal how well LSA values predict the learning outcomes of word meanings and usages, and differentiate themselves from the strength of contextual constraint. Another possible limitation is that no delayed recall test was conducted in Experiment 5A; it is possible that this experiment could not reveal a clear difference in the context effects of contextualized vocabulary learning. In fact, although the results demonstrated significant main effects of the semantic similarities, the effect sizes were relatively small. Experiment 5B complemented this finding, but it is important to examine what knowledge is retained after memory decay from the viewpoints of the semantic similarities based on LSA. In particular, a further study should consider context variations to investigate the incremental process of learning new words from context.

Similarly, in Experiment 5B, because incidental vocabulary learning from reading is an incremental process and is influenced by various factors (e.g., Bolger et al., 2008; Jiang, 2000), it is important to investigate its interaction with any possible factors, such as frequency of word encounter and individual learner differences. Frequency of input should especially be considered because the learning source of LSA is also massive input of language (i.e., large-scale corpora), following the usage-based model (e.g., Landauer et al., 1998). Future research, therefore, must focus on the time-dependent change in incidental word learning from multiple exposures to target words (Nation & Webb, 2011). This point is also critical to reveal how lexical knowledge gained from context is integrated into long-term memory.

In the area of second and foreign language learning, empirical research using LSA has only recently begun. Despite the set of limitations, a computational approach such as LSA is both necessary and beneficial. LSA represents real-world data based on natural language processing theory and psycholinguistic and corpus linguistic methods. As the word learning performance of Japanese EFL learners in this study was congruent with the prediction of LSA, more studies are required to consider natural language data and computational tools. Such approaches would allow us to improve the effectiveness of contextualized and incidental L2 vocabulary learning and provide valuable insights into discussing the theories regarding the relationship between discourse processing and L2 vocabulary learning.

References

Aizawa, K., & Mochizuki, M. (Eds.). (2010). *Eigo goishidou no jissen idea shu: Katsudourei kara test sakusei made* [Practical handbook for English vocabulary teaching]. Tokyo, Japan: Taishukan Shoten.

Amano, S., & Kondo, T. (Eds.). (1999). *NTT database series "Nihongo no goi tokusei" dai 1-kan tango shinmitsudo* [NTT database series "lexical properties of Japanese word" vol.1 word familiarity]. Tokyo, Japan: Sanseido.

Amano, S., & Kondo, T. (Eds.). (2000). *NTT database series "Nihongo no goi tokusei" dai 7-kan hindo* [NTT database series "lexical properties of Japanese word" vol.7 word frequency]. Tokyo, Japan: Sanseido.

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., & Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and Language, 39,* 187–205. doi:10.1016/0093-934X(90)90011-5

Balass, M., Nelson, J. R., & Perfetti, C. A. (2010). Word learning: An ERP investigation of word experience effects on recognition and word processing. *Contemporary Educational Psychology*, *35*, 126–140. doi:10.1016/j.cedpsych.2010.04.001

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, *52,* 323–363. doi:10.1111/0023-8333.00186

Barcroft, J. (2003). Effects of questions about word meaning during L2 Spanish lexical learning. *The Modern Language Journal*, *87,* 546–561. doi:10.1111/1540-4781.00207

Barcroft, J. (2009). Effects of synonym generation on incidental and intentional L2 vocabulary learning during reading. *TESOL Quarterly*, *43*, 79–103. doi:10.1002/j.1545-7249.2009.tb00228.x

Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading*, *7*, 15–32. doi:10.1111/j.1467-9817.1984.tb00252.x

Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes*, *45*, 122–159. doi:10.1080/01638530701792826

Bordag, D., Kirschenbaum, A., Tschirner, E., & Optiz, A. (2015). Incidental acquisition of new words during reading in L2: Inference of meaning and its integration in the L2 mental lexicon. *Bilingualism: Language and Cognition*, *18*, 372–390. doi:10.1017/S1366728914000078

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development*, *8*, 278–302. doi:10.1080/15475441.2011.614893

Borovsky, A., Kutas, M., & Elman, J. L. (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition*, *116*, 289–296. doi:10.1016/j.cognition.2010.05.004

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from contexts: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *96*, 671–681. doi:10.1037/0022-0663.96.4.671

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language*, *56*, 103–128. doi:10.1016/j.jml.2006.07.005

Chaffin, R. (1997). Associations to unfamiliar words: Learning the meanings of new words. *Memory & Cognition*, *25*, 203–226. doi:10.3758/BF03201113

Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning new word meanings from context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 225–235. doi:10.1037/0278-7393.27.1.225

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, *31*, 693–713. doi:10.1093/applin/amq031

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Crossley, S. A., Salsbury, T., McCarthy, P., & McNamara, D. S. (2008). Using latent semantic analysis to explore second language lexical development. *Proceedings of the twenty-first international FLAIRS conference*, *12*, 136–141. Retrieved from http://www.aaai.org/Papers/FLAIRS/2008/FLAIRS08-040.pdf

Crossley, S. A., Salsbury, T., Titak, A., & McNamara, D. S. (2014). Frequency effects and second language lexical acquisition: Word types, word tokens, and word production. *International Journal of Corpus Linguistics*, *19*, 301–332. doi:10.1075/ijcl.19.3.01cro

Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology*, *83*, 264–274. doi:10.1037/0022-0663.83.2.264

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*, 1–18. doi:10.1016/0749-596X(86)90018-5

Day, R. R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language*, *7*, 541–551. Retrieved from http://www.nflrc.hawaii.edu/rfl/PastIssues/rfl72day.pdf de Bot, K., Paribakht, T. S., & Wesche, M. B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition: Evidence from ESL reading. *Studies in Second Language Acquisition*, *19*, 309–329.

de la Fuente, M. J. (2006). Classroom L2 vocabulary acquisition: Investigating the role of pedagogical tasks and form-focused instruction. *Language Teaching Research*, *10*, 263–295. doi:10.1191/1362168806lr196oa

Deerwester, S., Dumais, S., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of American Society for Information Science*, *41*, 391–407. doi:10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121. doi:10.1038/nn1504

Dennis, S. (2007). How to use the LSA web site. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 57–70). Mahwah, NJ: Lawrence Erlbaum Associates.

Deschambault, R. (2012). Thinking-aloud as talking-in-interaction: Reinterpreting how L2 lexical inferencing gets done. *Language Learning*, *62*, 266–301. doi:10.1111/j.1467-9922.2011.00653.x

Dopkins, S., & Ngo, C. T. (2005). Role of recognition memory in anaphor identification. *Journal of Memory and Language*, *53*, 186–203. doi:10.1016/j.jml.2004.07.003

Dopkins, S., & Nordlie, J. (2011). Exploring a decrease in recognition performance for non-antecedents following the processing of anaphors. *Discourse Processes*, *48*, 432–451. doi:10.1080/0163853X.2011.564573

Educational Testing Service. (2007). *TOEIC Bridge® official guide & question collection*. Tokyo, Japan: Institute for International Business Communication.

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, *61*, 367–413. doi:10.1111/j.1467-9922.2010.00613.x

Elgort, I., Perfetti, C. A., Rickles, B., & Stafura, J. Z. (2015). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and event-related brain potential (ERP) indicators of learning. *Language*, *Cognition and Neuroscience*, *30*, 506–528. doi:10.1080/23273798.2014.942673

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*, 365–414. doi:10.1111/lang.12052

Ellis, N. C. (1994). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 211 –282). London, England: Academic Press.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 225–241. doi:10.1017.S0272263102002024

Ellis, N. C. (2008). The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The Modern Language Journal*, *92*, 232–249. doi:10.1111/j.1540-4781.2008.00716.x

Ellis, N. C., & Ferreira-Junior, F. (2009). Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal*, *93*, 370–385. doi:10.1111/j.1540-4781.2009.00896.x

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2013). Usage-based language: Investigating the latent structures that underpin acquisition. *Language Learning*, *63*, 25–51. doi: 10.1111/j.1467-9922.2012.00736.x

Ellis, N. C., & Wulff, S. (2015). Usage-based approaches to SLA. In B. van Patten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 75–93). New York, NY: Routledge.

Ellis, R. (2008). *The study of second language acquisition*. Oxford University Press.

Ericson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215–251. doi:10.1037/0033-295X.87.3.215

Fincher-Kiefer, R. (1995). Relative inhibition following the encoding of bridging and predictive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 981–995. doi:10.1037/0278-7393.21.4.981

Fincher-Kiefer, R. (1996). Encoding differences between bridging and predictive inferences. *Discourse Processes*, *22*, 225–246. doi:10.1080/01638539609544974

Fletcher, C. R. (1992). Assessing recognition memory for surface forms in discourse: A methodological note. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 199–203. doi:10.1037/0278-7393.18.1.199

Fletcher, C. R. (1994). Levels of representation in memory for discourse. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 589–607). San Diego, CA: Academic Press.

Fletcher, C. R., & Chrysler, S. T. (1990). Surface forms, textbases, and situation models: Recognition memory for three types of textual information. *Discourse Processes*, *13*, 175–190. doi:10.1080/01638539009544752

Frantzen, D. (2003). Factors affecting how second language Spanish students derive meaning from context. *The Modern Language Journal*, *87*, 168–199. doi:10.1111/1540-4781.00185

Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition*, *21*, 225–241.

Frishkoff, G. A., Perfetti, C. A., & Collins-Thompson, K. (2010). Lexical quality in the Brain: ERP evidence for robust word learning from context. *Developmental Neuropsychology, 35,* 376–403. doi:10.1080/87565641.2010.480915

Fukkink, R. G., Blok, H., & de Glopper, K. (2001). Deriving word meaning from written context: A multicomponential skill. *Language Learning, 51,* 477–496. doi:10.1111/0023-8333.00162

Fukkink, R. G., & de Glopper, K. (1998). Effects of instruction in deriving word meaning from context: A meta-analysis. *Review of Educational Research, 68,* 450–469. doi:10.3102/00346543068004450

Fukuta, J. (2015). Potential methodological biases in research on learning without awareness. *Applied Linguistics*. Advance online publication. doi:10.1093/applin/amv022

Gass, S., Behney, J., & Plonsky, L. (2013). *Second language acquisition: An introductory course* (4th ed.). New York, NY: Routledge.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. New York, NY: Psychology Press.

Gerrig, R. J., Love, J., & McKoon, G. (2009). Waiting for Brandon: How readers respond to small mysteries. *Journal of Memory and Language, 60,* 144–153. doi:10.1016/j.jml.2008.08.004

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition, 35,* 483–517. doi:10.1017/S0272263113000119

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48,* 163–189. doi:10.1146/annurev.psych.48.1.163

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395. doi:10.1037/0033-295X.101.3.371

Griffin, G. F. (1992). *Aspects of the psychology of second language vocabulary list learning*. (Doctor dissertation). Retrieved from http://wrap.warwick.ac.uk/36070/

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, *38*, 313–338. doi:10.1006/jmla.1997.2547

Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning*, *46*, 643–679. doi:10.1111/j.1467-1770.1996.tb01355.x

Haastrup, K. (1991). *Lexical inferencing procedures or talking about words: Receptive procedures in foreign language learning with special reference to English*. Tügingen, Germany: Gunter Narr.

Hama, M., & Leow, R. P. (2010). Learning without awareness revisited: Extending Williams (2005). *Studies in Second Language Acquisition*, *32*, 465–491. doi:10.1017/S0272263110000045

Hamada, A. (2011). Lexical inferencing and the acquisition of vocabulary depth: Focusing on strategy use and contextual information. *ARELE: annual review of English language education in Japan*, *22*, 313–328. Retrieved from http://ci.nii.ac.jp/naid/110009425264/

Hamada, A. (2012). How EFL learners process unknown words while reading: Effects of contextual constraints and recall instructions. *ARELE: annual review of English language education in Japan*, *23*, 249–264. Retrieved from http://ci.nii.ac.jp/naid/110009603308

Hasegawa, Y. (2012). Incidental vocabulary learning and memory of context: Interaction between informativeness and imageability. *ARELE: annual review of English language education in Japan*, *23*, 201–216. Retrieved from http://ci.nii.ac.jp/naid/110009603305

Hasegawa, Y. (2013). Contextualized vocabulary learning by two proficiency groups: Focusing on imagery instruction and context imageability. *ARELE: annual review of English language education in Japan*, *24*, 1–16. Retrieved from http://ci.nii.ac.jp/naid/110009806688

Hijikata, Y. (2012). *The chunking process of Japanese EFL readers focusing on verb bias, L2 reading proficiency, and working memory*. Saarbrücken, Germany: Lambert Academic Publishing.

Horiba, Y. (2000). Reader control in reading: Effects of language competence, text type, and task. *Discourse Processes*, *29*, 223–267. doi:10.1207/S15326950dp2903_3

Horiba, Y. (2012). Word knowledge and its relation to text comprehension: A comparative study of Chinese- and Korean-speaking L2 learners and L1 speakers of Japanese. *The Modern Language Journal*, *96*, 108–121. doi:10.1111/j.1540-4781.2012.01280.x

Horiba, Y. (2013). Task-induced strategic processing in L2 text comprehension. *Reading in a Foreign Language*, *25*, 98–125. Retrieved from http://nflrc.hawaii.edu/rfl/October2013/articles/horiba.pdf

Horiba, Y., & Fukaya, K. (2006). The effects of task conditions on English reading and incidental vocabulary learning. *Bulletin of St. Luke’s College of Nursing*, *32*, 11–19. Retrieved from http://hdl.handle.net/10285/485

Horiba, Y., & Fukaya, K. (2012). Effects of task instructions on text processing and learning in a Japanese EFL college nursing setting. In A. Shehadeh & C. A. Coombe (Eds.), *Task-based language teaching in foreign language contexts:*

Research and implementation (pp. 89–107). Amsterdam, the Netherland: John Benjamins.

Horiba, Y., & Fukaya, K. (2015). Reading and learning from L2 text: Effects of reading goal, topic familiarity, and language proficiency. *Reading in a Foreign Language*, *27*, 22–46. Retrieved from http://nflrc.hawaii.edu/rfl/April2015/articles/horiba.pdf

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*, 207–223. Retrieved from http://nflrc.hawaii.edu/rfl/PastIssues/rfl112horst.pdf

Hu, M. H., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*, 403–430. Retrieved from http://nflrc.hawaii.edu/rfl/PastIssues/rfl131hsuehchao.pdf

Huang, S., Eslami, Z., & Willson, V. (2012). The effects of task involvement load on L2 incidental vocabulary learning: A meta-analytic study. *The Modern Language Journal*, *96*, 544–557. doi:10.1111/j.1540-4781.2012.01394.x

Huckin, T., & Bloch, J. (1993). Strategies for inferring word-meanings in context: A cognitive model. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 153–176). Norwood, NJ: Ablex.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 113–125). London, England: Macmillan.

Hulstijn, J. H. (1993). When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *The Modern Language Journal*, *77*, 139–147. doi:10.1111/j.1540-4781.1993.tb01957.x

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition*, *27*, 129–140. doi:10.1017/S0272263105050084.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, *80*, 327–339. doi:10.1111/j.1540-4781.1996.tb01614.x

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *51*, 539–558. doi:10.1111/0023-8333.00164

Inohara, K., & Kusumi, T. (2011). Psychological validity of conceptual similarities based on latent semantic analysis: Merits and limitations in statistical analysis of corpuses. *Japanese Psychological Review*, *54*, 101–122. Retrieved from http://ci.nii.ac.jp/naid/40019161241

Inohara, K., & Kusumi, T. (2012). Effects of reading habits on vocabulary: Examination by latent semantic analysis. *Cognitive Studies*, *19*, 100–121. Retrieved from https://www.jstage.jst.go.jp/article/jcss/19/1/19_100/_pdf

JACET Basic Word Revision Committee (Ed.). (2003). *JACET list of 8000 basic words*. Tokyo, Japan: Author.

Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. van Patten (Eds.), *Research methods in second language psycholinguistics* (pp. 20–49). New York, NY: Routledge.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*, 160–212. doi:10.1111/lang.12034

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*, 47–77. doi:10.1093/applin/21.1.47

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, *24*, 617–637. doi:10.1017.S0272263102004047

Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal*, *88*, 416–432. doi:10.1111/j.0026-7902.2004.00238.x

Jiang, N. (2012). *Conducting reaction time research in second language studies*. London, England: Routledge.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354. doi:10.1037/0033-295X.87.4.329

Just, M. A., & Carpenter, P. A. (1992). A capacity of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149. doi:10.1037/0033-295X.99.1.122

Kappenman, E. S., & Luck, S. J. (2012). ERP components: The ups and downs of brainwave recordings. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 3–30). Oxford University Press.

Kida, S. (2010). The role of processing-resource allocation in incidental L2 vocabulary learning through reading. *ARELE: annual review of English language education in Japan*, *21*, 171–180. Retrieved from http://ci.nii.ac.jp/naid/110008512407

Kidd, E., Lieven, E. V. M., & Tomasello, M. (2010). Lexical frequency and exemplar-based learning effects in language acquisition: Evidence from sentential complements. *Language Sciences*, *32*, 132–142. doi:10.1016/j.langsci.2009.05.002

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *92*, 163–182. doi:10.1037/0033-295X.95.2.163

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*, 294–303. doi:10.1037/0003-066X.49.4.294

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, *29*, 133–159. doi:10.1016/0749-596X(90)90069-C

Klin, C. M. (1995). Causal inferences in reading: From immediate activation to long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1483–1494. doi:10.1037/0278-7393.21.6.1483

Klin, C. M., Murray, J. D., Levine, W. H., & Guzmán, A. E. (1999). Forward inferences: From activation to long-term memory. *Discourse Processes*, *27*, 241–260. doi:10.1080/01638539909545062

Koda, K. (2005) *Insights into second language reading: A cross-linguistic approach*. Cambridge University Press.

Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research*, *4*, 900–913. doi:10.4304/jltr.4.5.900-913

Krashen, S. D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal*, *73*, 440–464. doi: 10.1111/j.1540-4781.1989.tb05325.x

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. doi:10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205. doi:10.1126/science.7350657

Kutas, M., & van Petten, C. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). San Diego, CA: Academic Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240. doi:10.1037/0033-295X.104.2.211

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, *25*, 259–284. doi:10.1080/01638539809545028

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). What is latent semantic analysis? In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. ix–xii). Mahwah, NJ: Lawrence Erlbaum Associates.

Langacker, R. W. (2008). Cognitive grammar as a basis for language instruction. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 66–88). New York, NY: Routledge.

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. New York, NY: Routledge.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20–34). Cambridge University Press.

Laufer, B. (2006). Comparing focus on form and focus on forms in second-language vocabulary learning. *The Canadian Modern Language Review*, *63*, 149–166. doi: 10.3138/cmlr.63.1.149

Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*, 1–26. doi:10.1093/applin/22.1.1

Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal*, *28*, 89–108. doi:10.1177/003368829702800106

Leung, J. H. C., & Williams, J. H. (2011). The implicit learning of mappings between forms and contextual derived meanings. *Studies in Second Language Acquisition*, *33*, 33–55. doi:10.1017/S0272263110000525

Leung, J. H. C., & Williams, J. H. (2012). Constraints on implicit learning of grammatical form-meaning connections. *Language Learning*, *62*, 634–662. doi:10.1111/j.1467-9922.2011.00637.x

Li, X. (1988). Effects of contextual cues on inferring and remembering meanings of new words. *Applied Linguistics*, *9*, 402–413. doi:10.1093/applin/9.4.402

Love, J., McKoon, G., & Gerrig, R. J. (2010). Searching for Judy: How small mysteries affect narrative processes and memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 790–796. doi:10.1037/a0018989

Ma, T., Chen, B., Lu, C., & Dunlap, S. (2015). Proficiency and sentence constraint effects on second language word learning. *Acta Psychologica*, *159*, 116–122. doi:10.1016/j.actpsy.2015.05.014

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, *91*, 615–629. doi:10.1037/0022-0663.91.4.615

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*, 574–588. doi:10.1016/j.jml.2013.08.001

McKeown, M. G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly*, *20*, 482–496. doi:10.2307/747855

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*, 440–460. doi:10.1037/0033-295X.99.3.440

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1–43. doi:10.1207/s1532690xci1401_1

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). *Five papers on WordNet*. Princeton University, Cognitive Science Laboratory.

Mizumoto, A. (2006). Exploring the relationship between the use of vocabulary learning strategies and learning outcomes. *Konan Women's University Researches*, *42*, 93–107. Retrieved from http://ci.nii.ac.jp/naid/110004868193

Mizumoto, A., & Takeuchi, O. (2009). Examining the effectiveness of explicit instruction of vocabulary learning strategies with Japanese EFL university students. *Language Teaching Research*, *13*, 425–449. doi:10.1177/1362168809341511

Mondria, J.-A., & Wit-de Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics*, *12*, 249–267. doi:10.1093/applin/12.3.249

Morgan-Short, K. (2014). Electrophysiological approaches to understanding second language acquisition: A field reaching its potential. *Annual Review of Applied Linguistics*, *34*, 15–36. doi:10.1017/S026719051400004X

Morgan-Short, K., & Tanner, D. (2014). Event-related potentials (ERPs). In J. Jegerski & B. van Patten (Eds.), *Research methods in second language psycholinguistics* (pp. 127–152). New York, NY: Routledge.

Morimoto, Y. (2006). Concreteness effect on L2 learners' lexical inferencing. *ARELE: annual review of English language education in Japan*, *17*, 141–150. Retrieved from http://ci.nii.ac.jp/naid/110008512050

Muramoto, T. (2000). Daini gengo no bunshou rikai katei ni oyobosu shujukudo no eikyo: Bun sainin kadai niyoru jikken [The effects of second-language proficiency on text comprehension]. *The Science of Reading*, *44*, 43–50. Retrieved from http://ci.nii.ac.jp/naid/40002693398

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, *24*, 237–270. doi:10.3102/00028312024002237

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from contexts. *Reading Research Quarterly*, *20*, 233–253.

Nahatame, S. (2012). Latent semantic analysis (LSA) ni yoru kusho-hoju-gata dokkai test no kaimei: Bun level no imiteki-kanrendo wo kanten to shite [Investigating a cloze test using latent semantic analysis]. *STEP Bulletin*, *24*, 42–58. Retrieved from http://www.eiken.or.jp/center_for_research/pdf/bulletin/vol24/vol_24_p42-p58.pdf

Nahatame, S. (2015). *Making and revising predictive inferences in Japanese EFL learners' reading comprehension* (Doctor dissertation, University of Tsukuba).

Nakagawa, C. (2006). The effects of morphological and contextual clues on EFL readers' lexical inference. *ARELE: annual review of English language education in Japan*, *17*, 151–160. Retrieved from http://ci.nii.ac.jp/naid/110008512303/en

Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, *37*, 645–670. doi:10.2307/3588216

Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Modern Language Journal*, *90*, 387–401. doi:10.1111/j.1540-4781.2006.00431.x

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle Cengage Learning.

Nittono, H. (2005). *Shinrigaku no tameno jishou kanren den-i guidebook* [Event-related potentials guidebook for psychology]. Kyoto, Japan: Kitaohji Shobou.

Nordlie, J. W., Dopkins, S., & Johnson, M. (2001). Words in a sentence become less accessible when an anaphor is resolved. *Memory & Cognition*, *29*, 201–220. doi: 10.3758/BF03194930

Obunsha. (2010a). *Eiken jyun ichi kyu zen mondai-syu* [The collection of previous questions for the pre-first grade Eiken test]. Tokyo, Japan: Author.

Obunsha. (2010b). *Eiken ni kyu zen mondai-syu* [The collection of previous questions for the second grade Eiken test]. Tokyo, Japan: Author.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, *17*, 1212–1228. doi:10.1162/0898929055002436

Otten, M., & van Berkum, J. J. A. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, *45*, 464–496. doi:10.1080/01638530802356463

Otten, M., & van Berkum, J. J. A. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research*, *1291*, 92–101. doi:10.1016/j.brainres.2009.07.042

Paribakht, T. S., & Wesche, M. B. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp.174–200). Cambridge University Press.

Paribakht, T. S., & Wesche, M. B. (1999). Reading and “incidental” L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition*, *21*, 195–224.

Pellicer-Sánchez, A. (2015). Incidental L2 vocabulary acquisition *from* and *while* reading: An eye-tracking study. *Studies in Second Language Acquisition*. Advance online publication. doi:10.1017/S0272263115000224

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–383. doi:10.1080/10888430701530730

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1281–1292. doi: 10.1037/0278-7393.31.6.1281

Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquires. *Reading in a Foreign Language*, *5*, 271–275. Retrieved from http://nflrc.hawaii.edu/rfl/PastIssues/rfl52pitts.pdf

Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. doi:10.1111/lang.12079

Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *The Modern Language Journal*, *80*, 478–493. doi:10.1111/j.1540-4781.1996.tb05468.x

Pulido, D. (2007). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, *57*, 155–199. doi:10.1111/j.0023-8333.2004.00263.x

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Journal*, *56*, 282–308. doi: 10.3138/cmlr.56.2.282

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, *52*, 513–536. doi:10.1111/1467-9922.00193

Rapp, D. N. (2008). How do readers handle incorrect information during reading? *Memory & Cognition*, *36*, 688–701. doi:10.3758/MC.36.3.688

Rapp, D. N., & Kendeou, P. (2007). Revising what readers know: Updating text representations during narrative comprehension. *Memory & Cognition*, *35*, 2019–2032. doi:10.3758/BF03192934

Rapp, D. N., & Kendeou, P. (2009). Noticing and revising discrepancies as texts unfold. *Discourse Processes*, *46*, 1–24. doi:10.1080/01638530802629141

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*, 1457–1506. doi: 10.1080/17470210902816461

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z reader model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 720–732. doi:10.1037/0096-1523.30.4.720

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*, 595–626. doi:10.1111/lang.12010

Robinson, P. (1995). Attention, memory, and the “noticing” hypothesis. *Language Learning*, *45*, 283–331. doi:10.1111/j.1467-1770.1995.tb00441.x

Robinson, P. (2011). Task-based language learning: A review of issues. *Language Learning*, *61*, 1–36. doi:10.1111/j.1467-9922.2011.00641.x

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439. doi:10.1016/0010-0285(76)90013-X

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, *21*, 589–619.

Rott, S. (2005). Processing glosses: A qualitative exploration of how form-meaning connections are established and strengthened. *Reading in a Foreign Language*, *17*, 95–124. Retrieved from http://nflrc.hawaii.edu/rfl/October2005/rott/rott.pdf

Rott, S. (2007). The effect of frequency of input-enhancements on word learning and text comprehension. *Language Learning*, *57*, 165–199. doi:10.1111/j.1467-9922.2007.00406.x

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158. doi:10.1093/applin/11.2.129

Schmidt, R. W. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165–209). London, England: Academic Press.

Schmidt, R. W. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. W. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). University of Hawai‘i Press.

Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. London, England: Palgrave Macmillan.

Schoonen, R., van Gelderen, A., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning*, *53*, 165–202. doi:10.1111/1467-9922.00213

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, *31*, 577–607. doi:10.1017/S0272263109990039

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-related ERP components. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 397–439). Oxford University Press.

Swanborn, M. S. L., & de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, *69*, 261–285. doi:10.3102/003465 43069003261

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, *68*, 297–314. doi:10.1016/j.jml.2012.12.002

Till, R. E., Mross, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory & Cognition*, *16*, 283–298. doi:10. 3758/BF03197039

Tomasello, M. (2000a). Do young children have adult syntactic competence? *Cognition*, *74*, 209–253. doi:10.1016/S0010-0277(99)00069-4

Tomasello, M. (2000b). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, *11*, 61–82. doi:10.1515/cogl.2001.012

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tyler, A. (2010). Usage-based approaches to language and their applications to second language learning. *Annual Review of Applied Linguistics, 30*, 270–291. doi:10.1017/S0267190510000140

Ushiro, Y. (2010). Active role of the reader in the comprehension process. In H. Kimura, T. Kimura, & O. Shiki (Eds.), *Theory and practice in reading and writing: Nurturing independent learning* (pp. 44–58). Tokyo, Japan: Taishukan Shoten.

Ushiro, Y., Hamada, A., Hasegawa, Y., Nahatame, S., Kimura, Y., Shimizu, H., Takaki, S., Kobayashi, M., & Tanaka, N. (2013). Discourse-based lexical inferencing in EFL reading: Focusing on depth of vocabulary knowledge and cue availability. *ARELE: annual review of English language education in Japan*, *24*, 77–92. Retrieved from http://ci.nii.ac.jp/naid/110009806693

Ushiro, Y., Hamada, A., Kimura, Y., Nahatame, S., Hosoda, M., Kato, D., & Watanabe, Y. (2015). Building causally coherent mental representations and learning from expository texts among Japanese EFL readers. *JACET Journal*, *59*, 131–150. Retrieved from http://ci.nii.ac.jp/naid/110009908031

Ushiro, Y., Nahatame, S., Hasegawa, Y., Shimizu, H., Kimura, Y., Tanaka, Y., & Nakagawa, C. (2012). Activation and encoding of bridging and predictive inferences in EFL reading comprehension. *JACET Journal*, *54*, 33–52. Retrieved from http://ci.nii.ac.jp/naid/110009436822

van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, *64*, 88–107. doi:10.1016/j.jml.2010.08.006

van Berkum, J. J. A., Brown, C. M., Zwiserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 443–467. doi:10.1037/0278-7393.31.3.443

van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science*, *328*, 453–456. doi:10.1126/science.1182594

van den Broek, P., Fletcher, C. R., & Risden, K. (1993). Investigations of inferential processes in reading: A theoretical and methodological integration. *Discourse Processes*, *16*, 169–180. doi:10.1080/01638539309544835

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.

van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Simis, A., & Snellings, P. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first- and second-language reading comprehension: A componential analysis. *Journal of Educational Psychology*, *96*, 19–30. doi:10.1037/0022-0663.96.1.19

van Patten, B., Williams, J., Rott, S., & Overstreet, M. (2004). *Form-meaning connections in second language acquisition*. New York, NY: Routledge.

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*, 457–479. doi:10.1093/applin/ams074

Wang, H. C., Pomplun, M., Ko, H. W., Chen, M. L., & Rayner, K. (2010). Estimating the effect of word predictability on eye movements in Chinese reading using latent semantic analysis and transitional probability. *The Quarterly Journal of Experimental Psychology*, *63*, 1374–1386. doi:10.1080/17470210903380814

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*, 130–163. Retrieved from http://nflrc.hawaii.edu/rfl/October2003/waring/waring.pdf

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*, 33–52. doi:10.1017/S0272263105050023

Webb, S. (2007a). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research*, *11*, 63–81. doi:10.1177/1362168806072463

Webb, S. (2007b). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*, 46–65. doi:10.1093/applin/aml048

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*, 232–245. Retrieved from http://nflrc.hawaii.edu/rfl/October2008/webb/webb.pdf

Wesche, M. B., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, *53*, 13–40.

Wesche, M. B., & Paribakht, T. S. (2000). Reading-based exercises in second language vocabulary learning: An introspective study. *The Modern Language Journal*, *84*, 196–213. doi:10.1111/0026-7902.00062

Wesche, M. B., & Paribakht, T. S. (2010). *Lexical inferencing in a first and second language: Cross-linguistic dimensions*. Bristol, England: Multilingual Matters.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*, 1272–1288. doi:10.1162/0898929041920487

Williams, R., & Morris, R. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, *16*, 312–339. doi:10.1080/09541440340000196

Wolfe, M. B. W., Magliano, J. P., & Larsen, B. (2005). Causal and semantic relatedness in discourse understanding and representation. *Discourse Processes*, *39*, 165–187. doi:10.1080/0163853X.2005.9651678

Wolter, B. (2006). Lexical network structures and L2 vocabulary acquisition: The role of L1 lexical/conceptual knowledge. *Applied Linguistics*, *27*, 741–747. doi:10.1093/applin/aml036

Yamashita, J., & Shiotsu, T. (2015). Comprehension and knowledge components that predict L2 reading: A latent-trait approach. *Applied Linguistics*. Advance online publication. doi:10.1093/applin/amu079

Appendices

Appendix 1

Directions for L2 Reading Proficiency Tests

読解能力測定テスト

◇注意事項◇

① 試験開始の指示があるまで，問題冊子を開かないで下さい。

② テストは全部で大問 6 題から成り，合計 26 問で解答時間は 30 分です。

③ 問題用紙に書き込みをすることは構いません。

④ 辞書を使用することはできません。

⑤ 試験終了後，問題冊子を回収しますので，必ず名前を書いてから解答を始めて下さい。

所属　　　　　　学年　　　　氏名

Appendix 2

The Profile of L2 Reading Proficiency Tests Adapted From Obunsha (2010a, 2010b)

Text title	Grade	Words	FKGL	FRE
The Study of Latin	pre 1st	302	13.4	44
The Celtic Tiger	pre 1st	305	12.2	41
News Reporting in the U.S.	2nd	287	11.0	46
Sources of Intelligence	2nd	279	10.5	51
The Magic of Touch	2nd	279	9.9	55
Green Belts	2nd	307	9.5	61

Note. FKGL = Flesch-Kincaid Grade level. FRE = Flesch Reading Ease.

Appendix 3

Learning Sentences With LSA Values Used in Experiment 1

No.	HSS learning sentences	LSA
1	The zoo's rules did not allow visitors to give food to the **animals**.	.37
2	The minister was asked about his problem, but he never gave a direct **answer**.	.38
3	The soccer player strongly kicked the **ball**.	.69
4	The baseball player is practicing swinging the **bat**.	.70
5	Some children are in the sea, while others are playing with the sand on the **beach**.	.47
6	He felt cold and covered his knees with a **blanket**.	.48
7	He went to the library to get a **book**.	.41
8	A high-speed elevator was installed in that tall **building**.	.37
9	His coat was open because it was missing a **button**.	.43
10	If Tom wants to free the birds, he has to open the iron **cage**.	.37
11	George taught his son to drive a **car**.	.29
12	*Peter Rabbit* is a picture book for little **children**.	.22
13	Mrs. Carol said that Dennis was the loudest student in the **class**.	.26
14	They didn't know what time it was because they couldn't find a **clock**.	.24
15	The sun is no longer shining because it is hidden by a big **cloud**.	.33
16	As the weather had got warmer, Mark wore his jacket instead of his heavy winter **coat**.	.55
17	I installed new word processing software on my **computer**.	.30
18	On top of his head, the king wore an extremely expensive **crown**.	.47
19	The jeweler tried to determine whether it was a real or fake **diamond**.	.23
20	A faithful animal is not a cat but a **dog**.	.34
21	Always knock before you open my **door**.	.49
22	The princess wore a beautiful wedding **dress**.	.62
23	He couldn't hear well because of his illness in his left **ear**.	.34
24	She put contact lenses on her **eyes**.	.36
25	My uncle keeps many cows and pigs on his **farm**.	.35
26	Red and white color are used in the Japanese national **flag**.	.37
27	A kind of green animal that jumps around in a lake is a **frog**.	.26
28	After the hurricane, the area around the window was filled with broken **glass**.	.36
29	He tried to put the pieces of the broken plate back together with **glue**.	.35
30	To play rock 'n' roll at the festival, he brought his electric **guitar**.	.28
31	The bank robber aimed at the security officer and shot the **gun**.	.37
32	The pretty dancer wears a rose in her blond **hair**.	.37

(continued)

No.	HSS learning sentences	LSA
33	The fisherman attached the worm to the **hook**.	.48
34	Tom’s dog brought many tree branches into his **house**.	.29
35	He couldn’t unlock the door without the correct **key**.	.28
36	The skier fell and broke his **leg**.	.39
37	When you leave the room, don’t forget to turn off the **lights**.	.37
38	The musician held a charity concert to collect a lot of **money**.	.38
39	The astronauts of Apollo 11 landed on the **moon**.	.69
40	The climber planned to camp at the top of the **mountain**.	.40
41	Jim put a little cheese in the trap to catch the **mouse**.	.20
42	He opened a bottle of wine and listened to classical **music**.	.23
43	In the evening, my husband always watches the seven o’clock **news**.	.18
44	People listen with their ears and smell with their **nose**.	.49
45	In caring for patients, a doctor often needs the help of a **nurse**.	.61
46	Paul couldn’t write the message because he didn’t have a pencil or a piece of **paper**.	.38
47	To mark the answer sheet, the students needed a sharp **pencil**.	.37
48	The meal would taste better if you added some salt and **pepper**.	.35
49	Someone who flies an airplane is called a **pilot**.	.37
50	If you see a crime on the street, you have to call the **police**.	.39
51	For Halloween, they decorated an orange **pumpkin**.	.41
52	In Europe, Elizabeth II of Great Britain is the most famous **queen**.	.34
53	She put up an umbrella because of a sudden **rain**.	.21
54	He bought his girlfriend an engagement **ring**.	.32
55	Every morning, most children in Japan go to **school**.	.39
56	The rescue team found signs of the lost boy, so they continued the **search**.	.43
57	He found that a rainbow had appeared in the **sky**.	.29
58	That train always arrives on time at the **station**.	.51
59	There were long lines for the movie at the new **theater**.	.25
60	When it is stormy outside, you can see the lightning and hear the **thunder**.	.42
61	After taking a quick shower, she wiped her face with a **towel**.	.31
62	About seventy-five percent of the surface of the earth is covered by **water**.	.38
63	The referee stopped the football game by blowing his **whistle**.	.52
64	The poor bird couldn’t fly away with its broken **wing**.	.63

(continued)

No.	LSS learning sentences	LSA
1	She was quite surprised that Bill didn't keep his **animals**.	.14
2	As usual, he didn't do well because he had not prepared his **answer**.	.34
3	The violent monkey strongly kicked the **ball**.	.22
4	He learned the method for making wooden **bats**.	.13
5	Before the fireworks started, a lot of children sat and waited on the **beach**.	.23
6	On cold days, they provided each staff member with an extra **blanket**.	.35
7	He went to the convenience store to buy a **book**.	.14
8	He was packing all of his books and moving to that **building**.	.25
9	The woman wore a long coat with a fancy **button**.	.33
10	If Tom wants to see those animals, he has to find the special **cage**.	.32
11	The TV commercial was for a new **car**.	.17
12	She bought several colorful postcards to send to her **children**.	.21
13	No one wanted to make any decisions until after **class**.	.21
14	The class was so boring that all he could do was watch the **clock**.	.27
15	The drawing is not yet finished because he is still working on the **cloud**.	.18
16	She felt cold, so she reached for her **coat**.	.33
17	His seat in the small classroom was next to the **computer**.	.08
18	The Museum of London always displayed the extremely expensive **crown**.	.29
19	They installed a very good alarm system to protect their **diamonds**.	.17
20	At the park, she saw a child with a **dog**.	.15
21	The girl walked slowly toward the **door**.	.42
22	The girl wanted to buy a pretty **dress**.	.48
23	I'd like you to cut my hair just below the **ears**.	.21
24	I know the little dog with pretty **eyes**.	.25
25	My uncle did not pay much for this **farm**.	.23
26	The main streets were decorated with many red and white **flags**.	.43
27	A kind of small animal that lives in a lake is a **frog**.	.24
28	I could hear the loud telephone ringing through the **glass**.	.24
29	She walked across the large room to Mike's dirty desk and returned his **glue**.	.14
30	As a punishment, his parents immediately took away his **guitar**.	.15
31	Two security officers approached the strange car with their **guns**.	.28
32	The careless waiter spilled wine on her beautiful **hair**.	.30

(continued)

No.	LSS learning sentences	LSA
33	He found that the fish had swallowed the **hook**.	.61
34	She left the little baby boy alone in the **house**.	.37
35	In the mysterious box, they found a small **key**.	.22
36	He got burned on his right **leg**.	.38
37	When he approached the village, he found that none of the houses had **lights**.	.27
38	I want my children to learn how to manage their own **money**.	.25
39	The couple enjoyed looking at the shining **moon**.	.20
40	The boy painted the view from the **mountain**.	.18
41	He was so angry that he finally bought some poison to kill the **mouse**.	.22
42	She finally decided to go to college to study **music**.	.14
43	In the evening, my husband always records the **news**.	.18
44	My girlfriend said that she hated her **nose**.	.30
45	When she came back from the restroom, she found him talking with a **nurse**.	.23
46	You should write your name at the bottom of the **paper**.	.32
47	The teacher said that the students needed some **pencils**.	.25
48	The meal would taste better if you added some tomatoes and **pepper**.	.35
49	The young man sitting over there is a **pilot**.	.20
50	If you want a job, you can join the **police**.	.19
51	For her birthday, they made a cake out of the **pumpkin**.	.37
52	In Europe, that proud lady is the most famous **queen**.	.30
53	She wore heavy socks and boots to go walking in the **rain**.	.22
54	The jeweler put a diamond on the **ring**.	.39
55	The mayor is going to build a new **school**.	.30
56	Cathy is the firm’s newest clerk, and she always works overtime to finish her **search**.	.22
57	Before going fishing, the man always checked the **sky**.	.26
58	My friend sometimes arrives too late at the **station**.	.26
59	For a long time, the hand clapping could be heard throughout the **theater**.	.17
60	When she was standing outside, she could see the house and hear the **thunder**.	.34
61	While standing in line at the hotel desk, the young boy asked for another **towel**.	.26
62	They have to depend on the small river for their **water**.	.37
63	The end of the game was marked by the sound of a **whistle**.	.61
64	The scientist couldn’t take care of the broken **wing**.	.18

Note. Target words are bolded.

Appendix 4

Probes Used in the Plausibility Judgment Test in Experiment 1

No.	Plausible sentences	Rating	Implausible sentences	Rating
1	They killed the **animals**.	100%	They planted the **animals**.	94%
	He hunted the **animals**.	97%	She planned the **animals**.	84%
2	She knew the **answer**.	100%	They sat on the **answer**.	94%
	He had the **answer**.	88%	He visited the **answer**.	100%
3	He missed the **ball**.	84%	I started the **ball**.	97%
	She caught the **ball**.	91%	We caused the **ball**.	100%
4	I gripped the **bat**.	84%	They performed the **bat**.	88%
	He threw the **bat**.	91%	She respected the **bat**.	78%
5	They cleaned the **beach**.	97%	She threw her **beach**.	97%
	They lay on the **beach**.	100%	They drank the **beach**.	94%
6	They cleaned the **blanket**.	84%	She respected her **blanket**.	91%
	They moved the **blanket**.	91%	She told the **blanket**.	84%
7	He borrowed the **book**.	100%	He grew the **book**.	100%
	He wrote the **book**.	97%	He attended the **book**.	91%
8	She designed the **building**.	94%	He injured his **building**.	84%
	He destroyed the **building**.	94%	He met the **building**.	75%
9	He pushed the **button**.	97%	She visited the **button**.	100%
	He pressed the **button**.	75%	She invited the **button**.	100%
10	She built the **cage**.	75%	They fulfilled the **cage**.	97%
	He broke the **cage**.	97%	We paid the **cage**.	75%
11	She parked her **car**.	100%	They answered the **car**.	75%
	She crashed her **car**.	97%	He grew the **car**.	81%
12	She educated the **children**.	97%	He stored the **children**.	78%
	She trained her **children**.	94%	He wore his **children**.	100%
13	They began the **class**.	75%	She purchased the **class**.	91%
	He attended the **class**.	97%	He repaired the **class**.	91%
14	He fixed the **clock**.	75%	They planned the **clock**.	97%
	He repaired the **clock**.	100%	They cooked the **clock**.	100%
15	She watched the **cloud**.	88%	We picked the **cloud**.	97%
	She photographed the **cloud**.	75%	We talked to the **cloud**.	91%
16	She needed the **coat**.	97%	He went to the **coat**.	91%
	She bought the **coat**.	100%	He planted the **coat**.	100%

(continued)

No.	Plausible sentences	Rating	Implausible sentences	Rating
17	He moved the **computer**.	88%	She surprised the **computer**.	78%
	She used the **computer**.	100%	He greeted the **computer**.	100%
18	She received the **crown**.	88%	He believed the **crown**.	91%
	She lost her **crown**.	84%	He regretted the **crown**.	94%
19	They displayed the **diamond**.	100%	He taught the **diamond**.	100%
	He sold the **diamond**.	97%	They contacted the **diamond**.	91%
20	He played with the **dog**.	91%	They entered the **dog**.	97%
	He took care of the **dog**.	97%	She climbed the **dog**.	100%
21	They closed the **door**.	100%	She judged the **door**.	97%
	He repaired the **door**.	97%	He answered the **door**.	88%
22	She changed her **dress**.	84%	He drank the **dress**.	100%
	She made the **dress**.	75%	He served the **dress**.	100%
23	She doubted her **ear**.	75%	He bought the **ear**.	97%
	She cleaned her **ear**.	91%	He made her **ear**.	100%
24	She washed her **eyes**.	91%	He visited the **eye**.	75%
	She closed her **eyes**.	91%	They continued the **eye**.	91%
25	He managed his **farm**.	97%	We rode the **farm**.	100%
	I sold my **farm**.	84%	She dried the **farm**.	84%
26	They folded the **flag**.	88%	She read the **flag**.	84%
	We raised the **flag**.	84%	He tasted the **flag**.	100%
27	She disliked the **frog**.	100%	They improved the **frog**.	94%
	He caught the **frog**.	91%	They ruled the **frog**.	81%
28	They wiped the **glass**.	84%	He enjoyed the **glass**.	94%
	He broke the **glass**.	91%	He answered the **glass**.	75%
29	She used the **glue**.	94%	She grew the **glue**.	97%
	She spread the **glue**.	78%	He shot the **glue**.	91%
30	He lost his **guitar**.	100%	She asked the **guitar**.	75%
	He practiced the **guitar**.	100%	She spent the **guitar**.	94%
31	He carried the **gun**.	75%	She called the **gun**.	100%
	She handled the **gun**.	88%	We cooked the **gun**.	100%
32	I brushed my **hair**.	88%	She killed the **hair**.	97%
	She set her **hair**.	91%	We charged our **hair**.	84%

(continued)

No.	Plausible sentences	Rating	Implausible sentences	Rating
33	He removed the **hook**.	75%	She drove the **hook**.	97%
	They used the **hook**.	75%	She visited the **hook**.	100%
34	She left the **house**.	94%	He picked up the **house**.	78%
	He walked to the **house**.	100%	He followed the **house**.	88%
35	He inserted the **key**.	75%	She danced the **key**.	78%
	She copied the **key**.	81%	He excited the **key**.	94%
36	She crossed her **legs**.	100%	He hired the **leg**.	100%
	She lost her **leg**.	88%	They thanked the **leg**.	75%
37	They displayed the **lights**.	75%	She tasted the **lights**.	78%
	They created the **lights**.	75%	He informed the **lights**.	91%
38	We borrowed some **money**.	97%	We booked some **money**.	91%
	I earned my **money**.	97%	They visited some **money**.	100%
39	They approached the **moon**.	78%	He threw the **moon**.	100%
	She observed the **moon**.	81%	He received the **moon**.	100%
40	She walked through the **mountain**.	81%	He washed the **mountain**.	100%
	We visited the **mountain**.	94%	He sent the **mountain**.	94%
41	He hated the **mouse**.	88%	They mapped the **mouse**.	100%
	He disliked the **mouse**.	91%	He started the **mouse**.	100%
42	We performed some **music**.	88%	They kicked some **music**.	100%
	I played some **music**.	91%	They closed some **music**.	100%
43	They gathered the **news**.	78%	He burned the **news**.	81%
	She announced the **news**.	75%	He hurt the **news**.	91%
44	He pressed his **nose**.	75%	They greeted the **nose**.	100%
	He scratched his **nose**.	84%	He practiced the **nose**.	88%
45	They asked the **nurse**.	81%	They recycled the **nurse**.	100%
	She thanked the **nurse**.	88%	They cleaned the **nurse**.	94%
46	They recycled the **paper**.	100%	She looked after her **paper**.	97%
	They cut the **paper**.	97%	He laughed at the **paper**.	97%
47	She sharpened her **pencil**.	88%	He canceled the **pencil**.	97%
	She gripped her **pencil**.	100%	He runs the **pencil**.	91%
48	She bought the **pepper**.	100%	We posted the **pepper**.	88%
	She used the **pepper**.	100%	I supported the **pepper**.	94%

(continued)

No.	Plausible sentences	Rating	Implausible sentences	Rating
49	He supported the **pilot**.	94%	She signed the **pilot**.	81%
	We needed the **pilot**.	88%	He lived in the **pilot**.	100%
50	He avoided the **police**.	78%	They shared the **police**.	94%
	I helped the **police**.	97%	She studied the **police**.	78%
51	They cut the **pumpkin**.	100%	We answered the **pumpkin**.	97%
	He grew the **pumpkin**.	94%	I poured the **pumpkin**.	94%
52	We feared the **queen**.	91%	She read the **queen**.	100%
	They admired the **queen**.	88%	She replayed the **queen**.	97%
53	We needed some **rain**.	84%	I went to the **rain**.	97%
	He enjoyed the **rain**.	75%	She cut some **rain**.	100%
54	We exchanged the **rings**.	84%	He wrote the **ring**.	97%
	They wore the **ring**.	78%	I heard the **ring**.	78%
55	She attended the **school**.	81%	He used his **school**.	91%
	He finished the **school**.	75%	She threw the **school**.	100%
56	She started the **search**.	84%	They photographed the **search**.	84%
	She completed the **search**.	81%	He touched the **search**.	91%
57	He watched the **sky**.	84%	They dropped the **sky**.	75%
	He looked at the **sky**.	97%	He read the **sky**.	91%
58	We built the **station**.	97%	He invited the **station**.	91%
	They destroyed the **station**.	97%	He offered the **station**.	91%
59	They built the **theater**.	88%	He played with the **theater**.	91%
	He entered the **theater**.	97%	He picked up the **theater**.	88%
60	He feared the **thunder**.	100%	He began the **thunder**.	97%
	He noticed the **thunder**.	97%	He practiced the **thunder**.	100%
61	He washed the **towel**.	94%	He worked his **towel**.	78%
	He picked up the **towel**.	88%	He greeted the **towel**.	100%
62	He saved some **water**.	88%	He hit some **water**.	88%
	He drank some **water**.	91%	She completed some **water**.	100%
63	He heard the **whistle**.	88%	They traveled the **whistle**.	100%
	He touched the **whistle**.	78%	She followed the **whistle**.	75%
64	It had its **wings**.	100%	I replayed the **wing**.	91%
	It spread its **wings**.	97%	We waited for the **wing**.	88%

Note. The rating scores mean the (im)plausibility of each probe sentence.

Appendix 5

Learning Sentences and Target Words Used in Experiments 2A and 2B

No.		Learning sentences and target words	LSA
		Forward inference condition	
1	HSS	George taught / his son / to drive a *car*.	.29
	LSS	The TV commercial / was for a new *car*.	.17
2	HSS	Always knock / before you open / my *door*.	.49
	LSS	The girl moved slowly / toward the *door*.	.35
3	HSS	To fill in the mark sheet, / the student needed / a sharp *pencil*.	.44
	LSS	The student needed / a *pencil*.	.18
4	HSS	The bank robber / aimed at the security officer / and shot / the *gun*.	.37
	LSS	The robber had / a *gun*.	.15
5	HSS	The little puppy / grew up / to be a big *dog*.	.33
	LSS	At the park / she saw / a man / with a *dog*.	.19
6	HSS	Elizabeth II of Great-Britain / is the most famous *queen* / in Europe.	.34
	LSS	That proud lady / is the famous *queen* / in Europe.	.30
7	HSS	My husband always watches / the seven o'clock *news* / in the evening.	.17
	LSS	My husband always records / the *news* / in the evening.	.09
8	HSS	The pitcher was unable / to throw the ball / because of his broken *arm*.	.36
	LSS	He had / a scratch / on his *arm*.	.34
9	HSS	The surfers were attacked / by a dangerous *shark* / in the sea.	.35
	LSS	The group was surprised / by a large *shark* / in the sea.	.25
10	HSS	The train always arrives on time / at the *station* / in Tokyo.	.53
	LSS	My friend sometimes arrives too late / at the *station* / in Tokyo.	.28
		Backward inference condition	
1	HSS	Joe picked up / the *instrument* / and began / to play a melody.	.36
	LSS	Joe picked up / the *instrument* / and began / to walk home.	.24
2	HSS	Yesterday / the *doctor* examined / the pain / in Jack's stomach.	.74
	LSS	Yesterday / the *doctor* talked to Jane / about a number of issues.	.63
3	HSS	Last night / the *music* was performed beautifully / by the organist.	.71
	LSS	Last night / the *music* was something / most people were not familiar with.	.51
4	HSS	The day, / the *taxi* carried / the tourists / through the city streets.	.32
	LSS	The day / the *taxi* was taken / to town / by the group.	.07

(continued)

No.		Learning sentences and target words	LSA
5	HSS	Nancy finished / the *cocktail* / and asked / the bartender / for another glass.	.23
	LSS	Nancy took / the *cocktail* / and asked / her brother / for another glass.	.12
6	HSS	There was the *cloth* / that Ann's dressmaker made / for the party.	.43
	LSS	There was the *cloth* / that Ann's roommate had given her.	.37
7	HSS	Last night / the *bird* were singing / in the tree / outside my room.	.47
	LSS	Last night / the *bird* were in the yard / outside my room.	.14
8	HSS	Jim knew / that the *boots* would keep / his feet warm / in the snow.	.34
	LSS	Jim knew / that the *boots* would keep him / comfortable / in the winter.	.14
9	HSS	After her *house* was broken / by a thief, / Gloria always locked / the door.	.14
	LSS	After her *house* was checked, / she felt / a little better.	.02
10	HSS	There were *strawberries* / to pick for making jam / in the yard.	.16
	LSS	There were *strawberries* available / for those who wanted them / in the yard.	.11

Note. Target words are italicized. Slashes represent pause-chunks.

Appendix 6

The Segmentation Rules Into Chunks Adapted From Hijikata (2012, p. 38)

Basic rules

1. Sentence pattern [Subject + Verb]: Basically, there are no chunk boundaries.
2. Sentence pattern [Subject + Verb + Complement]: Basically, there are no chunk boundaries.
3. Sentence pattern [Subject + Verb + Object]: Basically, there is a chunk boundary before Object. When Object consists of only one word or two words, Object is connected with S + V; thus, the chunk is "S + V + O."
4. Sentence pattern [Subject + Verb + Indirect Object + Direct Object]: Basically, there is a chunk boundary before Indirect Object. When Indirect Object consists of only one word or two words, Indirect Object is connected with S + V.
5. Sentence pattern [Subject + Verb + Object + Complement]: Basically, there is a chunk boundary before Object. When Object consists of only one word or two words, Object is connected with S + V.

Exceptions

6. Big subjects and objects: Big subjects and objects, which contain three words or over are regarded as one chunk.
7. Punctuations: Basically, punctuations such as period, comma, and colon are signs of chunk boundaries. However, the comma representing "apposition" is an exception.
8. Adverbs:
   (a) Adverbs consisting of one word are regarded as one chunk if there is a comma after adverb.
   (b) Adverbs composed of more than two words are chunks by themselves.
   (c) Adverbs consisting of one word are connected to either the former chunk or the latter chunk, if the adverb is embedded in sentences.
9. Prepositions:
   (a) Prepositional phrases are independent chunks if the phrases are composed of more than two words.
   (b) Prepositional phrases that function as a complement follow the basic rules of "S V C."
   (c) Prepositional phrases without content words (e.g., *for him*) are integrated into the former chunks.
   (d) The preposition *of* and *as* are connected with the prior nouns (e.g., *A of B*). However, if the object of the pronoun *of* consists of more than two words, the preposition *of* and its object are independent chunks.
   (e) Relative clauses within prepositional phrases (e.g., *about the way they dress*) are contained in prepositional phrases.

(continued)

Exceptions

10. Conjunctions / Relatives:
    (a) Basically, conjunctions and relatives start new chunks.
    (b) If there are big Ss or big Os in *that* clauses, the rules of big S and big O are applied.
    (c) In the case of the combination of big S and *be* verbs, big S is regarded as one chunk (Big S | (v) that...).
11. Others:
    (a) If an antecedent consists of one word, relative pronouns and their antecedents are integrated as big S and one chunk (e.g., *women who say*).
    (b) To infinitives are basically addressed as one chunk. However, some of them can easily cross chunk boundaries, such as *want* + *to* + *do* or *deserve* + *to* + *do*.
    (c) Present participles with perception verbs (e.g., *see* + O + *doing*) are regarded as one chunk because it becomes difficult to interpret the meaning if the participles are independent.
    (d) The number of phrases containing content words is two, plus or minus one.

Note. S = subject, V = verb, O = object.

Appendix 7

Learning Sentences, Target Words, and Probe Words Used in Experiments 3A and 3B

No.		Learning sentences and probes	LSA
1	HSS	He is not quite awake yet because he still needs to drink a cup of black *windle* this morning.	.42
	LSS	He wants to stop for moment because he wants to buy a pack of this *windle* in the shop.	.30
		Specific: コーヒー [coffee] General: のみもの [drink]	
2	HSS	Elizabeth II of Great-Britain is the most famous *yoot* in Europe.	.34
	LSS	That proud lady is the most famous *yoot* in Europe.	.30
		Specific: じょおう [queen] General: じょせい [woman]	
3	HSS	If Kate wants to free the canary birds, she has to open the iron *mand* they are in.	.36
	LSS	If Kate wants to see those animals, she has to find the special *mand* they are in.	.30
		Specific: かご [cage] General: いれもの [case]	
4	HSS	If you want to stay in good health, you have to eat five pieces of *brench* every day.	.33
	LSS	If you are able to go to the supermarket, you have to buy a lot of *brench* for me.	.22
		Specific: くだもの [fruit] General: たべもの [food]	
5	HSS	The charming dancer wears a rose in her golden *mear* but she will remove it later.	.31
	LSS	He spilled wine on her *mear* but he cleaned it up in a few seconds.	.28
		Specific: かみ [hair] General: あたま [head]	
6	HSS	When they are on a holiday, they always sleep in a rich *mork* with a beautiful pool.	.39
	LSS	When they are in Brussels, they always pass by a beautiful *mork* with an impressive pool.	.35
		Specific: ホテル [hotel] General: たてもの [building]	

(continuted)

No.		Learning sentences and probes	LSA
7	HSS	Breaking and entering into a house is against the *cadle* in every country.	.25
	LSS	Politicians of the new party are talking about a *cadle* for their country.	.22
		Specific: ほうりつ [law] General: きまり [rule]	
8	HSS	My husband always watches the seven o'clock *jurg* in the evening.	.19
	LSS	My husband always records the *jurg* in the evening.	.18
		Specific: ニュース [news] General: ばんぐみ [program]	
9	HSS	The unfaithful man cheated on his *nase* and had absolutely no regrets.	.45
	LSS	Our new friend talked to his *nase* and told her the whole story.	.40
		Specific: つま [wife] General: じょせい [woman]	
10	HSS	When Gary was young, he always confused a goose and a *bick* when naming animals.	.32
	LSS	When Gary was young, he kept a rabbit and a *bick* in his room.	.28
		Specific: あひる [duck] General: とり [bird]	
11	HSS	The surfers were attacked by a dangerous *sind* in the sea.	.35
	LSS	The group was surprised by a large *sind* in the sea.	.25
		Specific: サメ [shark] General: さかな [fish]	
12	HSS	France is a beautiful *tance* that attracts many tourists.	.42
	LSS	Ben visited the beautiful *tance* which attracts many tourists.	.41
		Specific: くに [country] General: ばしょ [place]	
13	HSS	He tried to put the pieces of the broken plate back together with *palk*.	.35
	LSS	She walked across the large room to Mike's dirty desk and returned his *palk*.	.14
		Specific: ボンド [glue] General: どうぐ [tool]	
14	HSS	The little puppy grew up to be a big *parrow*.	.33
	LSS	At the park she saw a man with a *parrow*.	.19
		Specific: いぬ [dog] General: どうぶつ [animal]	

(continued)

No.		Learning sentences and probes	LSA
15	HSS	The people marched to the beat of a loud *blund*.	.46
	LSS	The musician pounded on a *blund*.	.40
		Specific: たいこ [drum] General: がっき [instrument]	
16	HSS	The bank robber aimed at the security officer and fired the *vack*.	.43
	LSS	The robber had a *vack*.	.31
		Specific: じゅう [gun] General: ぶき [weapon]	
17	HSS	The pitcher was unable to throw the ball because of his broken *tring*.	.36
	LSS	He had a scratch on his *tring*.	.34
		Specific: うで [arm] General: からだ [body]	
18	HSS	The bridesmaid wore an ugly *prink*.	.69
	LSS	The princess wore a *prink*.	.65
		Specific: ドレス [dress] General: ふく [clothes]	
19	HSS	The farmer milked the *bettle*.	.41
	LSS	The farmer had an old *bettle*.	.24
		Specific: うし [cattle] General: どうぶつ [animal]	
20	HSS	That train always arrives on time at the *greal* in Tokyo.	.52
	LSS	My friend sometimes arrives too late at the *greal* in Tokyo.	.28
		Specific: えき [station] General: ばしょ [place]	

Note. Target pseudowords are italicized.

Appendix 8

Target Words With L1 Equivalents and Learning Sentences With LSA Values Used in Experiments 5A and 5B

Target word	Part of speech	L1 equivalent	Instance	LSA
			HSS learning sentences	
lick[b]	verb	なめる	The dog jumped up and licked his face	.69
puppy[a]	noun	子犬	Her little puppy grew up to be a big dog.	.60
monarch[a]	noun	君主	Elizabeth II of England is the most famous monarch in Europe.	.48
sob[b]	verb	むせび泣く	She stopped crying in a big voice and began to sob.	.48
spear[b]	noun	槍	He was killed with the long hunting spear.	.46
cattle[a]	noun	家畜用の牛	The farmer milked the cattle.	.41
doze[b]	verb	うとうとする	She closed her eyes and dozed.	.37
mourn[b]	verb	嘆く	They continue to mourn for years after the death of their friend.	.36
mane[a]	noun	長髪	The pretty dancer wears a flower in her golden mane.	.32
reef[b]	noun	(サンゴ) 礁	The small boat went south around the reef.	.30
			LSS learning sentences	
pier[b]	noun	埠頭	My brother and I were at the end of the pier and trying to catch fish.	.23
vehicle[a]	noun	乗り物	George taught his son to drive the vehicle.	.22
boulder[b]	noun	巨石	The boulder was as large as a small house.	.21
mako[a]	noun	アオザメ	The surfers were attacked by a dangerous mako in the sea.	.17
recluse[b]	noun	世捨て人	He was a recluse and never came to the town.	.14
terminus[a]	noun	終着駅	The train always arrives on time at the terminus in Tokyo.	.10
headlines[a]	noun	重要ニュース	My husband always watches the seven o'clock headlines in the evening.	.09
pawn[b]	verb	質に入れる	He pawned his watch to buy some new clothes.	.08
crave[b]	verb	欲しがる	The girl craves expensive clothes and bags.	.07
abhor[b]	verb	ひどく嫌い	We really abhor his English class.	.06

Note. Sentences[a] were from Experiment 3A, and sentences[b] were from Webb (2007a).

Appendix 9

VKS Elicitation Scale: Self-Report and Scoring Categories, and Meaning of Scores

質問(１)～(10)を読み，自分に最も当てはまる数字を丸で囲んで下さい。Ⅲ・Ⅳを選んだ場合は，下線部に単語の意味，または簡単な英文（ミスしても OK）を書いて下さい。

Self-report categories

(1) **abhor**

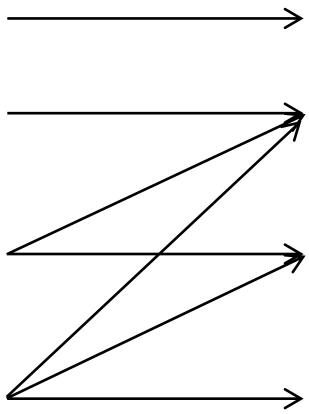
Ⅰ　この語は見たことがない。

Ⅱ　この語は見たことがあるが，どういう意味か分らない。

Ⅲ　この語は知っている。__________という意味である。

Ⅳ　この語は__________という意味で，この語を使って文を作れる。________________________

Self-report categories		Possible scores	Meaning of scores
Ⅰ		1	The word is not familiar at all.
Ⅱ		2	The word is familiar but its meaning is not known.
Ⅲ		3	A correct synonym or translation is given.
Ⅳ		4	The word is used with semantic appropriateness and grammatical accuracy in a sentence.



Appendix 10

Vocabulary Learning Strategies Questionnaire

1. Association (4 items)
   - (1) I associate a new word with a known word that looks or sounds similar to the shared part (e.g., I try to associate PRAY with PLAY).
   - (2) I remember a group of new words that share a similar part in sounds.
   - (3) I remember a group of new words that share a similar part in spelling.
   - (4) When I meet a new word, I search in my memory and see if I have any synonyms and antonyms in my vocabulary stock.

2. Imagery (4 items)
   - (5) I make a gesture of certain words (e.g., *stinking*) when I try to remember them.
   - (6) I act out a word meaning in order to remember it better.
   - (7) I create a mental image of the new word to help me remember it.
   - (8) I make a picture of a word meaning to help me remember it.

3. Word structure (3 items)
   - (9) I analyze words in terms of prefixes, stems, and suffixes (e.g., *unhappiness* → *un* + *happy* + *ness*).
   - (10) I deliberately study word-formation rules in order to remember more words.
   - (11) I memorize commonly used stems (e.g., *happy*) and affixes (e.g., *un*-, -*ness*).

4. Contextual encoding (3 items)
   - (12) When I try to remember a word, I remember a sentence in which the word is used.
   - (13) I remember the new word together with the context in which the new word occurs.
   - (14) I learn words better when I put them in contexts (e.g., phrases, sentences, etc.).

Note. Each statement was adapted from Gu and Johnson (1996) and Mizumoto (2006), and originally presented in Japanese.

Appendix 11

A Word List With Learning Sentences and Directions Used in Experiment 5A

- 下のリストにある（1）～（10）の単語を可能な限りたくさん覚えて下さい。**制限時間は 10 分**です。このページに書きこむことは構いません。
  この後すぐに**どれだけ単語を覚えられたかのテストをします**。

覚える単語	意味	例文
(1) mako	「アオザメ」	The surfers were attacked by a dangerous **mako** in the sea.
(2) abhor	「ひどく嫌い」	We really **abhor** his English class.
(3) cattle	「家畜用の牛」	The farmer milked the **cattle**.
(4) recluse	「世捨て人」	He was a **recluse** and never came to the town.
(5) spear	「槍」	He was killed with the long hunting **spear**.
(6) reef	「サンゴ礁」	The small boat went south around the **reef**.
(7) mourn	「嘆く」	They continue to **mourn** for years after the death of their friend.
(8) pawn	「質に入れる」	He **pawned** his watch to buy some new clothes.
(9) doze	「うとうとする」	She closed her eyes and **doze**d.
(10) puppy	「子犬」	Her little **puppy** grew up to be a big dog.

Appendix 12

Lexical Inference and MCG Tasks and Directions Used in Experiment 5B

■　これから行うテストでは，知らない単語の意味を文脈から推測します。

■　その後，３つの選択肢の中から，その単語の意味として適切だと思われるものを選びます。

■　最後に，全員で答え合わせをします。

＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊テストの回答方法＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊

●　推測テストの形式は次の通りです。

①　問題は全部で20問あります。

②　下線部が引かれた単語の意味を，周りの文脈から推測し，その**意味を日本語で**書きましょう。

例）The pitcher couldn’t throw a ball because of his broken **elbow**.

答え. elbow:　ひじ

●　選択テストの形式は次の通りです。

①　問題は推測テストと全く同じです。選択テストに進んだら，推測テストには戻れません。

②　(a), (b), (c) の意味の内，最も適切だと思う単語の**意味を日本語で**書きましょう。

例）The pitcher couldn’t throw a ball because of his broken **elbow**.

(a) ひじ　(b) むね　(c) 道具　答え. elbow:　ひじ

~~~~~~

分からない点はありませんか？

指示に従って課題を進めていきましょう。

Part A「推測テスト」

下線部の意味を周りの文脈から推測し，**日本語で**書きましょう。目標解答時間は 20 分です。

① The farmer milked the **cattle**.

答え. cattle: ____________

② He **pawned** his watch to buy some new clothes.

答え. pawn: ____________

③ The elephant was killed with the long hunting **spear**.

答え. spear: ____________

④ The **boulder** was as large as a small house.

答え. boulder: ____________

(continued)
~~~~~~

⑤ Her little **puppy** grew up to be a big dog.

答え. puppy: ________________

⑥ The girl **craves** expensive clothes and bags.

答え. crave: ________________

⑦ The dog jumped up and **licked** his face.

答え. lick: ________________

⑧ My brother and I were at the end of the **pier** and trying to catch fish.

答え. pier: ________________

⑨ Elizabeth II of England is the most famous **monarch** in Europe.

答え. monarch: ________________

⑩ He was a **recluse** and never came to the town.

答え. recluse: ________________

⑪ She stopped crying in a big voice and began to **sob**.

答え. sob: ________________

⑫ George taught his son to drive the **vehicle**.

答え. vehicle: ________________

⑬ She closed her eyes and **dozed**.

答え. doze: ________________

⑭ The surfers were attacked by a dangerous **mako** in the sea.

答え. mako: ________________

⑮ They continue to **mourn** for years after the death of their friend.

答え. mourn: ________________

⑯ The train always arrives on time at the **terminus** in Tokyo.

答え. terminus: ________________

⑰ The pretty dancer wears a flower in her golden **mane**.

答え. mane: ________________

⑱ He **pawned** his watch to buy some new clothes.

答え. pawn: ________________

⑲ The small boat went south around the **reef**.

答え. reef: ________________

⑳ We really **abhor** his English class.

答え. abhor: ________________

指示があるまで次のページ・前のページに進まない

(continued)

Part B「選択テスト」

(a), (b), (c) の内，最も適切だと思う**意味を日本語で**書きましょう。目標解答時間は 15 分です。

① The farmer milked the **cattle**.
(a) 牧草 (b) 牛 (c) やかん 答え.

② He **pawned** his watch to buy some new clothes.
(a) なくす (b) 手に入れる (c) 質に入れる 答え.

③ The elephant was killed with the long hunting **spear**.
(a) ハイエナ (b) 轟雷 (c) ヤリ 答え.

④ The **boulder** was as large as a small house.
(a) 巨石 (b) 肩 (c) 電柱 答え.

⑤ Her little **puppy** grew up to be a big dog.
(a) 父親 (b) 子犬 (c) かわいさ 答え.

⑥ The girl **craves** expensive clothes and bags.
(a) 欲しがる (b) 見つける (c) 洞窟に入る 答え.

⑦ The dog jumped up and **licked** his face.
(a) なめる (b) ぶつける (c) 噛みつく 答え.

⑧ My brother and I were at the end of the **pier** and trying to catch fish.
(a) 港 (b) 埠頭(ふとう) (c) 船 答え.

⑨ Elizabeth II of England is the most famous **monarch** in Europe.
(a) 国 (b) 女優 (c) 君主 答え.

⑩ He was a **recluse** and never came to the town.
(a) 世捨て人 (b) 町長 (c) 保安官 答え.

⑪ She stopped crying in a big voice and began to **sob**.
(a) 寝る (b) 怒る (c) むせび泣く 答え.

⑫ George taught his son to drive the **vehicle**.
(a) カブトムシ (b) 乗り物 (c) 馬 答え.

⑬ She closed her eyes and **dozed**.
(a) 休憩する (b) する (c) うとうとする 答え.

⑭ The surfers were attacked by a dangerous **mako** in the sea.
(a) アオザメ (b) 船 (c) 波 答え.

⑮ They continue to **mourn** for years after the death of their friend.
(a) 嘆く (b) 朝になる (c) 祈る 答え.

⑯ The train always arrives on time at the **terminus** in Tokyo.
(a) 時間 (b) 終着駅 (c) 期間 答え.

⑰ The pretty dancer wears a flower in her golden **mane**.
(a) 財布 (b) 男 (c) 長髪 答え.

⑱ He **pawned** his watch to buy some new clothes.
(a) 叩きつける (b) 見る (c) 質に入れる 答え.

⑲ The small boat went south around the **reef**.
(a) 葉っぱ (b) (サンゴ) 礁 (c) 島 答え.

⑳ We really **abhor** his English class.
(a) ひどく嫌い (b) 参加する (c) について 答え.

指示があるまで前のページに戻らない