Use of Semantic and Syntactic Cues in Sentence Parsing of EFL Learners

A Dissertation
Submitted to the University of Tsukuba
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Linguistics

Natsumi TANAKA

2016

Abstract of this Dissertation

Use of Semantic and Syntactic Cues in Sentence Parsing of EFL Learners

by

Natsumi TANAKA

Syntactic parsing is essential for readers to comprehend sentences, but English as a foreign language (EFL) learners have difficulties in processing syntactic cues, compared with native (L1) readers of English (Clahsen & Felser, 2006a, 2006b). To compensate for the syntactic deficits, learners intentionally choose to over-rely semantic or pragmatic information to process sentences. However, the issue of whether the differences between L1 and EFL sentence processing are fundamental or transactional has been controversial. Some studies insist that EFL sentence processing is inferior to L1 regardless of learners' English proficiency, cognitive resource, or processing speed. Others claim that EFL sentence processing is gradually changed into native-like processing.

To explain readers' disambiguated processing, previous studies have proposed and introduced various models and theories (e.g., the garden-path theory by Fraizer & Fodor, 1978; Fraizer & Rayner, 1982). These sentence processing models consider syntactic processing principles using the syntactic, semantic, pragmatic, lexical, and contextual cues provided by sentences. However, these models have rarely taken other factors into account, such as readers' features or situation or the task of sentence processing. Unlike in the case of L1 readers, there are various English proficiency levels of EFL learners. There are individual differences among these learners regarding explicit

and implicit knowledge of English, cognitive resources, or processing speed, and the learners are likely to process sentences differently. Additionally, it is possible that these learners use different strategies to process sentences.

The current dissertation investigated whether Japanese EFL learners utilize semantic or syntactic cues during sentence processing. To achieve this goal, the present study chose syntactically ambiguous sentences with a prepositional phrase (PP) attachment that had two possible attachment sites (e.g., *The little girl cut the apple with the plastic knife/the shiny coating*). One of the attachment sites was the verb modification (VP; *cut with the plastic knife*), and the other was the object noun phrase (NP; *apple with the shiny coating*). These sentences could be disambiguated using semantic, pragmatic or lexical information. In addition, the distance between the PP and its attachment site was manipulated in the sentences. In the long condition, learners have to maintain extra information (i.e., three-word phrase) until they process the disambiguated regions (i.e., after preposition) especially in the VP condition (e.g., *The little girl cut the large juicy green apple with the plastic knife*): The long condition can cause processing loads.

Study 1 consisted of two experiments regarding PP-attachment sentence processing (Experiments 1 and 2) that investigated learners' use of semantic and syntactic cues during or after reading sentences based on previous studies. Experiment 1 examined whether learners preferentially attached the PP to high attachment sites or low attachment sites with or without long-phrase intervening verbs and the object NP, using a multiple-choice preference and translation tasks. Additionally, the effects of the tasks on sentence-processing performance were examined. The results of the multiple-choice preference task indicated that the participants preferred to attach the PP to the object NP, which was a nearer option than the VP condition. The translation task did not show attachment preferences. When the verb types of each sentence were analyzed separately, the verb

biases of attachment sites were found. Distance effects were not found in Experiment 1, but the translation task alone demonstrated length effects. These inconsistent results between tasks might be due to task demands, whether or not a task required to choose syntactically ambiguous attachment sites, and whether or not the task required interpretation of the additional phrase.

Experiment 2 utilized word-by-word self-paced reading tasks to investigate learners' processing during reading (on-line). The participants answered comprehension questions that were not related to PP-attachment structures. In addition, a revised off-line preference task was conducted again to confirm the results of the preference task in Experiment 1. In the revised preference task, participants were required to report their confidence rates as well as point out proper attachment sites. The results suggested that self-paced reading did not show attachment preferences, in contrast to the L1 study (Thornton, MacDonald, & Arnold, 2000). The revised preference task showed a significant interaction between attachment sites and length condition, especially processing difficulties in the VP-long condition. The result was not exactly the same as in L1 studies, but the distance effects that occurred in L1 on-line processing appeared in EFL off-line processing.

In Study 2, two experiments were conducted to disclose how to construct or modify learners' detailed sentence representation using the think-aloud method. The think-aloud method was rarely used for sentence processing studies, because the oral protocols hardly reflected lower-level processing. However, the method can still provide researchers with representations or strategic processing of readers. First, Experiment 3 utilized local ambiguous sentences such as reduced and unreduced relative clause (RRC and URC) sentences to establish the think-aloud method for sentence processing. The types of sentences were not identical to other experimental sentences with PP attachment,

but obvious garden-path effects could be easily observed. The participants reported on what they thought or on the meanings of sentences while reading the sentences. They were also required to provide the L1 translation of each sentence as a final interpretation. Their individual skills were investigated by making them perform processing unit spans that were assessed by a timed sentence recall task (Hijikata, 2012). The results indicated that learners with large processing units infrequently reported syntactic or semantic processing; instead, they processed the sentence more effectively and implicitly than learners with small processing units.

Experiment 4 applied the think-aloud method to PP-attachment sentence processing. In addition, the verification task after reading was changed. The revised task required readers to resolve ambiguities and provided corrective feedback to ensure that learners disambiguated sentences. Readers also took the grammar proficiency test and processing unit tests. The results of the verification task showed that learners did not have bias regarding attachment sites and were not affected by three-word phrase intervening main verbs and the object NP. Additionally, the performance differed according to their grammar proficiency. The think-aloud protocols indicated some individual differences including processing unit span affected sentence processing. Other strategies such as using considerable amounts of semantic or pragmatic information were adapted by learners with small processing units, and their comprehension were not inferior to other learners with large processing units.

The main findings of the present study can be summarized as the follows:

(a) Learners' attachment preferences were highly affected by non-syntactic information such as lexical information of verb types, and usage of the preposition *with*. They displayed structural attachment preferences or biases if

they could not use the subcategorization information, for instance, the recency principle where learners attach phrases to the most recently processed.

(b) Learners' attachment preferences of preposition phrases varied according to their individual differences. EFL learners did not always prefer high attachment sites (i.e., VP attachment sites), which was inconsistent with the studies of L1 readers. However, the attachment preference gradually changed to native-like processing according to improvement of learners' grammar proficiency, or processing speed. In other words, EFL learners' processing was not restricted to a single representation.

(c) Learners did not respond sensitively to distances between constituents in on-line processing, but they responded to the sentence distance in off-line processing, although the reading pattern was not identical to L1 readers. Their syntactic processing was delayed compared to L1 readers. The additional phrase itself disturbed the whole sentence comprehension in off-line processing, even when the attachment sites and PP were not intervened by the long phrase.

(d) The task after reading sentences affected learners' processing. Tasks that required readers to resolve ambiguities promoted learners' reanalysis of sentences, while accuracy rates were not improved by merely assigning tasks.

(e) The think-aloud method was partially effective for sentence processing studies. Automatized initial processing and processing of learners with highly automatized syntactic processing were not shown, but the processing difficulties of EFL learners and their initial incorrect representations were revealed.

Based on these findings, some pedagogical implications for EFL sentence processing were provided. The present dissertation was meaningful, as it examined

learners' individual strategies of sentence processing, which were not focused on in previous studies. Additionally, the present study applied the think-aloud method to sentence processing. Although this method is not always effective for sentence processing studies, more detailed representations can be obtained through this method.

Acknowledgements

I am deeply indebted to many people for their invaluable guidance and assistance over the course of my studies at the Graduate School of Humanities and Social Sciences at University of Tsukuba. Although it would be difficult to acknowledge everyone who has contributed in various ways to the research reported in this dissertation, the following people deserve special recognition.

First and foremost, I would like to express sincere thanks to my chief adviser, Professor Yuji USHIRO, for an abundance of helpful advice and kind encouragement. He has presented me with many challenging tasks that seemed difficult to complete at first, although now I realize that these tasks contributed to who I am today as a student and teacher of English education. His theoretically and educationally beneficial comments have assisted to complete the doctoral dissertation. I will always be grateful to him for believing in me.

I am thankful to Professor Hirosada IWASAKI at the University of Tsukuba who supplied many useful suggestions with regard to vocabulary and grammar; because of his lectures, I was able to improve my academic writing skills. Professor Akira KUBOTA at the University of Tsukuba, who provides me the implications for chunking studies through his broad knowledge of the second language acquisition. His lectures have always awakened my interest in reading materials. I am also deeply grateful to my external supervisor, Professor Shuhei KADOTA from Kwansei Gakuin University. His research showed me how daily questions while learning English could be used for an educational study on English. He provided me with expert advice based on his own studies.

Many thanks are due to my senior associates, Dr. Shingo NAHATAME, Dr. Yusuke HASEGAWA, Dr. Yukino KIMURA, and Dr. Akira HAMADA who taught me how to research and what a researcher should be. I also appreciate Mr. Yoshinobu MORI, Mr. Masaya HOSODA, and all of my colleagues in Professor Ushiro's seminar who constantly provided crucial and beneficial comments, and always encouraged me during times of adversity.

I also greatly appreciate the contributions made by 155 undergraduate and graduate students at University of Tsukuba who participated in my experiments. The present study would not have been completed without their cooperation.

Last but not least, my special gratitude is extended to my family, who always stand by me with understanding, encouragement, and warm support. My parents have often given keen comments related to my studies, even though they majored in different subjects. Their passionate attitudes toward study and education are worthy of the greatest respect.

Natsumi TANAKA

Tsukuba, Japan

December, 2016

Table of Contents

Chapter 4

Study 2: Learners' Use of Strategy in Sentence Processing

List of Tables

Chapter 4

List of Figures

List of Appendices

Chapter 1
Introduction

1.1 Overview of the Current Problems

In traditional English-as-a-foreign-language (EFL) classrooms in Japan, the grammar translation method was widely used to teach English (Japan Society of English Language Education [JASELE], 2014; Mochizuki, 2010). The method focused on the comprehension of the meaning of English sentences in the learners' native language (L1, i.e., Japanese), where the teachers explained the grammar explicitly. As the importance of communicative skills became more emphasized, these methods gradually went out of use. However, some learners still tend to believe that comprehending the sentences is equal to translating the sentences into their own language. Comprehending sentences is one of the most basic and important skills for readers engaging in academic tasks. Even in a simple sentence, readers use and activate diverse information. For example, readers have to integrate semantic, syntactic, prosodic, and contexual information to comprehend sentences. However, readers can have more or less difficulty in syntactic parsing (Clahsen & Felser, 2006a, 2006b). Such tendencies are particularly strong among learners of a second language (L2) or foreign language (EFL), and native speakers (L1) also show the same priority of semantic information over syntactic information (Ferreira & Patson, 2007; Lim & Christianson, 2013; Tabor, Galantucci, & Richardson, 2004).

The broadly known thesis of learners' sentence processing posits that learners have fundamental difficulties in parsing syntactic information: Their syntactic representation is shallower and less detailed than L1 readers' representation, and the differences between native speakers and learners are not attributable to slow processing

time, limited cognitive capacity, or lack of grammatical knowledge (Shallow Structure Hypothesis, Clahsen & Felser, 2006a, 2006b). Instead, learners construct their sentence representations with semantic and pragmatic information to compensate for their shallow syntactic information. Other studies arguing the opposite view suggest that learners' poor performance is due to native language transfer and lack of processing speed, and the disadvantage reduces as their proficiency grows (Jackson, 2008; Juffs, 1998; Lim & Christianson, 2013; Rah & Adone, 2010b). They have also demonstrated that learners' performance in on-line tasks might be poorer than that of L1 readers, but their performance in off-line tasks is of the same level (e.g., Lim & Christianson, 2013). Arguments between the two different views have continued. They have agreed that successful sentence comprehension for learners depends on the semantic or pragmatic information that they can derive from sentences. This study will verify the question of whether the semantic and pragmatic cues are available to EFL learners; even when the syntactic structure becomes complex.

A plenty of prior studies have examined L1 or L2/EFL readers' sentence processing using universal or linguistically specific principles or models (e.g., garden-path model by Fraizer & Fodor, 1978); however these models attempt to apply to general sentence processing, regardless of readers' individual factors or situation where the readers pursue sentence comprehension. The sentence comprehension will be affected by various factors which might be not always linguistic principles, but also readers' intentional strategies use regarding sentence parsing or comprehension.

In order to focus on readers' syntactic parsing, prior studies have utilized syntactically ambiguous sentences as experimental materials. The present study utilized the syntactically ambiguous sentences that include two attachment sites of an element. For example, let us look at the following ambiguous sentences with prepositional phrase

(PP). The underlined phrase are different in each sentence. In Sentence (1a), the sentence is meaningful both *the criminal shot with the pistol* and *the cop with the pistol*. In Sentences (1b) and (1c), on the other hand, the most plausible attachment site was determined on the basis of the semantic or pragmatic information. Sentence (1b) is interpreted as *the criminal shot with the pistol*. Sentence 1c changed the second underlined phrase to *the collar*, thus it means that *the dog with the collar*.

(1a) The criminal shot <u>the approaching cop</u> with <u>the pistol</u> yesterday late at night.

(1b) The criminal shot <u>the approaching dog</u> with <u>the pistol</u> yesterday late at night.

(1c) The criminal shot <u>the approaching dog</u> with <u>the collar</u> yesterday late at night.

These PP-attachment sentence structures were chosen as an experimental sentences in this study, because the structure is relatively simple and understandable for EFL learners. The simplicity will ensure learners' understanding of sentence structure. Additionally, Japanese language systems do not contain PP attachment structures. Thus the effects of L1 transfer can be excluded from learners' performance.

To observe learners' sentence representation, many studies have investigated sentence processing using reading time or the readers' eye movements (e.g., Frenck-Mestre, 2005; Juffs, 1998). These methodologies have been confirmed, and precise data have been derived; however, these methods cannot elucidate the detailed representations of the readers unless researchers assume the readers' possible mistakes or interpretations beforehand. This study adapted the think-aloud method with a translation task or comprehension questions in order to observe the readers' diverse syntactic representations.

1.2 Organization of This Dissertation

This dissertation consists of the following five chapters: Introduction (Chapter 1: the present chapter), Related Literature Review (Chapter 2), Study 1 (Chapter 3), Study 2 (Chapter 4), General Discussion (Chapter 5), and Conclusion (Chapter 6), as shown in Table 1.1.

The present chapter reviewed the current problems of sentence processing studies in Section 1.1. Chapter 2 introduces the related areas of this dissertation: (a) general sentence processing models, (b) specific models for L2/EFL readers, (c) readers' individual factors that affect sentence comprehension, (d) experimental methods that have established sentence processing models, (e) distance and length effects on sentence comprehension, and (f) effects of post-reading task on sentence comprehension. This study focused on the syntactic processing of attachment structure, especially prepositional phrase (PP) attachment sentences; thus the studies on PP-attachment sentence were also introduced.

Chapters 3 and 4 include the main four experiments of sentence comprehension. In Chapter 3, Study 1 focuses on learners' processing of PP-attachment sentences. It included two off-line and on-line experiments conducted after experimental sentences are constructed in two kinds of norming studies. In the off-line experiment, the results of multiple question task and translation tasks are compared. In the next on-line experiment, participants are assigned the word-by-word self-paced reading of PP-attachment sentences to gain the reading times of each region. They answer the comprehension questions that do not require the attachment site to be decide.

In Chapter 4, Study 2 investigates the process and strategy of EFL learners' interpretations using the think-aloud method, which requires readers to report orally what they think when reading texts or sentences using local and global ambiguous sentences.

In the third experiment with the local ambiguous sentences, participants perform the think-aloud method during reading, and after that, they translate sentences into Japanese. The garden-path sentences used in this experiment allows for the reanalysis processing of learners to be observed. The final experiment also requires participants to explain sentences orally before they answer comprehension questions with corrective feedbacks.

Chapter 5 discusses the findings from the four experiments and pilot study from the following five perspectives: (a) learners' use of semantic and syntactic cues, (b) effects of learners' individual factors on syntactic parsing strategies, (c) post-reading task demands and differences, and (d) effectiveness of the think-aloud method in sentence processing studies. Finally, in Chapter 6, the main findings regarding sentence processing are summarized before concluding remarks are made denoted the limitations and suggestions for further research. Chapter 6 also offers the pedagogical implications to Japanese EFL classrooms, methodological implication in EFL sentence studies, and theoretical implications to EFL sentence models.

Table 1.1

Summary of the Structures in the Present Dissertation

Chapter 2 **Study 1: Learners' Off- and On-line Processing of PP-attachment Sentences**
Experiment 1: Off-line processing of PP-attachment sentences ■ Design: Attachment sites × Length of sentences (within-factors) ■ Materials: Globally ambiguous PP-attachment sentences ■ Measures: Multiple-choice preference task and translation task
Experiment 2: On-line processing and detailed off-line processing of PP-attachment sentences ■ Design: Attachment sites × Length of sentences (within-factors) ■ Materials: Globally ambiguous PP-attachment sentences ■ Measures: Reading time data (with comprehension questions) and confidence rates of preference
Chapter 3 **Study 2: Learners' Use of Strategy in Sentence Processing**
Pilot study for processing unit tests
Experiment 3: Learners' use of strategy in locally ambiguous sentence processing ■ Design: Reading proficiency, processing unit (between-factors) × Structures of sentences (within-factor) ■ Materials: Locally ambiguous RRC and URC-attachment sentences ■ Measures: Processing unit test, reading proficiency test, think-aloud method with translation
Experiment 4: Learners' use of strategy in globally ambiguous sentence processing ■ Design: Reading proficiency, processing unit (between-factors) × Attachment sites × Length of sentences (within-factors) ■ Materials: Globally ambiguous PP-attachment sentences ■ Measures: Processing unit test, grammar proficiency test, think-aloud method with comprehension questions

Note. PP = Prepositional Phrase, RRC = Reduced relative clause, and URC = Unreduced relative clause.

Chapter 2

Literature Review

2.1 Sentence Processing and Sentence Parsing Strategies

When comprehending sentences, readers pursue complex processing using syntactic, semantic, pragmatic, and lexical cues (Bornkessel-Schlesewsky & Schlesewsky, 2009; Juffs & Rodríguez, 2015; van Gompel & Pickering, 2007) in addition to the referential context, visual and prosodic cues (Dekydtspotter, Donaldson, Edmonds, Fultz, & Petrush, 2008; O'Brien, 2014; Pan & Felser, 2011; Ying, 1996; Yokokawa, 1994). Some sentences are processed easily, and readers have difficulties processing the certain kinds of sentences or sentence structures due to lack of information of sentences, or lack of knowledge of readers. Before reviewing the sentence processing models, this section introduces why and how some sentences become ambiguous to readers.

Comprehenders incorporate each word into the preceding syntactic structure when they encounter it rather than construct syntactic structure after processing the whole sentence or clause (van Gompel & Pickering, 2007). This *incremental* parsing style requires comprehenders to reanalyze the sentence or clause, once initial parsing turns out to be incorrect in the semantic, syntactic, or contextual way. It seems to be ineffective process, but comprehenders' parser do not wait for end of the sentences to begin parsing them in order to save their cognitive resource. Using simple-structure sentences that comprehenders do not have to reanalyze makes it difficult to observe comprehenders' processing; thus, researchers focus on ambiguous sentences that allow readers to interpret more than one syntactic structure or semantic interpretation (Juffs & Rodríguez, 2011; McRae & Matsuki, 2013; van Gompel & Pickering, 2007).

Based on incremental processing, some types of ambiguous sentences require comprehenders to reanalyze interpretation after the initial interpretation. This includes *garden-path* sentences that have strict contradictions compared with other ambiguous sentences. These types of ambiguous sentences are called *local ambiguities*. The ambiguity will be resolved when comprehenders complete reading these ambiguous sentences, because part of the structure will make them confuse and reanalyze the sentences. For example, the following garden-path sentence introduced by Bever (1970) is the most well-known as the main verb/reduced relative ambiguity:

(2a) The horse raced past the barn yesterday.

(2b) The horse raced past the barn fell.

As Sentence (2a) include no ambiguity, readers do not experience any processing difficulty to comprehend that *raced* is the main verb of this sentence. In contrast, in Sentence (2b), readers will initially interpret that the first verb *raced* is the main verb of this sentence, but they will notice their mistake after they encounter the second verb *fell*. At this point of ambiguity, they have to reanalyze the sentence and resolve that the first verb must be included in a reduced relative clause (RRC). Readers have difficulties in comprehending these kinds of sentences that include deep ambiguities.

Another type of ambiguous sentence is categorized as *global ambiguity*. Global ambiguities occur not due to incremental processing, but due to syntactic ambiguities: thus such sentences are called *syntactically/structurally ambiguous sentences*. In those sentences, for example, sentences include the multiple possible attachment sites as shown in the following sentences: *The criminal shot the approaching cop with the pistol yesterday late at night*. The sentences do not have a correct attachment

site or interpretation, but comprehenders sometimes prefer one of the either options using semantic, pragmatic, lexical or referential context cues. In the case that the information that can disambiguate sentences are not given, comprehenders resolve ambiguities based on the structural preferences (van Gompel & Pickering, 2007; Witzel, Witzel, & Nicol, 2012).

2.1.1 General sentence processing models

In order to explain sentence processing, many researchers have proposed various models. The models are divided into two types: one type predicts linguistic rules using several principles (Fraizer & Fodor, 1978; Fraizer & Rayner, 1982; van Gompel & Pickering, 2007; van Gompel, Pickering, & Traxler, 2001), and the other is based on comprehenders' experience of language use (Rah, 2009; Snedeker, & Trueswell, 2004). The most commonly cited example of the first model type is the garden-path model introduced by Fraizer and her collaborators (Fraizer & Fodor, 1978; Fraizer & Rayner, 1982). Experience-based models are represented by the constraint-based model (MacDonald, 1994; Snedeker, & Trueswell, 2004). The garden-path model intends to be cross-linguistic or universal, and it has two stages: First, processor adopts syntactic analysis, and then other analyses were employed. The model includes two principles: *minimal attachment* and *late closure*. In the minimal attachment principle, comprehenders attach incoming material to the phrase-marker being constructed using the fewest nodes consistent with the well-formedness rules of the language, while late closure predicts that comprehenders attach incoming lexical items to the clause or phrase currently being processed, when possible (Fraizer & Rayner, 1982, p.80). A similar model to this garden-path model was the Race model by van Gompel et al. (2001) which also assumed that comprehenders attempt to parse single syntactic structures at the same time. When they

construct the structure incrementally and notice any difficulty in parsing, comprehenders should discard the preceding structure and adapt other structure from the beginning in order to reanalyze and reach the correct structure. The electrophysiological study also supported the two-stage processing models, it claimed that the syntactic processing has two steps: initial processing is automatic and late processing is more controlled according to the results of event-related brain potential measures (Hahne & Friederici, 1999).

On the other hand, the constraint-based model's focus was on interactions of semantic, pragmatic, and referential information as well as syntactic parsing. The available information is parsed immediately and simultaneously. The model assumed five principles (McRae & Matsuki, 2013; Snedeker & Trueswell, 2004): (a) multiple sources of information (i.e., constraints) are used for comprehending sentences and resolving ambiguities; (b) there is little or no delay in information availability; (c) there is no delay in information usage once it becomes available; (d) multiple potential alternative interpretations are activated probabilistically in parallel; and (e) anticipation or expectation of structure and contents are frequently assumed. In addition, the model assumed that processing is guided by linguistic specific principles in contrast with the garden-path model.

The garden-path model and the constraint-based model have been compared widely, as the two models assume different sentence processing in serial or parallel (van Gompel & Pickering, 2000). The garden-path model assumes that readers repeat the process of examining and discarding one interpretation at a time. The constraint-based model, on the other hand, predicts that readers activate multiple interpretations at the same time, and these interpretations compete with each other. As parsing proceeds, the more appropriate and plausible interpretation is activated and the less appropriate one is suppressed. Finally, the most appropriate and plausible one survives. There is evidence

of competition among interpretation results in ambiguous sentences with good cues that take longer reading time.

The models or theories as reviewed above predicted and assumed how comprehenders process syntactic, semantic or other cues of sentences. The *good-enough theory* suggests that readers do not always syntactically parse ambiguous sentences (Christianson, 2016; Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Tabor et al., 2004). When context, semantic, and syntactic information that can resolve ambiguity are not provided, readers construct surface level or shallow structures of a sentence as shown in Figure 2.1.

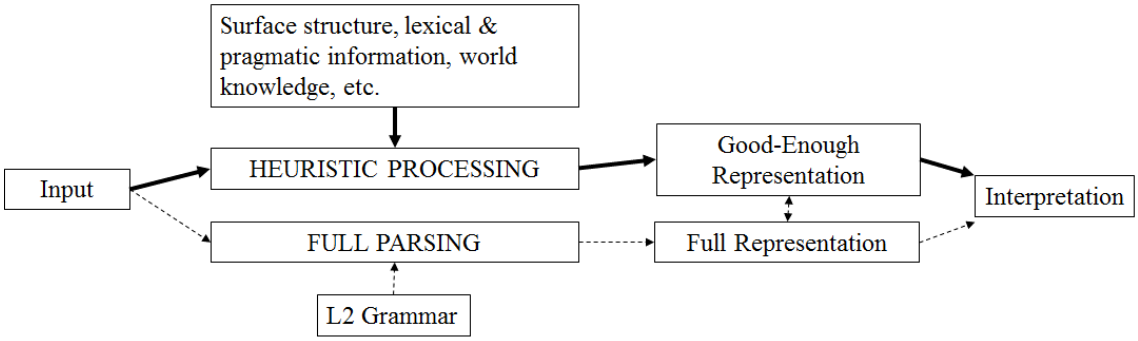

Figure 2.1. Models of good-enough hypothesis adopted by Lim and Christianson (2013).

According to this hypothesis, the parser has a heuristic processing (i.e., arrows of heavy line above in Figure 2.1) that is not always consistent with the inputted structures of sentences (Ferreira & Patson, 2007). Even L1 comprehenders inconsistently parse simple structure sentences such as passive sentences without a time limitation, which is called the effect of *local coherence* (Ferreira & Patson, 2007; Lim & Christianson, 2013; Tabor et al., 2004). When they read the following simple passive sentence, for example, many L1 readers mistakenly interpret that *the dog bit the man*, because their heuristic processing preferentially parses the first subject as the agent.

(3) The dog was bitten by the man.

The effect of local coherence frequently occurs when sentences include garden-path structure. Tabor et al. (2004) demonstrated that when readers encounter the following reduced relative clause/main clause ambiguous sentence: *The coach smiled at the player tossed a frisbee by the opposing team*, they tended to parse "the player tossed a frisbee," which interferes with the global interpretation. As seen so far, the good-enough theory assumes that comprehension errors in sentences are based on the particular preferences of readers (Christianson, 2016).

2.1.2 Sentence processing models of L2/EFL learners

The same tendency of difficulties in syntactic processing can be observed in L2 or EFL sentence comprehension. The tendency might even be stronger than in native speakers due to EFL learners' specific situation. Many previous studies have argued that learners have difficulties in processing syntactic structure automatically and they heavily rely on other lexical or semantic cues in order to compensate for these gaps for comprehension (Clahsen & Felser, 2006a, 2006b; Felser, Roberts, Marins, & Gross, 2003; Juffs, 1998; Lim & Christianson, 2013; Marnis, Roberts, Felser, & Clahsen, 2005; Papadopoulou & Clahsen, 2003). However, whether the differences between native speakers and learners are due to a fundamentally different approach to syntactic representation, or for other reasons such as the lack of working memory capacity and processing speed of learners is still a controversial issue.

Regarding the former position, Clahsen and Felser's Shallow Structure Hypothesis (2006a, 2006b) is well known. This theory claims that even if learners possess the same

level of grammatical knowledge, processing speed, and working memory capacity as L1 readers, learners' syntactic processing would still somehow be shallow and non-detailed, compared with L1 readers (Clahsen & Felser, 2006a, 2006b; Felser, et al., 2003; Marnis, et al., 2005; Papadopoulou & Clahsen, 2003). Therefore, shallow processing and full parsing take fundamentally different routes as shown in Figure 2.2, and the integration between them do not occur unlike good-enough theory. The notion of shallow parsing involves "identifying parts of speech, segmenting the input strings into meaningful chunks, and determining what relations these chunks bear to the main verb" (Clahsen & Felser, 2006a). To compensate for syntactic processing problems, learners might prefer short-distance over long-distance relations.

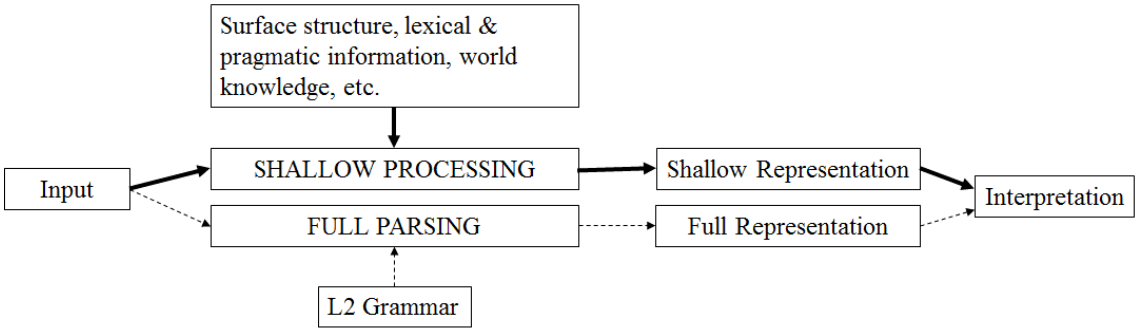

Figure 2.2. Shallow structure hypothesis adopted by Lim & Christianson (2013).

The latter transitional position contends that the differences between L1 and L2/EFL learners' processings are "quantitative," namely L2/EFL learners are inferior to L1 readers in the speed of processing or retrieving syntactic information. L2 learners with relatively high proficiency can retrieve morphosyntactic information and reach native-like processing patterns, though their processing speed is inferior to that of L1 readers (Jackson, 2008; Juffs, 1998; Lim & Christianson, 2013; Rah & Adone, 2010b; Witzel et al., 2012). The lack of processing speed is gradually resolved according to learners'

advance in proficiency (Dekydtspotter et al., 2008). The study which controlled the processing speed of EFL learners and L1 readers even claimed that learners processed sentences faster than L1 readers (Kaan, Ballantyne, & Wijnen, 2015). Some off-line data of learners has been found to be almost identical to that of native speakers. Researchers have argued that learners could reach the same final interpretation as native speakers, except that they have difficulties in constructing syntactic representation during reading. For example, Rah and Adone compared on-line and off-line performances of native speakers, and upper and intermediate learners of English. They focused on the ambiguity between reduced relative clause (RRC) structure and main clause using the following sentences:

(4a) The brown sparrow noticed on an upper branch pecked at an insect. [ambiguous RRC sentence with a good cue]

(4b) The brown sparrow noticed almost every day pecked at an insect. [ambiguous RRC sentence with a poor cue]

They assigned self-paced reading task (on-line) and grammaticality judgement tasks (off-line). The off-line scores of learners were almost the same as native speakers, but learners took more time to read sentences whose syntactic structures were different from those of the learners' mother tongue (i.e., first language interference). Similarly, Lim & Christianson (2013) measured L1 readers and EFL learners' on-line and off-line syntactic representation using traditional self-paced reading and translation tasks in order to compare their integration of semantic and syntactic processing of sentences. They manipulated RRC structures (e.g., subject and object relative clauses) and plausibility (i.e., plausible and implausible). Syntactic processing pattern of L1 readers and EFL learners

were quite similar, excepting the processing speed. They concluded that EFL learners could use syntactic information, but they had difficulties in integrating syntactic and semantic routes.

2.1.3 Readers' individual factors of sentence processing

This section will review the individual differences that affect sentence processing. Sentence processing is affected by many different EFL sentence-processing factors. Studies have attempted to control for individual factors such as reading proficiency and cognitive capacity (i.e., working memory capacity), but these studies do not focus on the effects of individual differences (Roberts, 2012). The major factors of individual differences are working memory capacity, explicit and implicit grammatical knowledge, processing speed, and learners' proficiency in English.

Grammatical knowledge

Both explicit and implicit grammatical knowledge are essential to parse syntactic structures of sentences, but these have some distinguishing features according to Ellis (2005): Implicit knowledge involves automatic processing, whereas explicit knowledge entails controlled processing. In addition, Ellis showed that implicit knowledge cannot be verbalized, while explicit knowledge is potentially verbalized (Bowles, 2011). It is controversial whether explicit knowledge can be converted into implicit knowledge (interface position) or not (non-interface position), or whether it can in-between, but some researchers with an interface position insist that implicit knowledge can be measured by tasks with time limitation (Ellis, 2005; Hulstijn, 2005).

Working memory capacity

Sentence comprehension requires readers to process syntactic, semantic, pragmatic, lexical, and context information, or multiple interpretations at the same time. In addition, syntactic processing transforms a linear sequence of words into a nonlinear syntactic structure, and this transformation requires the temporary storage of word representations during left-to-right processing of a sentence (King & Just, 1991). In order to allot readers' limited cognitive resources, *working memory* plays a critical role (Baddeley, 1986; Daneman & Carpenter, 1980). The two main functions of working memory are simultaneous storage and processing of information (Baddeley, 1986; Daneman & Carpenter, 1980; Osaka, 1998). When task requirements exceed resource capacity, a trade-off relationship comes into existence between processing and storage functions, as cognitive resources available in working memory, also known as *working memory capacity* (WMC), are limited. The shortage of capacity negatively affects task performance, through factors such as slower processing and information loss. For example, when learners read difficult sentences, they engage their working memory capacity to ascertain the meaning of the passage and, as a result, fewer available resources remain for storage. Working memory capacity differs between individuals, and working memory capacity is closely related to linguistic proficiency especially in L2 or EFL (Friedman & Miyake, 2005; Osaka, 1998). With respect to reading, a *phonological loop* is a subsystem of the working memory model (Baddeley, 1986) that processes information related to language. A phonological loop has two important roles: decoding written words in phonological form for storage and rehearsal, and storing phonological information for a few seconds and refreshing the stored sounds with sub-vocalization (*inner speech*). The inner speech basically collapses only in two minutes after the parser inputs linguistic information in case that rehearsal does not occur. Thus, the processing speed or

automatized processing is also required for effective parsing, as described below.

The effects of working memory capacity on sentence processing have been demonstrated again and again using sentences with especially complex structures such as RRC structures (Hopp, 2006; King & Just, 1991). The relationships between WMC and sentence processing are two-fold: structural complexity and ambiguity of sentences (Koda, 2005). The structurally complex sentences requires readers to retain preceding information until the structure complete. Similarly, the structurally ambiguous sentences induces more than one interpretation; readers examine these possible interpretations, and their WMC are restricted. For EFL sentence processing, Nakanishi (2012) investigated the effects of individual working memory capacity on sentence comprehension performance. The study found that reading time of sentence comprehension of EFL learners is highly affected by working memory capacity. While this study will not use WMC, but it closely related to the following factors.

Sentence processing unit and speed

Processing unit and speed are the next factor that affects sentence comprehension. These factors of processing unit span and speed are closely connected each other: Readers who process the sentences with large processing unit, can process the whole sentence faster than learners with small processing unit. *Chunking* during reading is one of the things readers can do for more fluent reading (Ellis, 2001; Schmidt, 1992). When readers parse texts, they integrate several amounts of information (i.e., group of words) into one meaningful unit such as a clause, phrase, or even sentence, which is called chunk (Ellis, 2001; Miller, 1956). Chunking the bringing together a set of already formed chunks in memory and welding them together into a large unit, is a basic associative learning process which can occur in all representational systems (Ellis, 2001). This process of

chunking allows readers to save their available WMC for lower-level processing, and to allot more WMC from higher-level processing, which leads to a fluent and proficient reading as a whole.

Readers with fast processing speed can perceive sentences with large unit. Therefore, similarly to the processing speed, processing unit will be the factors that affect sentence comprehension. Hijikata (2012) mentioned that the processing unit might be extended to the sentence unit with a different methodology. Hijikata adopted the eye-voice span test that assesses how many words readers look ahead during reading aloud; and the written recall task that requires to write down what learners remember in reading passages, in order to measure the span of learners' perceived processing units: She presented a sentence for five seconds to her participants, Japanese EFL university students. After five seconds, the sentence disappeared from the screen and the participants wrote down what they remembered in English. When they could read and write the whole sentence within five seconds, their chunking spans were seen as large enough to read in the given time. The results supported the notion that Japanese EFL learners perceive a phrase on a larger level as a processing unit. Additionally, sometimes the size of processing units can be extended to the sentence level, as learners can recall the whole sentence in the test. Hijikata also found that the processing unit is related to the learners' reading proficiency. Concerning individual differences, the comprehension and reading times of more proficient learners or learners who use larger chunking units suffer from smaller unit chunking such as at the word level (Hijikata, 2012; Yamashita & Ichikawa, 2010).

2.1.4 Experimental methods for the sentence processing study

Researchers have used different off-line and on-line methods to investigate

readers' sentence processing (Juffs & Rodríguez, 2015; Keating & Jegerski, 2014; van Gompel & Pickering, 2007). The off-line methods focus on readers' sentence representation after they process the sentence several times (both initial and reanalysis processing). As participants complete to read the whole sentence, the ambiguities due to the incremental processing have been resolved when they undertake the off-line method; thus the off-line methods are not sensitive to reanalysis process. The typical off-line methods are giving grammatical judgment tasks and translation tasks. The grammatical judgment task requires readers to decide whether each experimental sentence includes grammatical errors or not. When the task set strict time limit to answer, it is regarded to measure on-line processing of readers. Another way to know learners' specific representation is through the translation method. In the translation method, readers translate L2/EFL sentences into their L1, or L1 sentences into L2/EFL. This method makes it possible to know readers' specific representation after their initial and second processing (Lim & Christianson, 2013). The translation task, however, includes three-steps: comprehending sentences in original language, switching two languages, and producing in a targeted language (Lim & Christianson, 2013; Macizo & Bajo, 2006). These process were demanding, and it needs more cognitive capacity than such simple tasks as repetition. When readers make some errors in translation protocols, it is regarded that they have difficulties in processing the sentences.

Investigating how readers parse or comprehend sentence during reading is another way to learn about their sentence processing as well as representation. Readers construct sentence representation using various information, and when they encounter syntactically or semantically ambiguous sentences, they sometimes reread the ambiguous region, or retrieve cues from their own constructing representation in order to resolve the ambiguities. To examine such shifting representation in on-line, researchers have used

self-paced reading tasks using non-cumulative, moving-window technique; and eye-tracking paradigms (e.g., Frenck-Mestre, 2005; Juffs, 1998); or brain response such as event-related potentials (Tanenhaus, 2004), and functional magnetic resonance imaging (fMRI) (Love, Haist, Nicol, & Swiney, 2003). In self-paced reading tasks, participants read in a word-by-word or phrase-by-phrase style; namely, one segment (usually word or phrase) appears on the computer screen at a time. Readers press keys to precede the sentence at their own rate, but they cannot look back at any previous parts once they move on to the next segment. Eye-tracking methods do not include such limitations of lookbacks or re-reading, and so participants can read sentences in a natural way. In particular, the eye-tracking method can discriminate early measure (i.e., initial processing effects) from late measure (i.e., reanalysis processing effects), and figure out readers' sentence processing on time course; though the self-paced reading task cannot do so. In both methods, reading/processing times (i.e., fixation times in eye-tracking study) of a given region are recorded for comparison between conditions or regions. Processing difficulties or making inferences are assumed to be reflected in slower reading times, as readers will not take time to read parts without processing difficulties. Both methods sometimes requires comprehension or verification questions after reading in order to confirm that participants do not precede reading texts without paying attention to the contents, because the reading/processing times themselves do not guarantee readers' comprehension. These methods in sentence processing studies have been used to investigate anomaly detection, ambiguity resolution, and syntactic dependency formation.

These methods provide reading time data at the precious point, and they are believed to tap readers' implicit knowledge of language. However, at the same time, experimental design flaws can affect participants' response data in ways that researchers do not expect (Keating & Jegerski, 2014). Additionally, researchers have pointed out that

the comprehension question after reading might affect performance, as it works like a probe question (Christianson et al., 2001). Another problem is that experimental design has to focus on the statistic differences as to whether readers process one segment faster than another segment, and the data do not indicate what readers interpret the sentence in detail and what in real time.

To resolve the problem, the *think-aloud method* could focus on readers' specific thoughts during reading. Think-aloud is a research method that requires learners to report what they are thinking while performing a task (e.g., Olson, Duffy, & Mack, 1984). It focuses on the processes or strategies that participants use during tasks: reading or writing a passage, or learning vocabulary. Some criteria of categories were introduced for L1 or L2/EFL readings (e.g., Horiba, 1996, 2013). Horiba originally set the following categories: (a) *graphophonemic/graphomorphemic analysis*, (b) *word recognition*, (c) *syntactic/ semantic analysis of clause or sentence*, (d) *in-text inference*, (e) *reader response*, (f) *self-monitoring*, (g) *comment on text structure*, and (h) *others*. She counted the number of each category in think-aloud protocols, and the correlation with other indices were analyzed. Advantages of the think-aloud method is to report the readers' detailed text representation during reading. Although few studies used the think-aloud method to examine sentence comprehension as some limitations were found when applying this method for sentence processing (Olson et al., 1984). First of all, the process of think-aloud are generally possible to affect readers' comprehension process. Reading with think-aloud activates readers' self-monitoring, and their comprehension will be promoted, compared with other tasks. Additionally, oral protocols did not always reflect readers' on-line thoughts, especially because lower-level processing such as sentence parsing was based on unconscious and rapid processing, unlike passage reading. Finally, the think-aloud method is a dual task that requires more attentional resources during

reading. Olson described the possibility of using the think-aloud method in sentence processing research as follows: “Syntactic and semantic analysis of sentences may or may not be usefully analyzed.” This is why only a few studies have applied this method to the syntactic parsing situation. One of them targeting a multi-lingual speaker whose L1 was Polish, L2 was English, and L3 was Italian used the think-aloud method to reveal her own processing strategy (Gabryś-Barker, 2008). However, this study had limitations that the number of participants were only one. Other sentence studies with the think-aloud method are needed.

2.2 Relationships Between Syntactic and Semantic Cues in Sentence Processing

2.2.1 Distance and length effects in sentence processing

The distance of sentence processing elicits processing difficulties such as filler gap construction. This has been argued by dependency locality theory (DLT; Gibson, 1998, 2000). The theory indicated that on-line processing becomes difficult with the great distance between constituents in sentences (i.e., locality effects). DLT suggested the two loads: one of them is structural integration, consisting of integrating new input words into the currently existing syntactic and discourse structures; and another one is storage of the syntactic categories that are necessary to complete the current input string as a grammatical sentence (Gibson, 1998; Gibson & Pearlmutter, 1998). So far, previous studies have focused on the distances of difficult structures in L1 and L2 sentences and found out that the distance between constituents in such constructions impedes readers’ comprehension. For example, Gibson and his colleagues used the filler gap dependency structures including relative clauses (Gibson, 1998, 2000; Grodner & Gibson, 2005). While these studies focused on complex sentence structures, Bartek, Lewis, Vasishth, and Smith (2011) demonstrated that L1 readers have difficulty in processing even simple

structures such as subject-verb. The difficulties were shown during the initial processing according to their eye-tracking experiment (Experiment 3). In addition, the constituent length also caused the processing load during sentence comprehension (Hemforth, Colonna, Petrone, & D'lmperio, 2013). Considering the limited cognitive resources of L2/EFL learners, the distant itself will influence learners' sentence comprehension, even if the distant is related to the whole sentence structure (e.g., filler-gap structure) or the distance includes some complex structures such as relative clauses.

2.2.2 Attachment structures and prepositional-phrase attachment sentences

Attachment structures in sentence processing studies.

To examine the relationships between syntactic and semantic information, some studies have addressed the question of whether the merits of semantic or pragmatic information could override syntactic complexities that cause syntactic memory or integration load. Before explaining the relationships between syntactic and semantic information, the attachment structure that is targeted in this study will be introduced. Phrase or clause attachment are some of the most frequently used structures in sentence processing studies, and relative clause (RC) attachment and prepositional phrase (PP) attachment have been focused on to date. The EFL processing of RC-attachment sentences are affected by their mother tongue, because many languages possess the RC attachment structures. The famous sentence modelling RC-attachment is shown below (Jun, 2003):

(5a) Someone shot <u>the servant</u> of <u>the actress</u> <u>who was on the balcony</u>. [English]

(5b) dareka-ga barukonii-ni-iru joyuu-no mesitsukai-o utta. [Japanese]

In the globally ambiguous sentence (5a), the RC (*who was on the balcony*) can both structurally and semantically be attached to the two noun phrases that are underlined. When the plausibility of the two attachment structures are the same, L1 readers of English preferentially choose the close attachment site: *It was actress that was on the balcony* (Juffs & Rodríguez, 2015). The Japanese language also has a similar structure, as shown in Sentence (5b), and native speakers of Japanese initially indicated a low-attachment preference in such RC-attachment structures. However, after they finished reading the entire sentence, their comprehension was revised to the high-attachment one (Kamide, Mitchell, Fodor, & Inoue, 1997).

On the other hand, comprehension of PP-attachment sentences do not interfere with readers' L1 attachment preferences, because the Japanese language does not have prepositions. Some studies focused on PP attachment ambiguities in L1 processing (Frenck-Mestre & Pynte, 1997; Schütze & Gibson, 1999; Thornton, et al., 2000) and in EFL processing (Nakanishi, 2012; Ying, 1996) to explore the relationship between readers' syntactic and semantic or pragmatic processing. These sentences contain two possible attachment sites of a *with*-PP, as exemplified by the following sentences (see Figure 2.3):

(6a) The criminal shot <u>the approaching cop</u> with <u>the pistol</u> yesterday late at night.

(6b) The criminal shot <u>the approaching dog</u> with <u>the pistol</u> yesterday late at night.

(6c) The criminal shot <u>the approaching dog</u> with <u>the collar</u> yesterday late at night.

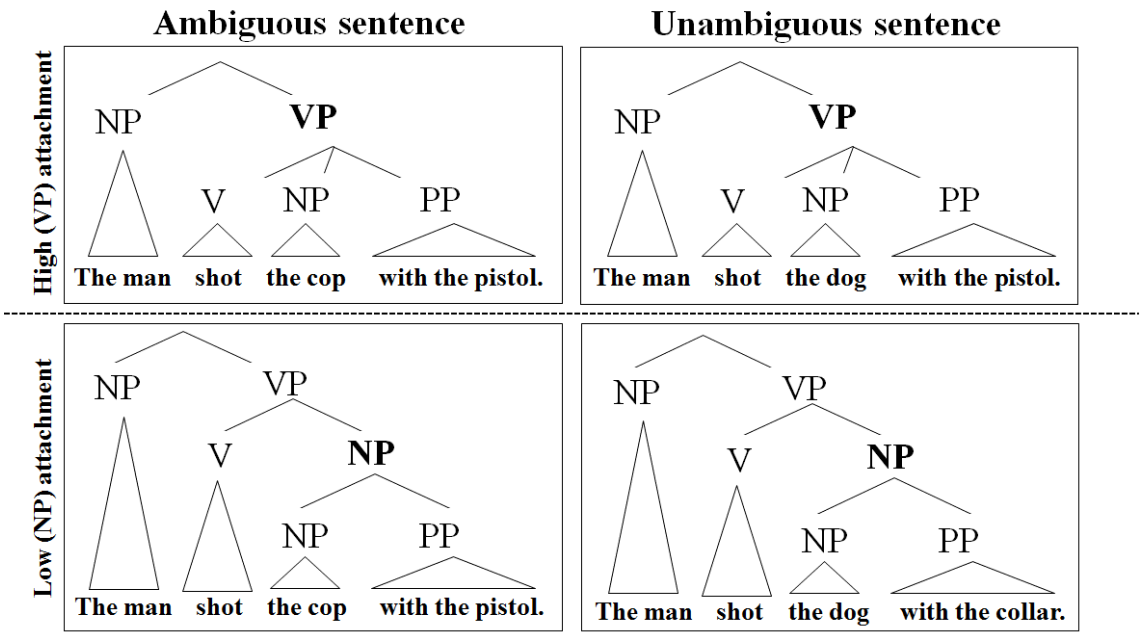

Figure 2.3. Tree diagram of PP-attachment sentence of Sentence 6a, b, and c.

In ambiguous sentence (6a), readers could regard the PP *with the pistol* as a modification of both the verb phrase (VP) *shot the approaching cop* and the second noun phrase (NP) just before the PP *the approaching cop*. Additionally, they could not resolve the ambiguity syntactically or semantically, except in cases where sentences were presented in other contextual discourse. Meanwhile, in the other two sentences, readers could decide one of two attachment sites on the basis of sentence meanings such as pragmatic or semantic cues, although syntactic cues were not available. The PP in sentence (6b) can be interpreted as a modification of VP *shot the approaching cop* (distant site/high attachment). In this case, the *with*-phrase indicated instruments of actions, and the user of the instruments was limited to the subject of the sentence. Similarly, sentence (6c) is biased toward NP attachment; thus, readers could interpret the PP as a modification of the second NP just before the PP *the approaching dog* (local site/low attachment).

Using these kinds of sentences, previous studies have discussed readers'

preferred attachment site with such off-line data as preference decisions and such on-line data as reading-time or eye-tracking data, as shown in Table 2.1. When readers encounter ambiguous sentences such as (6a), some studies have found that they preferably choose to attach PP to a distant site, as readers can parse the sentence with less nodes than in NP interpretation according to the minimal attachment in both L1 (e.g., van Gompel et al., 2001) and EFL (Ying, 1996). This preference is predicted by *minimal attachment* in the garden-path model, which describes the universal tendency of the parser to build the simplest structure as long as the rules of the language allow (e.g., Fraizer & Rayner, 1982). However, several research studies have denied the VP preference in PP sentences. One such study, Frenck-Mestre and Pynte (1997), compared the reading time of ambiguous PP sentences between L1 readers and bilingual (L2) learners, the result being that L2 learners preferred to attach incoming information to the most recently processed constituent. These results were supported by *late closure* (Fraizer & Rayner, 1982) and *recency*, which were originally intended for native English readers. Other factors such as the definiteness of nouns in NP affected readers' attachment decision (Spivey-Knowlton & Sedivy, 1995).

In temporary ambiguous sentences such as Sentences (6b) and (6c), previous studies have mostly examined attachment preference using on-line data such as reading time (for L1 reader, Schütze & Gibson, 1999; Spivey-Knowlton & Sedivy, 1995; Thornton et al., 2000; for L2/EFL, Nakanishi, 2012; Rah, 2009; Yokokawa, 1994) and eye movement (for L1 readers, Boland & Blodgett, 2006; Frenck-Mestre & Pynte, 1997), as shown in Table 2.1. Some of the studies decided the preferred attachment sites that sentence includes higher comprehension in off-line tasks (Nakanishi, 2012). When the reading or fixation times in the target region (i.e., disambiguated or hereafter regions) in Sentence (6b) were faster than (6c), for example, the result suggests that the readers

initially construct the VP-preference interpretation in their minds. Previous studies have not achieved a consensus on preferred attachment sites. On the one hand, some L1 studies have claimed that the VP-attachment preference is due to minimal attachment principle, which is similar to that of ambiguous sentences (van Gompel et al., 2001), and other EFL studies have replicated the findings (e.g., Nakanishi, 2012; Ying, 1996; Rah, 2009). On the other hand, some studies have discussed the NP preference in the biased sentence (Schütze & Gibson, 1999; Spivey-Knowlton & Sedivy, 1995). Their research designs were peculiar to the verb manipulation of PP-attachment sentences. The EFL study that did not find minimal attachment also used the some types of verbs (Yokokawa, 1994). Both studies with VP or NP attachment evidences were based on principles of the garden-path models, but the results were not consistent with each other. Furthermore, the shallow structure hypothesis predicted the null preference in PP-attachment sentence processing; the preferred attachment sites have been confounding.

Some other factors affect the attachment preference of PPs. One of the most important factors is a matrix verb type of sentences (Kidd & Bavin, 2005; Spivey-Knowlton & Sedivy, 1995): When PP-attachment sentences include the preposition *with*, VP attachment preference becomes strong when sentences include action verbs (e.g., wave), while NP attachment preference is dominant in sentences with perception and psycho verbs (e.g., see). One PP-attachment sentence study found that L1 readers preferred high attachment when the effects of verb were eliminated (Kim, Carbary, & Tanenhaus, 2013). They used PP-attachment sentences whose attachment sites were both nouns (e.g., *The FBI agent noticed* <u>*the mirror*</u> *on* <u>*the wall*</u> <u>*with crack*</u>). The primary purpose of the study was not to determine the preferred attachment sites but to examine the effects of comprehension priming task without lexical overlaps. Their study conclude that verb-specific information is essential to construct sentence representation.

Comprehenders not only process input information but also predicate the next elements such as transitivity and frequency data (Arai & Keller, 2013; Hijikata, 2012). Referential information such as definiteness of object NP and NP in PP (i.e., the object NP and NP in PPs as shown in Sentence 6b; The criminal shot the approaching dog with the pistol yesterday late at night.) also affect readers' decision of attachments (Spivey-Knowlton & Sedivy, 1995). When the NP includes the definite article, readers indicate the VP attachment preference, while the NP with an indefinite article leads the NP attachment preference, though the effect of verb types is dominant over effect of NP definiteness. Other referential information is discourse context before PP-attachment sentences.

Compared with frequent data, however, there are inconsistencies between the frequencies of the corpus and readers' behavioral data (Gibson & Pearlmutter, 1998; Spivey-Knowlton & Sedivy, 1995). Spivey-Knowlton and Sedivy searched the Brown Corpus for PP attachment sentences using the term *with*, which indicated that the use of NP attachment was found twice more than that of VP attachment. Gibson and Pearlmutter (1998) also found that on-line psycholinguistic results were not identical to Brown corpus frequencies when they focused on the argument/adjunct status of PP attachment.

Table 2.1

Summary of L1 and L2/EFL Studies With PP-attachment Sentences

Study	◈ Participants, ☐ materials and ■ methodology/task	Main findings
	VP-attachment preference with L1 readers	
Thornton et al. (2000)	◈ L1 university students ☐ Syntactically ambiguous sentences with PP-attachment (VP, NP) ☐ Manipulation of length of object NPs (i.e., distance between main verb and PP) ■ Reading time data using word-by-word self-paced reading	The interaction between length and attachment sites appeared: In the short condition, the VP conditions were processed faster than the NP conditions, but superiority of VP conditions disappeared in the long condition.
Trueswell, Sekerina, Hill, & Logrip (1999)	◈ L1 children and adult readers ☐ Syntactically ambiguous sentences with PP-attachment *Put the frog on the napkin in the box*. (Ambiguous) *Put the frog that's on the napkin in the box*. (Unambiguous) ☐ Participants listened to the target sentences and were given one or two visual contexts (pictures) as a cue. ■ Visual eye-tracking study	L1 children initially interpret the PP as VP-attachment preference, and their representation were not modified when the additional visual contexts that disambiguate sentences as NP-attachment were given.
van Gompel et al. (2001) Experiments 1 and 2	◈ L1 university students ☐ Forced disambiguated PP-attachment sentences (only VP attachment) for Experiment 1 ☐ Syntactically ambiguous sentences with PP-attachment for Experiment 2 ■ Eye-tracking study	In Experiment 1, L1 readers processed the VP-attachment sentences and they had difficulties in processing interpretation as a NP attachment sentence. They processed easily (rapidly) in VP-attachment sentences using fully ambiguous sentences in Experiment 2.

(Continued)

Study	◈ Participants, ☐ materials and ■ methodology/task	Main findings
	VP-attachment preference with L2/EFL learners	
Nakanishi (2012)	◈ Japanese EFL university students ☐ Syntactically ambiguous PP-attachment sentences *The woman visited the man / (early this morning) / with a car / instead of / walking*. (minimal attachment) *man with a bike*. (Non-minimal attachment) ■ Self-paced reading	Sentence comprehension performance was affected by the learners' working memory capacity, but their decisions were consistent with minimal attachment, regardless of working memory capacity.
Rah (2009) Experiment 6	◈ German L2 learners of English and L1 readers ☐ Syntactically ambiguous RC and PP-attachment sentences *Lily offered the CD that she had stolen to John once again.* (high) *Lily offered the CD that she had stolen from John once again.* (low) ■ Self-paced reading	Learners processed attachment sentences of RC and PP like L1 readers. The results did not show the recency preference that is specific to learners.
Rah & Adone (2010a)	◈ L1 readers and German L2 learners (intermediate and advanced groups) ☐ Syntactically ambiguous PP-attachment sentences ☐ Manipulations of attachment sites (NP, VP), verb types (transitive, ditransitive), and length (short, long) ■ Sentence complement task that provided subjects and main verbs of sentences	Sentence production of L1 and L2 readers indicated the overall VP-attachment preference that was consistent with sentence comprehension studies. Learners with lower proficiency showed slightly different results, but the effect of length was weak in EFL sentence production.
Ying (1996) Experiment 1	◈ ESL university students ☐ Fully ambiguous sentences with PP-attachment (VP, NP) without contexts ■ Off-line preference task	Participants chose VP attachment more frequently than NP attachment. They preferred the VP attachment structures in ambiguous sentences.

(Continued)

Study	◈ Participants, □ materials and ■ methodology/task	Main findings
	NP-attachment or null preference with L1 readers	
Schütze & Gibson (1999)	◈ L1 readers □ Syntactically ambiguous PP-attachment sentences *The company lawyers / considered employee demands / for a raise* (NP argument), *for a month* (VP modifier) */ but they…* ■ Self-paced reading	Each condition was highly comprehended. NP arguments were read more rapidly than VP modifier attachments. VP attachment preference was not determined, and thus minimal attachment was not supported.
Spivey-Knowlton & Sedivy (1995)	◈ L1 readers □ Syntactically ambiguous PP-attachment sentences □ Effects of definiteness of NPs (referential information) and verb types (lexical information) on attachment preferences *The fireman / smashed down / the (a) door / with the (a) rusty lock.* (NP attachment) *with the heavy axe.* (VP attachment) ■ Corpus analysis, self-paced reading, and sentence complement task with ambiguous sentences	Verb types (lexical information) decided attachment preference, and the strong VP-attachment preference was due to lexical information. The definiteness overrode NP-attachment bias induced by verbs. Readers used the lexically specific bias and referential presupposition at the same time. Thus, this study supported the constraint-based model.
	NP-attachment or null preference with EFL learners	
Frenck-Mestre & Pynte (1997) Experiment 1	◈ French L2 learners of English (bilinguals) and monolingual speakers of French □ PP-attachment sentences in French with monotransitive (1-complements are needed) and ditransitive verbs ■ Eye-tracking data	Bilinguals tended to resolve local ambiguities to attach incoming information to the most recently processed constituent; though this decision was affected by lexical information of verb.
Yokokawa (1994)	◈ Japanese EFL university students □ Syntactically ambiguous PP-attachment sentences with preposing context (neutral, VP-biased, and NP-biased) *The policeman saw the spy with a telescope.* ■ Reading time data using word-by-word self-paced reading	The biases towards VP attachment were not always determined. The effects of contexts appeared according to the sentence structure.

Interactions between syntactic and semantic processing.

In order to examine L1 readers' syntactic and semantic or pragmatic processing, Thornton et al. (2000) added information to phrases intervening between attachment sites and PP using the self-paced reading method in word-by-word style. Phrase length is one of the factors that complicates syntactic parsing and causes heavy processing loads (Gibson, 1998). Participants read four types of sentences including two attachment sites and two length conditions:

(7a) The salesman glanced at the [amazingly rude young (Long condition)] customer with <u>suspicion</u> and then walked away. [VP attachment]

(7b) The salesman glanced at the [amazingly rude young (Long condition)] customer with <u>ripped jeans</u> and then walked away. [NP attachment]

In the long version which was in parentheses above, readers gained additional information for NP just before the targeted PP, but the syntactic construction became complex at the same time. The findings suggest that participants read VP-attachment sentences faster than NP-attachment sentences in cases where the sentence is short, but the VP preference disappears in the long-version sentence. This interaction is regarded as the distance effect. In addition, readers detect anomalies in long condition sentences that might be attributed to the length effect.

However, EFL studies have attained different results. Nakanishi (2012) focused on the relationship between length effect and EFL learners' working memory capacities, although the additional phrases in that study did not provide informational contents. His experimental sentences intervened between both the attachment sites and PP in the long-sentence condition:

(8a) The doctor / examined / the child / [last Monday morning] / with a pen / instead of / another tool. [Minimal attachment condition]

(8b) The doctor / examined / the child / [last Monday morning] / with a pen / instead of / the child without a pen. [Non-minimal attachment condition]

The participants read the sentence in phrase-by-phrase style; the slashes in Sentences (8a) and (8b) indicate the region that presented on computer screen. The results indicated no length effect on processing, and they also found that EFL learners showed VP preference in the same manner as in L1 studies. Since Nakanishi's study did not intend to investigate the interaction between length and attachment sites as discussed in Thornton et al. (2000), it was natural to lack the distance effect that dominance of VP attachment preference disappeared with the long phrase. However, it was surprising that he did not find the length effects attributable to the long phrase. The possible reason is that Nakanishi's study adopted the phrase-by-phrase reading style, and the participants easily recognized the long part as a phrasal chunk. In addition, since the materials was not identical to the L1 study, the results could not be compared. As seen so far, the relationship between informational contents and phrase length for EFL learners has been rarely investigated. Because using pragmatic or semantic cues is imperative for successful syntactic parsing of L2/EFL learners, it is necessary to examine length manipulation with some semantic information added to the original sentences.

RRC sentence studies have suggested that attachment preferences are affected by learners' proficiency (Dekydtspotter, et al., 2008). The study used the French relative-clause attachment sentences and preposing context information to choose ambiguous attachment sites targeting American learners of French. When they read contexts that

disambiguated sentence as high attachment, they initially attached to low phrases in order to minimize parsing costs in on-line processing. Learners with high proficiency modified their initial low-attachment interpretation into high attachment in off-line processing (i.e., comprehension questions), but low proficient learners maintained their unrevised and wrong interpretations. These results suggest that attachment preferences cannot be generalized but might be different from learners' proficiency.

2.3 Effects of Post-reading Task on Sentence Comprehension

As reviewed in Section 2.1.4, on-line sentence processing studies generally use dual tasks to investigate readers' sentence comprehension (e.g., end-of-sentence verification questions, grammatical judgment tasks, translation tasks, and stop-making-sense tasks). As comprehension is goal-driven process, the goal of comprehension affects retrieval of the specific information and global comprehension (Grabe & Stoller, 2002; Jegerski & van Patten, 2014; Lim & Christianson, 2013; Swets, Desmet, Clifton, & Ferreira, 2008; Traxler, Pickering, Clifton, 1998). In particular, when the dual task requires performing meta-linguistic tasks as well as mere input processing, differences of readers' or learners' linguistic proficiency or cognitive factors become prominent (Roberts, 2012).

Readers are affected by comprehension questions or other tasks after reading sentences, which is usually a self-paced reading task. Reading comprehension questions regarding ambiguous regions draw readers' attention toward resolving ambiguities, while readers intentionally leave ambiguous attachment sites unresolved in general comprehension questions in which they do not have to resolve ambiguities. This strategy is called *underspecification* (Lim & Christianson, 2013; Swets et al., 2008). Swets et al. (2008) examined the effects of comprehension questions on sentence comprehension,

especially focusing on the frequencies of providing comprehension questions to readers. Their study provided three conditions: (a) In the relative clause condition, participants had to answer questions regarding the attachment sites of relative clauses (targeted structures) in the RC-attachment sentences, and relatively difficult comprehension questions in filler sentences; (b) In the superficial condition, comprehension questions irrelevant to the RC structure were provided in every trial; and (c) In the occasional superficial condition, superficial comprehension questions appeared randomly. The results showed that the readers did not determine the attachment sites unless they were required to answer; this strategy involves the faster processing of ambiguous sentences than that of unambiguous sentences as well as the lack of resolution of ambiguity when only surface comprehension is needed. This is supported by the good-enough theory (Traxler et al., 1998) and the construal model (Fraizer & Clifton, 1998), indicating that it is ineffective to disambiguate sentences with little information: when they have to reanalyze or reconstruct syntactic construction with additional information that can disambiguate sentences, the cost of their initial interpretation becomes meaningless.

It should be noted that not only the tasks after reading the sentences, but also the research methodology during reading affects readers' comprehension. These effects were sometimes regarded as a methodological limitation. For example, word-by-word self-paced reading task violates readers' natural chunk boundaries and disturbs their sentence comprehension (Yamashita & Ichikawa, 2010). The problem of reading time data has been resolved with the eye-tracking study. The think-aloud method also interferes with readers' comprehension. Since the think-aloud method draws readers' attention to self-monitoring the information that they are currently processing, the method tends to promote comprehension or noticing, as mentioned in Section 2.1.4.

However, since the effects of post-reading questions on sentence comprehension

and processing have not been examined in detail, more studies are required.

2.4 Summary of Past Findings and Purpose of Current Research

As seen so far, previous research on sentence processing in both L1 and L2/EFL have been reviewed in this chapter. Although there are massive research about sentence processing with syntactic and semantic cues, the following four points have not been fully investigated, or not reached a consensus in previous research:

Firstly, learners' syntactic processing is inferior to native speakers, although the issue of whether the difference between them is fundamental or not has been controversial (Clahsen & Felser, 2006a, 2006b; Felser et al., 2003; Juffs, 1998; Lim & Christianson, 2013). In order to compensate for syntactic processing, learners use semantic information. To investigate the relationships between syntactic distance and semantic processing, PP-attachment sentences were utilized in previous studies. In this structure, the site wherein readers attach PPs has been argued upon, for more than 30 years. As for the attachment sites of fully ambiguous PP-attachment sentences, L1 readers prefer VP attachment on the basis of minimal attachment, and this tendency is also seen in unambiguous PP-attachment sentences (i.e., forced disambiguated sentence). The results were based on the minimal attachment principle in the garden-path model. Similarly, readers have VP-attachment preference in both ambiguous and unambiguous sentences, but there are some exceptions or constraints, which make them decide to attach object NP immediately before the PP. If the distance between attachment sites and PP is given to L1 readers, they sensitively notice the distance between constituents in a sentence in on-line processing. The distance or length will also affect learners' sentence processing, but the distance effect in PP-attachment sentences has not been discussed fully. Therefore, this study used PP-attachment sentences to investigate learners' syntactic and semantic processing.

Secondly, as learners' cognitive resources are restricted compared to L1 readers, their sentence processing might differ according to their grammar knowledge size, processing automaticity, and cognitive resources. When sentence processing are different, it is supposed that learners' processing strategies can also differ. Previous studies have determined readers' overall tendency in sentence processing through on-line methods such as reading time or eye-movement, and the effects of individual differences on sentence processing (e.g., Roberts, 2012). However, the effects of individual differences on strategies for processing sentences has not been discussed enough. This study focused on the individual differences that affect learners' sentence processing strategies.

Thirdly, in order to observe readers' strategy for on-line processing, the previous studies utilized self-paced reading or eye-tracking data. These data could not provide direct readers' representation, while off-line tasks such as translation cannot visualize how readers construct sentence representations in progress. To resolve such dilemmas, the present study adapted the think-aloud method to observe readers' strategy use of sentence processing.

Forth and finally, a few studies have investigated the effects of the types of post-reading tasks or questions on sentence comprehension performance. Though the investigations into post-reading tasks are highly suggestive for EFL classrooms or the acquirement of sentence processing skills, many studies regarded these tasks and questions as a measurement of comprehension. The present study will utilize two types of tasks and compared learners' sentence processing performance.

To resolve the problems discussed above, four experimental studies and pilot studies were conducted in this research. Based on findings from previous studies, the present research focuses on learners' use of syntactic and semantic cues using attachment-

structure sentences in off-/on-line processing. Study 1 investigates the differences between off- and on-line processing of Japanese EFL learners. Study 2 examines the detailed processing and strategy use of learners using the think-aloud method.

The purposes of this study will be to:

(1) investigate how learners process and use semantic information to decide attachment sites of PP-attachment sentences in off- and on-line processing,

(2) explore how learners recover from initial incorrect interpretations in global and local ambiguous sentences,

(3) examine the effects of learners' individual differences on sentence comprehension,

(4) determine the effectiveness of the think-aloud method to observe learners' detailed representation while sentence processing, and

(5) investigate the effect of tasks after reading on learners' sentence comprehension.

Chapter 3

Study 1: Off- and On-line Processing of PP-attachment Sentences of EFL learners

3.1 Experiment 1: Off-line Processing of PP-attachment Sentences

3.1.1 Purpose and research questions of Experiment 1

The aim of the present study is to investigate the relationship between semantic and syntactic cues for EFL learners. To investigate these relations, the study adapts the four sentence structures in Thornton et al. (2000) with two different attachment sites and two length manipulations as shown in Figure 3.1. These sentences are force-disambiguated, and readers decide the attachment site of PP using semantic information. The double underlined phrase appears only in the long conditions; the arrows indicated the attachment site of each attachment condition. The solid arrow indicates VP attachment and the dashed arrow indicates NP attachment.

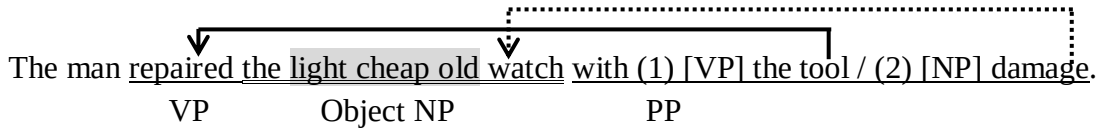

Figure 3.1. Example material in Study 1.

In Thornton's study, L1 readers performed better in VP condition without the shaded additional three words, while differences between attachment conditions disappeared when three words were given. The distance effect occurs when the interaction between attachment and length suggests the least performance in VP-long condition, for the distance between attachment site and PP becomes long only in VP-long condition. In NP-long condition, the distance between attachment site and PP is the same as in NP-

short condition. The present study predicts that the length manipulation with informational contents will negatively affect learners' sentence comprehension due to the length effect that was reported in Thornton et al. (2000).

The reason why this study focuses on *with*-PP attachment sentences was twofold: The first one was concerned with the simpleness of the structure. Apart from such syntactic structures as relative clause attachment, learners rarely have difficulties in syntactically parsing PP attachment, so the effect of an added part (i.e., length part) will be evident. The simple structure will prevent learners from misunderstanding sentence structures even in the short condition. Another reason regards as the argumenthood of PP. Some PPs are used as a modifier, not as prepositional attachment to some previous antecedent (e.g., *in the night*). The processing of these adverbs differed from usual PPs at the initial processing according to the L1 reading time study (e.g., Gibson & Pearlmutter, 1998; Schütze & Gibson, 1999). In order not to make differences between the attachment conditions, this study chose PP-attachment structure using *with*.

To investigate the learners' syntactic processing with semantic cues, this study focused initially on the off-line processing data. For these purposes, the following three research questions were addressed:

RQ1-1: Do EFL learners show attachment preference in off-line processing?

RQ1-2: Do EFL learners sensitively respond to material distance in off-line processing?

RQ1-3: Do the question types affect EFL learners' off-line processing?

As for RQ1-1, native speakers of English and L2/EFL learners prefer the VP-attachment preference according to the prior studies, and performed better than in NP-attachment

condition. Therefore, it is predicted that EFL learners in this study also prefer the VP attachment, though attachment-decision process of learners' is likely to differ from that of native speakers. Lim and Christianson (2013) pointed out that outputs of off-line task between L1 and learners had similar reading pattern, because learners' difficulties in syntactic parsing were due to slow retrieval of syntactic information.

Regarding RQ1-2, effect of distance between attachment site and PP has been examined by on-line task in previous studies, especially reading time paradigm, while studies of off-line tasks that focused on the length effect using PP-attachment sentence have not yet been seen. The previous studies that do not support or partially support the shallow structure hypothesis concluded that learners showed the same reading pattern as L1 readers, but the appearance of the pattern was not so immediate as L1 readers (Jackson, 2008; Lim & Christianson, 2013; Rah & Adone, 2010b). Considering these findings, off-line tasks for learners will show the length effect (i.e., the simple main effect of length in VP-attachment sentence) as in on-line task results for L1 readers.

RQ1-3 focuses on the types of comprehension questions after reading: preference task and translation task. Both of them are used to investigate readers' encoded sentence representation, but some differences are predictable: Preference task asks directly for the attachment site of PP, so readers have to parse the structure in question. Translation task, on the other hand, will draw readers' attention to semantic representation of whole sentences.

3.1.2 Methods of Experiment 1

3.1.2.1 Participants

A total of 27 Japanese EFL undergraduates and graduates participated in Experiment 1. All of them were native speakers of Japanese, and had received formal

instruction in English for more than six years in Japan. They had intermediate and advanced levels of English proficiency according to their reports. Their majors varied. They participated in this study on a paid basis, and all of them were unaware of the purpose of this experiment. One of the participants failed to complete the tasks, so his data was excluded from the following analysis. The remaining data of the 26 participants were analyzed.

3.1.2.2 Material sentences

Four types of sentences were prepared in this study: Length (Short and Long) × Attachment (VP and NP attached). The participants read 79 sentences in total, which included 39 experimental sentences and 40 filler sentences. They were given one of four counterbalanced material sets, and thus encountered eight or nine sentences in each condition (for the experimental sentences, see Appendix A). Short sentence pairs of NP and VP conditions were derived from previous studies of prepositional phrase attachment (O'Brien et al., 2014; Rayner, Carlson, & Frazier, 1983; Schütze & Gibson, 1999; Spivey-Knowlton & Sedivy, 1995; Taraban & McClelland, 1988; Thornton et al., 2000; van Gompel et al., 2001). Since some of these sentences targeted English L1 readers, the sentences were modified into simple ones based on the following criteria: (a) the sentence excepting PP should be identical between NP/VP conditions, (b) the number of words in a whole sentence is approximately 10 words, (c) when context is attached to the original sentence, the experimental sentences should be independent without any context, (d) articles of NP in both object NP and PP were modified to definite article when possible and (e) the words in a sentence should be easier than the level of 3,000 words in JACET 8,000 (JACET, 2003).

To create long-version sentences, three words (three adjuncts, or one adverb and

two adjuncts) were added before the NP region: this causes the distance between VP and PP in only long condition, so VP-Long condition possessed a greater syntactic processing load than the other conditions (see Tables 3.1 and 3.2). Three added modification words should not affect learners' preference decision between NP/VPs. Two native speakers of English checked these revised sentences.

Table 3.1

Examples of Experimental Sentences

Condition		Sentences	Choices	
Short	VP	The little girl cut the apple <u>with the plastic knife</u>.	(a) girl*	(b) apple
	NP	The little girl cut the apple <u>with the shiny coating</u>.	(a) girl	(b) apple*
Long	VP	The little girl cut the **large juicy green** apple <u>with the plastic knife</u>.	(a) girl*	(b) apple
	NP	The little girl cut the **large juicy green** apple <u>with the shiny coating</u>.	(a) girl	(b) apple*

Note. The target PPs were underlined. The additional phrases in Long condition were shown in bold. The correct answers were denoted by asterisks.

The 44 filler sentences were derived from various previous studies that focused on attachment structures (Frazier & Rayner, 1982; Jun, 2010; O'Brien et al., 2014; Rayner et al., 1983; Schütze & Gibson, 1999; Spivey-Knowlton & Sedivy, 1995; Taraban & McClelland, 1988; van Gompel et al., 2001), while some filler sentences had the same structure as experimental sentences. The filler sentence list included the PP-attachment sentences without *with*, ambiguous sentences with PP attachment, and relative-clause attachment sentences. The filler sentences included the ambiguous and biased PP

sentences using *with* and other prepositions such as *in* and *over*. In addition, the relative clause attachment sentences of both high and low attachment were also used (for the more materials, see Appendix B). These sentences also have two attachment sites as in PP-attachment sentences (i.e., distant and local attachment sites), which makes it possible to adapt the two-choice task.

Table 3.2

Length Counts of Experimental Sentences (Word Number Per Condition)

		DP		VP		NP		PP	
Condition		*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Short	VP	2.13	0.33	1.10	0.30	2.00	0.00	2.03	0.60
	NP	2.13	0.33	1.10	0.30	2.00	0.00	2.28	0.53
Long	VP	2.13	0.33	1.10	0.30	5.00	0.00	2.03	0.60
	NP	2.13	0.33	1.10	0.30	5.00	0.00	2.28	0.53

3.1.2.3 Norming study 1

To verify the plausibility of experimental sentences, the norming study was carried out to target some EFL learners as follows:

Participants.

In the norming study, the plausibility of sentences among conditions was rated by four graduate and under-graduate students who did not take any experiment in the dissertation. All of them majored in English education, indicating that their English proficiencies were upper-intermediate level or higher.

Materials.

The 39 sentence-sets of short condition used in Experiment 1 were passivized using *by*-clause on the basis of Schütze and Gibson (1999) and van Gompel et al. (2001) as seen in examples (6a-d). This transformation of sentences allows for maintaining the meaning and lexical content while removing temporal syntactic ambiguity. In NP attachment condition, PPs always immediately followed the head noun of the subject; in VP attachment condition, they appeared immediately following the passive participle. In addition, to confirm the validities of experimental items, the false versions of the sentences were constructed by replacing *with*-clause with the PP in the other attachment condition, as shown in the following sentences:

(9a) The apple with the shiny coating was cut by the little girl. [NP attachment-true]

(9b) The apple was cut with the plastic knife by the little girl. [VP attachment-true]

(9c) The apple with the plastic knife was cut by the little girl. [NP attachment-false]

(9d) The apple was cut with the shiny coating by the little girl. [VP attachment-false]

Procedure.

Four participants each read four versions of passive sentences (see Sentences 6a to 6d above); they read a total of 156 sentences including 39 sentences for each condition. The participants rated the plausibility of each sentence using a 7-point Likert scale (1: *highly plausible*, 4: *I cannot decide the plausibility of this sentence*, 7: *highly implausible*). The presentation order was randomized, but no two passive sentences in the same set never appeared in a row.

Scoring and analysis.

For each set, six types of rating were calculated for each item set: the rating of (a) NP-attachment-true, (b) VP-attachment-true, (c) NP-attachment-false, (d) VP-attachment-false, (e) result of (a) – (c), and (f) result of (b) – (d). To confirm validation of sentences, a repeated ANOVA test was conducted for average rating scores of (a) to (d). A 2 (Attachment [Within]: NP vs. VP) × 2 (Correctness [Within]: True vs. False) two-way ANOVA was conducted. The item analysis was adapted because the number of participants was limited. In addition, to measure the plausibility between attachment conditions in the same material set, repeated *t* test was conducted for scores of (e) and (f).

Results.

Table 3.3 shows the average rating scores of plausibility. Two-way ANOVA showed the significant main effect of Correctness ($p < .001$), insignificant main effect of Attachment ($p = .321$), and insignificant interaction between Correctness and Attachment ($p = .712$). Another ANOVA for participant analysis was conducted and resulted in the same findings.

Table 3.3

The Average Rating Score of Plausibility With Standard Deviation

Attachment	True condition	False condition	(True) – (False)
NP attachment	5.91 (0.59)	3.40 (1.28)	2.51 (1.28)
VP attachment	5.69 (0.87)	3.31 (0.87)	2.38 (1.62)

Note. Standard deviations are parenthesized. The maximum possible score = 7 (*highly possible*).

To avoid any ambiguities in the experimental sentences, the four sentences were excluded from further analyses, because these were inappropriately biased: These are assumed to cause bias in each condition in the previous studies, but the participants in the norming study did not interpret these sentences as the force-disambiguated condition, as seen in Sentence (10). In the sentence, the verb *admired* were regarded as an unknown word for some of participants.

(10) The couple admired the (really big traditional) house with a garden [NP]/ a friend [VP].

3.1.2.4 Norming study 2

To confirm the lexical cohesion of each condition, the second norming study was conducted. The PP-attachment sentences in Study 1 required readers to use semantic or pragmatic information to disambiguate. If learners could find the apparent lexical cohesion between one of the attachment sites and PP, their decisions would be affected regardless of their syntactic processing. The second norming study was needed to rule out the possibilities that learners did not use the syntactic structure and rely on the lexical information, though it is not likely to happen because attachment sites (i.e., VP or NP) in this study were of various parts of speech.

Procedure and Analyses.

In order to investigate the lexical cohesion or semantic relatedness, Latent Semantic Analysis (LSA) was utilized. According to LSA, the semantic relatedness between terms or sentences are calculated (Landauer, McNamara, Dennis, & Kintsch, 2007).

The LSA scores of the following pairs were computed in each sentence set: (a) verb and PP in VP-attachment condition, (b) object NP and PP in VP-attachment condition, (c) verb and PP in NP-attachment condition, and (d) object NP and PP in NP-attachment condition. A 2 (Attachment [Within]: NP and VP) × 2 (Category [Within]: Verb and NP) two-way ANOVA was conducted for these scores.

Results.

Table 3.4 shows LSA scores between the two phrases. For clarification's sake, Figure 3.2 also demonstrated the result with example sentences. The number denoted the LSA scores between each constituents and PP in the figure. Two-way ANOVA found significant interaction between Attachment × Category ($p < .001$): Subsequent analysis showed the significant simple main effect of attachment in verb region ($p = .001$) and in object region ($p = .014$); significant simple main effect of category in the VP condition ($p < .001$) and insignificant effect in the NP condition ($p = .405$). The results indicated that the scores between VP attachment site and PP were higher than those between NP attachment site and PP in verb region, and vice versa in object NP region. In VP attachment condition, however, the same results as the NP condition were not found out.

Table 3.4

*LSA Scores Between Attachment Sites and PP in the Attachment Conditions (*k *= 39)*

	Verb	Object NP	Long phrase
VP-attachment sentence	21.59 (12.91)	23.97 (14.80)	26.13 (15.40)
NP-attachment sentence	29.03 (13.91)	17.00 (10.93)	29.13 (15.23)

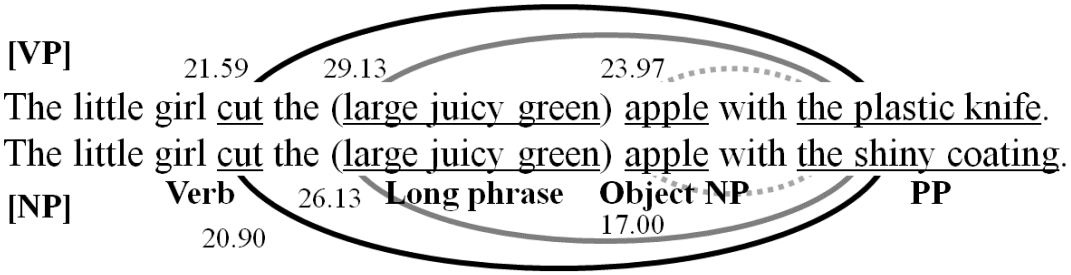

Figure 3.2. LSA scores between each constituent and PP with example sentence.

The results of the two norming studies confirmed that the experimental sentences were appropriately manipulated. Based on these results, the experiment was carried out with the following procedure.

3.1.2.5 Procedure

This experiment was conducted individually or in a group of a few participants. At the beginning of the experiment, the experimenter explained general aims of this study, and also explained that the participants were going to read sentences and perform some tasks. They were not told the detailed tasks beforehand. After that, they were randomly given one of the four booklets of the preference task. In the booklet, the condition and presentation order of experimental sentences were counterbalanced. After finished the preference task, the participants brought back the booklet, and the experimenter gave them another booklet of translation task. At the beginning of each task, a practice session was carried out to confirm the procedure. They were not allowed to use paper/electronic dictionaries, or cellphones during the experiment. The whole procedure took one or one and a half hours depending on participants.

Preference task.

In the preference task, participants read the list of sentences and chose one of

two choices as a modification answer. In the practice session, the pair of sentences of NP and VP conditions was presented as shown in Figure 3.3:

英文の意味を良く考えて、下線部が結びつくものとして適切な解答を以下の選択肢から選び〇をつけてください。
According to the meaning of the sentences, choose the most appropriate option to which the underlined words attach.
1: Grandmother didn't see any articles <u>on Jupiter</u>. (a) Grandmother (b) article
2: Grandmother didn't see any articles <u>on the flight</u>. (a) Grandmother (b) article
Note that more than one sentence in the same set will not be presented repeatedly in the preference task session.

Figure 3.3. Sample instructions and materials of practice session.

After the experimenter confirmed the participants' responses, feedback was provided (i.e., the correct answers are 1 = b, and 2 = a). The reason for the answer was always explained to make them understand the procedure of the proceeding preference task, regardless of whether the participants answered correctly or not.

Sometimes, sentence ambiguity made it impossible to decide which one the PP should modify, in which case they were asked to report that the sentence was ambiguous after they chose one of two options (i.e., an ambiguity decision report) instead of opting for nonresponse. Note that the normal preference task in ambiguous-sentence studies does not have any correct or incorrect answers, because the sentence is ambiguous; the preference task in Experiment 1, however, had a correct answer to choose. The participants were expected to answer the subject of the VP-attachment sentence in VP

conditions and the object of the NP-attachment sentence in the NP conditions based on Yokokawa (1994). The main preference task did not provide feedback to participants contrary to the practice session. After the participants had completed the preference task, they were asked to report unknown words in a whole sentences list by circling them.

Translation task.

After the preference task, the translation task in Japanese was conducted in order to confirm learners' experimental sentence representation. The experimenter did not tell the participants the existence of this task beforehand. The experimental sentences were presented in lists, but their presentation order was re-randomized to differ from the previous preference task. The participants wrote down the translation of each sentence in Japanese and also the reason for miscomprehension when they could not reach the final interpretation or they had difficulties in sentence comprehension.

3.1.2.6 Scoring

Working individually, two raters scored the same 30% of the products in the translation task based on the following criteria: One point was given when the translation products included the correct structure of PP, even if learners made mistakes in comprehending the meaning of some words, following the example of van Gompel et al. (2001). This study added the following scoring criteria: (a) learners' slight mistakes regarding verb were accepted, (b) when learners confused the preposition (i.e., *with*) with another one, the responses were not regarded as a correct structure of PP. The other structures such as additional phrases in long condition were not targeted. Scoring inconsistencies between the two raters were discussed to decide the revised criteria. Using this criteria, one of the two raters scored the remaining 70% of data.

3.1.2.7 Analyses

Three types of 2 (Attachment; NP vs. VP) × 2 (Length; Short vs. Long) two-way ANOVAs of repeated measures were conducted for (a) the number of correct answers, (b) the score of the translation task, and (c) ambiguity decision report in the preference task. The significant *p*-value in this dissertation is .05; the marginal significance value will be taken into consideration depending on the effect size (if any).

3.1.3 Results of Experiment 1

3.1.3.1 Preference task

Table 3.5 shows the descriptive statistic of the ambiguous decision in the preference task. The overall number of reporting ambiguity decision was quite low, suggesting that the experimental sentences satisfied each condition and, also, that each sentence did not include ambiguities. The results of ANOVA for the ambiguity decision scores did not indicate any significant main effects; $F(1, 25) = 0.03$, $p = .861$, $\eta^2 = .00$ for Attachment; $F(1, 25) = 2.11$, $p = .159$, $\eta^2 = .08$ for Length; nor significant interaction, $F(1, 25) = 0.61$, $p = .443$, $\eta^2 = .02$. The remaining number of sentence was 35 based on the two norming studies, and each condition contained eight or nine sentences, as a result.

The overall correct answer rate was relatively high, suggesting that the participants comprehended the sentences with the semantic information. As shown in Table 3.6, the ANOVA for the scores of the preference task showed insignificant interaction between Attachment and Length, $F(1, 25) = 2.68$, $p = .114$, $\eta^2 = .01$; the significant main effect of Attachment, $F(1, 25) = 4.55$, $p = .043$, $\eta^2 = .07$ and insignificant main effect of Length, $F(1, 25) = 1.52$, $p = .229$, $\eta^2 = .03$.

This finding that learners performed best in the NP conditions was somehow unexpected, because many previous studies predicted the initial VP preference among

English native readers and L2/EFL learners.

On the other hand, some studies demonstrated that L1 readers and EFL learners preferred the NP-attachment sentence, or showed no preference (Yokokawa, 1994). The other unexpected result was that the EFL learners answered correctly more often in long sentences than in short sentences. The possible reason is that learners might not have such difficulties in processing additional phrases in the long condition, as they could read back to VP or NP freely even after disambiguating PP.

Table 3.5

*Descriptive Statistics of Correct Rates in Preference Tasks (*N = *26)*

		Preference task (%)		Translation task (%)	
		M (*SD*)	95%CI	*M* (*SD*)	95%CI
Short	VP	82.22 (14.84)	[76.23, 88.22]	82.05 (14.92)	[76.02, 88.08]
	NP	85.64 (10.60)	[81.36, 89.92]	81.41 (16.83)	[81.15, 91.42]
Long	VP	83.21 (17.89)	[75.98, 90.43]	86.28 (12.72)	[75.98, 90.43]
	NP	91.79 (10.31)	[87.63, 95.96]	88.12 (13.63)	[82.61, 93.63]

3.1.3.2 Translation task

The descriptive statistics of the translation task shows that learners comprehended the sentences well (also see Table 3.5). According to Table 3.6, the results of ANOVA for the translation task's scores differed from that of the preference task. It showed no significant main effect of attachment, $F(1, 25) = 0.09$, $p = .763$, $\eta^2 = .00$; or interaction $F(1, 25) = 0.27$, $p = .607$, $\eta^2 = .01$; however, marginal significance was found in the main effect of length, $F(1, 25) = 3.49$, $p = .073$, $\eta^2 = .12$. The results showed no attachment preference and high scores in the short sentence regardless of attachment

condition (see Figure 3.4).

Table 3.6

*Summary Table for Two-way ANOVA of the Effects of Attachment and Length on the Preference and Translation Tasks (*N *= 26)*

Source	*Df*	*SS*	*MS*	*F*	*p*	η^2
		Preference task (ambiguity decision report)				
Attachment (A)	1	0.00	0.01	0.03	.861	.00
Length (L)	1	0.01	0.01	2.11	.159	.03
A × L	1	0.00	0.00	0.61	.443	.01
Error (A)	25	0.11	0.01			
Error (L)	25	0.13	0.01			
Error (A × L)	25	0.16	0.01			
		Preference task (accuracy rate)				
Attachment (A)	1	0.09	0.09	4.55	.043	.07
Length (L)	1	0.03	0.03	1.52	.229	.03
A × L	1	0.02	0.02	2.68	.114	.01
Error (A)	25	0.52	0.02			
Error (L)	25	0.31	0.01			
Error (A × L)	25	0.29	0.01			
		Translation task				
Attachment (A)	1	0.00	0.00	0.09	.763	.00
Length (L)	1	0.08	0.08	3.49	.073	.06
A × L	1	0.00	0.00	0.27	.607	.00
Error (A)	25	0.25	0.01			
Error (L)	25	0.56	0.02			
Error (A × L)	25	0.37	0.02			

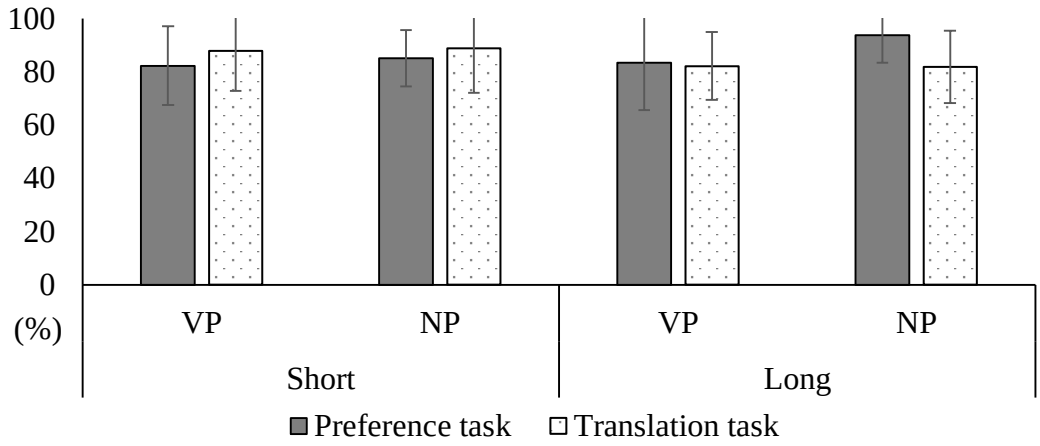

Figure 3.4. Accuracy rates of the preference and translation tasks (%).

In order to figure out reasons for learners' incorrect answers in the translation task, error analysis was conducted. The errors were categorized into five types: (a) misunderstanding of verb, (b) object NP, (c) any word in PP, (d) problems of syntactic structure including usage of *with*, and (e) no response regarding decision of attachment site, as shown in Table 3.7. Some responses were based on the participants' report, and others were categorized by raters. Typical Error (d) included the cases in which learners interpreted all words appropriately but construction was incorrect, and this type of errors was discriminated from (a), (b), and (c). Error (e) indicated the answers that did not refer to the attachment site in translation or did not report the reason why they could not answer. When multiple factors were possible to lead to errors in translation, the raters labeled the cause most likely to affect answers. Table 3.5 also shows descriptive statistics of translation scores. The following part discusses overall tendencies and shows some examples in each cause. In addition, some sentences with specific error pattern will be investigated in detail. The original responses of the translation tasks in this section are provided in Appendix C. The sentence number denoting the experimental sentences corresponds with the number in the sentence list of Appendix A.

Table 3.7

The Number of Error in Each Cause of Translation Task in Experiment 1 (%)

Condition		A	B	C	D	E
Short	VP	28.30	13.21	33.96	15.09	0.00
	NP	12.96	14.81	33.33	29.63	0.00
Long	VP	24.49	8.16	26.53	32.65	0.00
	NP	15.38	12.82	33.33	30.77	2.56
Total		20.51	12.31	25.64	26.67	0.51

Note. *N* = 26.

Error type A: Troubles of verbs.

Errors occurred in the VP condition almost twice as much as in the NP conditions. The results suggest importance of verb in sentences. Especially, Sentence (11a) was a typical case that had more errors in the VP condition, and did not include any error in the NP condition. In the translation answer of Sentence (11b), the participant seemed to confuse meaning of the verb and compensate for the sentence meaning using another verb. As other participant made error, verb *follow* might work as an initiating factor of errors.

(11a) The tourists followed the (wild hungry yellow) fox with the car. [Sentence 20]

(11b) The tourist guide brought the wild hungry yellow fox to somewhere by his/her car. [Translation 1 in VP-long condition]

In Sentence 36, some participants produced errors of verb in only VP-short condition (*k*

= 3). Two of the participants who wrote incorrect translations tried to interpret VP-attachment, though they reported that they did not know the meaning of the verb *glance*. The perception verb induced NP-attachment interpretation according to the previous studies, but some participants did not know the meaning of a verb, and thus the predominance of NP-attachment was eliminated. Their answers' translations are shown below:

(12a) The salesman glanced at the customer with caution/dirty shirt. [S 36]

(12b) The salesman with attention (do something) to the customer… /I do not know the word *glance*. [VS]

(12c) The salesman draw the customer's attention. /I did not know the word *glance*. [Translation 4]

Note that the protocols after the slashes mean the participants' own comments about the reason they hardly understood the sentence. These comments were not included as a translation answer. In Examples (12b) and (12c), the participants expected that the PP should attach to the verb, though they did not know the meaning of the verb.

Error type B: Troubles of objective NP.

Object NPs were attachment sites of PP in the NP conditions. If readers did not know some words in object NPs, they could possibly make more errors in the NP conditions than in the VP conditions. However, the rate of error due to object NPs was relatively low even in the NP conditions, and the difference between the NP and VP conditions was not salient, unlike Error type A, which is also the attachment site of the VP condition. The many errors in Type B, trouble of objectives, appeared in Sentence 35.

In Sentence 35, some participants had difficulties in comprehending the word *crown* in the object NP regardless of attachment conditions, while this sentence contained other kinds of problems (in detail, see section of Error type D). One of the examples is shown:

(13a) The king saw the (remarkably beautiful expensive) crown with the glasses/large diamonds. [Sentence 35]

(13b) The king showed the large diamond to the queen. [NS condition]

In Example (13b), the participant confused the meaning of *crown* with queen. While the main verb *saw* is a familiar word for EFL learners, the participant's comprehension seemed to be affected by the objective NP. Even when the sentence included queen and not crown, his translation was still incorrect. It is possible that the learner attempted to construct the sentence representation using the lexical information that he grasped from the sentence, regardless of the original sentence structure.

Error type C: Troubles of PP.

Words in PP were different between attachment conditions, so that errors concentrated in either one of the attachment conditions. For example, NP conditions of Sentence 17 include many errors due to the word *cracks* in PP. Note that the tendencies between the length conditions were different: Most of the answers did not mention the attachment sites or any interpretation as in Example (14b).

(14a) The bad boy damaged the (costly nice delicate) wine glass with the cracks. [the NP conditions of Sentence 17]

(14b) The bad boy hit and damaged the wine glass. / I do not know the word

cracks. [NS condition]

(14c) The bad boy damaged (something) using *cracks*. / I do not know the word *cracks*. [NS]

(14d) The bad boy damaged the costly and very delicate wine glass at the place where clerks were around. / I do not know the phrase *the costly nice*. [NL]

In Protocol (14c), the learner interpreted cracks (PP) as a VP attachment, as he was not familiar with the word. The participant who answered (14d) reported that his unknown words were only modification parts, but he also confounded the word *cracks* with *clerks*. The next example (15a) includes errors due to words in PP in each attachment condition. The PP words in two attachment conditions seemed to be equally unfamiliar; thus, the biased attachment preferences or tendency were not found out:

(15a) The old man listened to the important funny German opera with six acts/a hearing aid. [Sentence 38]

(15b) The old man listened to the opera using his sixth sense. [NS]

(15c) The old man listened to the opera which was played by six actors. [NL]

(15d) The old man, with the aim of listening, listened to the opera. [VS]

(15e) The old man listened to the opera which was sung by the deaf German. [VL]

In these examples, (15c) and (15e) implied NP-attachment interpretation, and (15b) and (15d) implied VP-attachment interpretation. Therefore, Protocols (15c) and (15d) were correct in the attachment sites, but their sentence representations were incorrect. The other patterns are “the opera with something’s aid, the opera which had a healing effect (voice),

or the man with his wound healing" (VP) and "the opera with six cases" (NP). Some of them were obviously affected by the original words in PP, but they could not amend their somehow weird interpretation using contextual information that general passages usually possess. In sum, learners appeared to be sensitive to the semantic or lexical information of each phrase, but not to the structural effects.

Error type D: Trouble of structure.

In this type, the answers showed that participants probably knew the meaning of each word, yet they produced serious errors in translation since they could not understand sentence structures. These few errors occurred not because of ambiguities of attachment sites, but the structure interpretation itself. A mistake in example (16b) of Sentence 13 was caused by the active or passive. The verb *kill* was possibly a familiar and clear word for learners, as it is an intransitive word both in English and Japanese. The possible reason for the learner making an error is that the additional long phrase might have affected the construction of sentence representation. The long phrase implied that *lion* in the object NP might be stronger than human, and the participant might have determined that the representation of the hunter killed by the lion was more plausible than the lion killed by the hunter. In this case, the learner placed priority on semantic information, or plausibility; plausibility interfered with comprehension, and so the final response is not consistent with the original sentence (Ferreira & Patson, 2007; Lim & Christianson, 2013).

(16a) The hunter killed the (horribly big dangerous) lion with the big face/gun. [Sentence 13]

(16b) The hunter was killed with the gun by the big dangerous lion. [VL]

(17a) The king saw the (remarkably beautiful expensive) crown with the large diamonds (the glasses). [Sentence 35]

(17b) King was remarkably beautiful and expensive, and he saw the crown which was stuck to glasses. [VL]

In contrast, (17b) showed that learners did not attempt to visualize their sentence representation. As a result, the protocol in L1 was quite meaningless. Participant seemed to compose this Japanese sentence using the translation of original words in the sentence regardless of plausibility in this case.

Error type E: Non-response.

The task instruction said that participants should report why they have difficulties in comprehending each sentence; thus many of them managed to point out the reason even if they had not reached the full interpretation. Therefore, the number of errors due to non-response was just one.

The error analyses of translation were summarized as follows:

(1) Learners have difficulties in processing VP-condition sentences without knowing the verbs, rather than processing NP-condition sentences without knowing words in object NP.

(2) When learners did not know any word in PP, the tendency of attachment site was not biased.

(3) When learners did not know any word in PP, they attempted to infer the meaning using some lexical clues. At that point, if they somehow associated weird or irrelevant words in the sentence, the whole situation model of the sentence were

collapsed because they could not use other cues such as preposing contexts. In that case, they constructed an implausible sentence with the original structure ignored.

(4) Some learners probably understood all words in sentences, but yet they could not construct sentences properly. However, these cases were rare. The errors were clearly caused by the failure of situation models of sentences. It implied that these participants translated each phrase into Japanese, and their sentence representations were constructed in Japanese at a quite earlier stage. They also did not reanalyze the syntactic structure, which caused mistakes in translation task.

These findings give us some explanation of why the participants in this study prefer NP attachment to VP attachment. The advantages of the NP conditions in preference task might be affected by the verbs, though that was not the only cause. In addition, the learners did not always display bias towards attachment site even when they did not know some words in the sentence. Therefore, that experiment sentence became ambiguous for them.

3.1.3.3 Follow-up analysis 1: Filler sentence analysis

To look for any clues as to the reason for learners' NP attachment preference, it was useful to examine the filler sentences, though the experimental design did not intend to analyze these sentences. Neither the number of sentences nor the counterbalance was controlled. The filler sentences included four types of construction: (a) fully ambiguous sentences using *with*-PP attachment, (b) unambiguous sentences using PP attachment excepting preposition *with*, (c) fully ambiguous sentence using reduced relative clause

(RRC) attachment, and (d) unambiguous sentences using RRC attachment. In unambiguous sentences, the attachment site were constrained by the meaning of sentences just like experimental sentences in Experiment 1. In RRC-attachment sentences, two attachment sites (i.e., low and high attachment sites) were presented to the participants as options to choose. The example filler sentences were shown as follows (see Appendix B for a list):

(19a) The old man pointed at the girl with the stick. [fully ambiguous sentences using *with*-PP attachment]

(19b) The police detective conducted a search for a weapon. [unambiguous sentences using PP attachment excepting preposition *with* (low attachment)]

(19c) The journalist criticized the coach of the runner who was drinking too much. [fully-ambiguous sentence using RRC]

(19d) The nurse trusted the doctors of the teacher who were preparing to go home. [unambiguous sentences using RRC attachment]

As for fully ambiguous sentences (i.e., 19a and 19c), the answer rates of attachment site (i.e., high/low attachment in RRC sentences and NP/VP attachment in PP sentences) were compared in each sentence construction. The more frequently answered attachment site was regarded as a preference attachment site of learners. In unambiguous sentences, the correctly answered rates were compared between high/low (for PP attachment, VP/NP) attachment conditions. Four paired *t* tests were conducted for each ambiguity condition and each type of sentence construction. The descriptive statistics are shown in Table 3.8.

Table 3.8

*Results of Number of Answer Rates of Attachment Site in Ambiguous and Unambiguous Sentences (*N = *26)*

	High (%)		Low (%)				
Structure	*M*	*SD*	*M*	*SD*	*t*(25)	*p*	Cohen's *d*
			Ambiguous sentences				
PP attachment	76.92	23.79	23.08	23.79	5.77	< .001	2.26
RC attachment	35.58	22.00	64.42	22.00	-3.34	.003	1.31
			Unambiguous sentences				
PP attachment	65.73	17.69	83.57	12.60	4.13	< .001	1.16
RC attachment	69.23	37.62	69.23	24.81	0.00	1.00	0.00

Note. PP-attachment sentences did not include *with*-PP.

Results of *t* tests with ambiguous sentences found that (a) learners preferred the VP attachment in ambiguous sentences of *with*-PP attachment; and (b) learners preferred low attachment in ambiguous sentences of RC attachment. These results contradicted each other; learners chose high attachment in PP construction, but they preferred low attachment in RC construction (c.f., The option of VP attachment corresponds to high attachment of RC construction, and NP attachment corresponds to low attachment). Some studies turned out to be the low attachment preference of RC construction when learners' proficiency was relatively low (e.g., Rah & Adone, 2010a). The results showed that the other factor might affect learners' decision of attachment apart from mere length or distance between constituents. The reason must be the differences of sentence construction or lexical information of verb (Frenck-Mestre, 2005).

The first result was also contradicted by results of experimental sentences, but

this was the same tendency as native speakers' response (e.g., Schütze & Gibson, 1999; Thornton et al., 2000). In that point, the results demonstrated that learners show different preference from L1 studies not because they misunderstand the experimental instruction, but because they decide based upon their nature. The reason why the participants' preferences were inconsistent of experimental sentences with ambiguous filler sentences might be the lexical features of ambiguous sentences.

As for unambiguous sentences, *t* tests found out that (c) the learners had NP preference in PP-attachment sentences, but (d) they showed a null preference in RC-attachment sentences.

From the filler sentence analysis, the possibilities that learners' decision about attachment preference were affected by lexical information of main verb came up. The prior studies pointed out that the main verb of PP-attachment sentence affect readers' attachment decision (Kidd & Bavin, 2005; Spivey-Knowlton & Sedivy, 1995). In the following section, the data in Sections 3.1.3.1 and 3.1.3.2 are reanalyzed with verb types.

3.1.3.4 Follow-up analysis 2: Analyses with verb types

According to prior studies, verbs that affect the attachment decision of *with*-PP construction can be categorized into two types: action verbs and, perception and psych verbs. Since this comparison between verb types was also out of the scope of this experimental design, the number of each verb type was not equal in each counterbalanced booklet. However, the analyses were worthy to conduct in order to explain the reasons of learners' attachment preferences.

Overall, experimental sentences consisted of 30 action verbs and 9 perception verbs. The 2 (Attachment: NP vs. VP) × 2 (Length: Short vs. Long) × 2 (Verb: Action vs. Perception) three-way ANOVAs of repeated measures were conducted for correct answer

rate of preference task and translation task scores. Since the number of each verb sentence was not controlled, the scores were recalculated using arcsine transformation (see Table 3.9 for scores of both tasks).

Table 3.9

Descriptive Statistics of Preference and Translation Tasks With Two Types of Verbs

Condition		Action verb			Perception verb		
		M	*SD*	95%CI	*M*	*SD*	95%CI
				Preference task			
Short	VP	73.65	14.38	[63.43, 75.05]	66.08	25.88	[55.62, 76.53]
	NP	69.24	14.39	[63.43, 75.05]	80.79	17.57	[73.70, 87.89]
Long	VP	73.11	18.73	[65.54, 80.67]	74.93	26.17	[64.35, 85.50]
	NP	77.83	17.66	[70.69, 84.96]	80.87	14.57	[75.08, 86.65]
				Translation task			
Short	VP	73.71	15.23	[67.56, 79.87]	51.98	33.52	[38.44, 65.52]
	NP	67.66	14.07	[61.98, 73.35]	62.33	29.59	[50.37, 74.28]
Long	VP	69.29	17.16	[62.36, 76.22]	66.35	28.15	[54.98, 77.72]
	NP	72.79	17.52	[65.71, 79.87]	72.51	24.79	[62.50, 82.53]

Note. The data were arcsine transformed, as the number of each verb type was not the same.

Preference task.

The three-way ANOVA for correct answer rates found the significant two-way interaction between Attachment × Verb type. The subsequent analyses revealed the following points: (a) the insignificant simple main effect of attachment in action verb,

$F(1, 25) = 0.03$, $p = .957$, $\eta^2 = .00$; (b) the significant simple main effect of attachment in perception verb, $F(1, 25) = 7.00$, $p = .014$, $\eta^2 = .22$; (c) the significant simple main effect of verb type in the NP condition, $F(1, 25) = 14.89$, $p = .001$, $\eta^2 = .37$; and (d) the insignificant simple main effect of verb type in VP condition, $F(1, 25) = 0.52$, $p = .480$, $\eta^2 = .02$. Neither the other interactions nor the main effect of Length turned out to be insignificant; $F(1, 25) = 2.76$, $p = .109$, $\eta^2 = .10$ for three-way interaction; $F(1, 25) = 0.00$, $p = .971$, $\eta^2 = .00$ for interaction between Attachment × Length; $F(1, 25) = 0.01$, $p = .925$, $\eta^2 = .00$ for interaction between Length × Verb type; $F(1, 25) = 2.86$, $p = .103$, $\eta^2 = .10$ for the main effect of length. In summary, the interaction for scores of the preference task was attributed to the high scores of NP-attachment condition with perception verbs.

Translation task.

Another three-way ANOVA for translation task score revealed the significant interaction between Length × Verb type and marginally significant interaction Attachment × Verb type: $F(1, 25) = 5.45$, $p = .028$, $\eta^2 = .179$; $F(1, 25) = 3.73$, $p = .065$, $\eta^2 = .13$, respectively. The subsequent analyses for interaction between Length × Verb type showed (a) the insignificant simple main effect of length in action verb, $F(1, 25) = 0.02$, $p = .901$, $\eta^2 = .00$; (b) the significant simple main effect of length in sentence with perception verb, $F(1, 25) = 6.08$, $p = .021$, $\eta^2 = .20$; (c) the insignificant simple main effect of verb type in the long condition, $F(1, 25) = 0.16$, $p = .694$, $\eta^2 = .01$; and (d) the significant simple main effect of verb type in the short condition, $F(1, 25) = 9.64$, $p = .005$, $\eta^2 = .28$. The marginally significant interaction Attachment × Verb type showed (e) the insignificant simple main effect of attachment in action verb, $F(1, 25) = 0.21$, $p = .649$, $\eta^2 = .01$; (f) the marginally significant simple main effect of attachment in sentence with perception verb, $F(1, 25) = 3.60$, $p = .070$, $\eta^2 = .13$; (g) the insignificant simple main

effect of verb type in the NP condition, $F(1, 25) = 0.50$, $p = .488$, $\eta^2 = .02$; and (h) the significant simple main effect of verb type in VP condition, $F(1, 25) = 8.18$, $p = .008$, $\eta^2 = .25$. The significant two-way interaction was not found, $F(1, 25) = 1.39$, $p = .249$, $\eta^2 = .05$; and interaction between Attachment × Length was not significant, $F(1, 25) = 0.36$, $p = .552$, $\eta^2 = .01$.

The results indicated that the reason of interaction between Length × Verb types was attributed to the preferences of the long condition in perception-verb sentences and the preferences of the action verbs in the short condition. Attachment × Verb type occurred due to the high score of NP attachment sentences with perception verb, and action verb with VP attachment sentence. In summary of translation task, the significant results suggest that learners performed better in the NP condition than in the VP condition when the sentences included the perception verb, and they also performed better in sentences with action verb than those with perception verb when they read short-condition sentences.

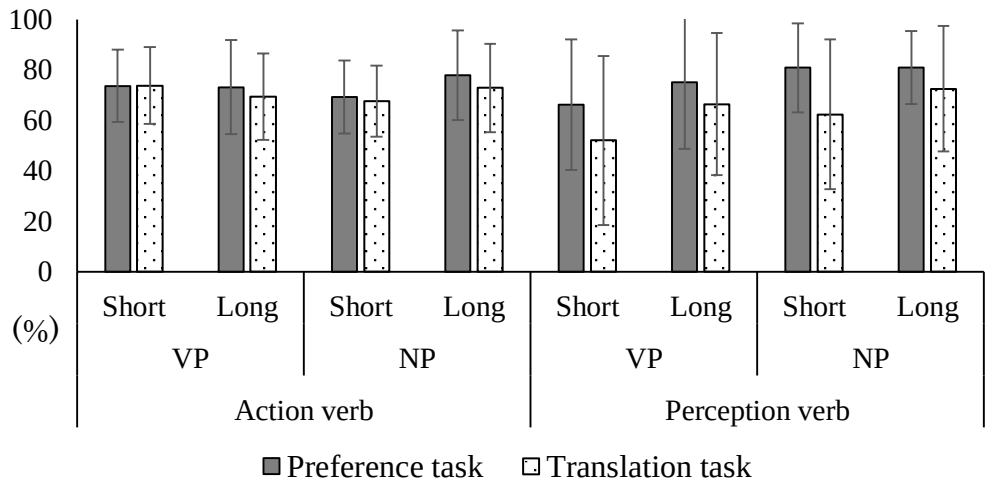

Figure 3.5. Accuracy rates of the preference and translation tasks with verb types.

The results of the translation and preference tasks were consistent with the observation that learners performed better in the long condition with perception-verb sentences (see Figure 3.5). While neither tasks showed the Attachment × Length × Verb type two-way interactions, the phrase intervening VP and PP may have been essential in the NP-attachment interpretation. The sentence with perception verbs essentially biased towards the NP attachment (e.g., Kidd & Bavin, 2005; Spivey-Knowlton & Sedivy, 1995); however, the verb (i.e., the attachment site of VP-attachment condition) is located immediately before the object NP (i.e., the attachment site of NP-attachment condition). Thus, readers were unable to judge the attachment sites from only the syntactic information of distance. However, when three words interrupted VP attachment in the long condition, they were able to rule out the options of VP-attachment interpretation that were supported by both lexical and syntactic information, as seen in Figure 3.6.

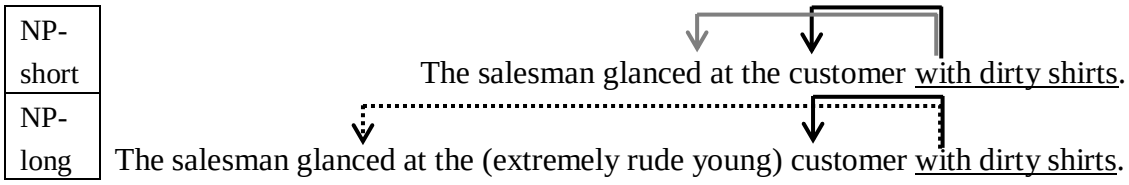

Figure 3.6. Sentence in NP-attachment condition with a perception verb.

3.1.4 Discussion of Experiment 1

3.1.4.1 Attachment preference in off-line tasks (RQ1-1)

Only the preference tasks showed NP attachment bias in the experimental sentences, while the translation task did not indicate any preference. When the verb types of each sentence were analyzed separately, learners identified the strong NP-attachment in perception-verb sentences in both tasks, while they identified the null attachment in action-verb sentences (RQ1-1). Previous studies rarely focused on syntactically

ambiguous sentence processing in off-line methods, because these sentences included no attachment ambiguity or processing difficulties. Although, even considering that the methods of this experiment did not exactly correspond with L1 studies (Schütze & Gibson, 1999; Thornron et al., 2000), it is awkward that EFL learners' NP or non-attachment preference was inconsistent with previous studies. There are two possible reasons for this: (a) non-minimal attachment and recency principles, and (b) effects of the main verbs of sentences.

Regarding the first reason, learners would not adapt minimal attachment in sentence processing. When the minimal attachment principle is adapted, comprehenders should choose high-attached options to construct the simplest structure (Fraizer & Fodor, 1978; Fraizer & Rayner, 1982). The overall trends in this experiment contradict this assumption. Instead, they suggest that learners may prefer low attachment sites, as indicated by the preference and translation tasks, given that learners adapted their structural attachment decisions. The first follow-up analysis of filler unambiguous PP-attachment sentences supported that learners performed better in the low attachment (i.e., NP) condition in the PP-attachment sentences. The results implied that the learners comprehended better in low-attachment than in high-attachment sentences, with or without the preposition *with*. Learners with limited cognitive resources avoided the high-attachment preference that required the retainment of information during processing object NPs.

However, when learners' interpretation of sentences did not signify syntactic parsing, the principle would not be available. For example, in the translation task, Error type D was not consistent with the syntactic structures of sentences even though the learners' comprehension of each word was not problematic. Hence, some learners did not attempt to parse the sentences syntactically, and instead, wanted to construct meaningful

representation in Japanese, their native language. In situations where the processing was based on the meaning, or in their L1, learners did not need to allot their cognitive resources to the structure simplification of sentences, as discussed in the minimal attachment principle, nor did they have to save their cognitive resources as indicated in the recency principle. When learners did not show a structural preference, the recency principles could not be adapted as well as the minimal attachment principle.

The second reason is related to the subcategorization information of the main verbs in sentences. Error analysis of translation tasks showed the disadvantage of the VP condition with unknown verbs, which caused more serious problems than when learners did not know the object NP in the NP conditions. The overall NP-attachment preference was affected by the results of perception verbs. The NP-attachment preferences in sentences with perception verbs can be explained using previous research (Kidd & Bavin, 2005; Spivey-Knowlton & Sedivy, 1995).

For example, in the data from the first follow-up analysis, the results of fully-ambiguous, PP-attachment sentences in filler sentences indicated VP attachment preferences that were inconsistent with the results of experimental sentences. This was because the main verbs of all the ambiguous sentences were action verbs that were biased towards a VP attachment. The experimental sentences included perception verbs as well as action verbs, and consequently, the trends for those were different from ambiguous sentences. Therefore, learners were intuitively affected by the verb categories, which led to confusing results. When the verb categories were omitted, learners generally chose the low attachment sites. The second follow-up analysis indicated that learners performed better in NP and long conditions when sentences included perception verbs, while they did not indicate any preference with the action verb sentences. According to previous studies, the types of verbs affected L1 readers' attachment decisions: Readers showed a

high attachment preference when sentences included the action verbs, while they showed the low attachment preference in the perception- or psycho-verb sentences (Gibson & Pearlmutter, 1998; Kidd & Bavin, 2005; Spivey-Knowlton & Sedivy, 1995). Nevertheless, the result that learners did not prefer the VP attachment even in action-verb sentences is not inconsistent with studies on L1 sentence processing. The reason for the discrepancies can be explained via a detailed analysis of the translation task. Learners' performances suffered considerably when they were unable to identify the meaning of the main verbs. In such cases, the verb type did not affect the attachment sites.

Overall, learners seemed not to use minimal attachment that made syntactic sentence representation simple. At the same time, they did not always attach the PPs to the most recently processed phrases. Therefore, their attachment sites were not consistent through this off-line experiment. Instead of structural parsing, learners sensitively focused on verbs categories or on plausibility of sentences, namely, semantic information of sentences. As a result, the learners strongly indicated an NP preference, especially in sentences with perception verbs. These results were quite different from L1 readers' patterns, which showed strong VP attachment preferences (Thornton et al., 2000), although it is worth noting that the example did not suggest that learners cannot use any syntactic processing. Since the remaining translations highlighted the attachment sites or each part of speech in sentences accurately, learners could perform the shallow parsing, such as segmentation (Clahsen & Felser, 2006a).

3.1.4.2 Distance effect in off-line tasks (RQ1-2)

In off-line tasks, EFL learners did not respond to the distance effect that occurred only in the VP condition (RQ1-2); while the factor of length itself affected performance of translation task. At this point, there were some discrepancies between the preference

and translation tasks in the main effect of length. The performance in the preference task did not suffer from the long conditions. In the translation task, the performance was marginally better in the long condition than in the short condition, which was an unexpected result. To interpret these results, the advantage of long condition should be considered. The additional information in the long condition might be an advantage while constructing sentence representation. Readers in the long conditions could construct more detailed sentence situations than in the short conditions due to three adverbs that were added to the object NPs. Although the three adverbs in the long condition were chosen not to affect attachment decision, the additional information positively or negatively affected learners' performance; for example, Protocol (16b) showed that a participant translated it into neither a syntactically correct nor a plausible sentence (the original sentence: *The hunter killed the horribly big dangerous lion with the gun*, and translation: The hunter was killed by big dangerous lion), where the three additional words could change the meaning of the sentence.

The additional words did not include any syntactic boundary, such as a relative clause attachment or PP attachment (e.g., [NP1] + of + [NP2]), and so they did not require syntactic parsing to comprehend the long part. This indicates that learners could regard the three words as one phrase, or chunk. Learners could process the long-condition sentences without additional cognitive resources or working memory capacities to resolve the long NPs. The preference task, however, did not indicate any difference between the length conditions. Participants did not require the information contained in the three additional words to answer the attachment sites of PP in the preference task. Therefore, they intentionally left out the redundant information in their sentence representation and their comprehension was not affected by the length factor.

Thornton et al. (2000) showed that even the short distance (i.e., three adverbs in

this study) made readers experience processing difficulties in on-line processing. The most important difference is that this experiment adopted the off-line task. The nature of off-line tasks that provide plenty of time might strengthen the tendency wherein learners' cognitive loads of processing long conditions are low. However, learners actually did utilize the phrase length information. Follow-up analysis 2 shows the interaction between verb types and length condition, and between verb types and attachment condition. The additional long phrases worked as separators between attachment sites and irrelevant parts (e.g., the attachment site of object NP and irrelevant verb in the NP condition) as well as processing loads during reading. When learners could access two types of information such as verb types and semantic disambiguation (i.e., lexical and semantic information); or verb types and separating phrase between attachment sites (i.e., lexical and syntactic information), their performance tended to improve.

3.1.4.3 Differences between the preference and translation tasks (RQ1-3)

Comparing the results of the translation task with those of the preference task showed some differences in attachment biases and length manipulation. Before discussing the effects of the task, it should be noted that the scoring of the two tasks: The scoring criteria of the translation task were never related to the additional phrases in the long condition, and it focused on the same point as the preference task, namely, whether or not learners comprehended *with*-phrase attachment, suggesting that the possible reason for the difference of results between tasks was due to the task effect.

As for attachment preferences, the preference task showed the biases towards the NP attachment, while the accuracy rates of the translation task were not biased. The differences between the two tasks might be due to learners' sentence processing strategies and task demands. In the multiple-choice preference task, learners needed to choose the

attachment sites, therefore, their decisions were biased to the interpretable option. While the learners could report the ambiguity of sentences when they noticed it, they still needed to choose the most plausible one. In situations where they had to choose the attachment site, their decisions were biased. However, in the translation task where learners constructed the sentence representation, and did not choose the attachment sites, their accuracy rates were not biased. The latter finding was consistent with the strategy of underspecification, which involved learners not attempting to disambiguate the attachment sites (Swets et al., 2008). The effect of the task could change or affect sentence processing or even the degree of ambiguity resolution.

Regarding task complexities, there were significant differences between tasks. In the translation task, learners needed to produce detailed sentence representation in their L1. However, the preference task did not require such cognitive capacities, especially since the task required learners to choose one attachment site out of only two choices. It is possible that participants in Experiment 1 simply guessed when they did not know the correct answers.

The effects of length was also different in the two tasks. Since learners in this experiment referred to disambiguated regions (i.e., PP after *with*) and were allowed to read back to the cues regions freely in both length conditions during the task, they did not have to retain the previous phrases in their minds. Thus, their comprehension did not suffer due to the distance even in the long condition, unlike in previous on-line sentence processing studies (Thornton et al., 2000).

3.1.5 Conclusion of Experiment 1

The purpose in Experiment 1 was to investigate (a) whether EFL learners show attachment preference in off-line processing as shown in previous studies, (b) whether

EFL learners respond to material distance in off-line processing, and (c) whether question types affect EFL learners' off-line processing focusing on preference and translation tasks.

First, the findings of the off-line task showed that EFL learners preferred NP-attachment sentence when they processed sentences off-line, especially in sentences with perception verbs, while EFL learners performed well in VP-attachment sentences when they were given sentences with the action verb according to the filler sentence analysis. Taking into accounts these contradicting results, learners used the lexical information of the main verb of sentences and were influenced by this information to decide on attachment sites (RQ1-1). Experiment 1 supported neither minimal attachment principle, nor recency principle in EFL sentence processing.

Second, the present study indicated that learners' sentence comprehension did not suffer from distant conditions in the off-line task (RQ1-2). Null distance effect was not consistent with the previous study; since the off-line tasks did not require many cognitive resources as they did not include time limitations. The length itself affected learners' off-line performance, not by placing processing loads, but by providing additional information or indicating a separator.

Third and finally, the differences between the preference task and translation task were based on whether the attachment site of PP was biased or not, and also whether the distant condition affected performance or not (RQ 1-3). These differences were due to task characteristics, and it could affect the learners' sentence processing.

However, there were three methodological limitations in Experiment 1, which are as follows: (a) the multiple-choice preference task could not take into account ambiguous sentences and induced blind guesses, and (b) off-line methods were not used in Thornton et al. (2000). To resolve these problems, Experiment 2 adapted more detailed off-line preference tasks and on-line reading-time methods.

3.2 Experiment 2: On-line Processing and Detailed Off-line Processing of PP-attachment Sentences

3.2.1 Purpose and research questions of Experiment 2

Experiment 1 found that EFL learners' off-line process of PP-attachment sentences was different from native speakers in the following two points: (a) Learners had been affected by constituent length as well as minimal attachment, and (b) learners' data did not show the distance effect.

To investigate the on-line processing in the long condition, the learners performed the on-line self-paced reading task in Experiment 2. This experiment was also designed on the basis of Thornton et al. (2000).

RQ2-1: Do EFL learners show attachment preference in on- and off-line processing?

RQ2-2: Do EFL learners sensitively respond to material distance in on- and off-line processing?

In addition, to resolve the methodological limitation of the preference task in Experiment 1, another preference task was carried out in a different manner. The preference task of Experiment 1 compelled participants to choose either one of the two options, thus not showing the learners' uncertainness. In that case, the learners were able to answer with blind guesses. They were allowed to choose the ambiguity decision (i.e., when participants could not decide where to attach PPs, they were asked to report the item as an ambiguous sentence), but the decisions were still in a coercive manner. The revised preference task that was adopted from the previous study (Schütze & Gibson, 1999) will assess learners' detailed performance by including their uncertainness and eliminating their blind guessing.

As for RQ2-1, Experiment 1 (off-line preference and translation tasks) found out

that learners' performances were better in the NP attachment sentences, contrary to the prediction and other previous studies. There were three possible results in the on-line task: NP attachment preference, VP attachment preference, and no preference. If the learners did not have constraint of minimal attachment, but they were affected by the distance of attachment site as discussed in Experiment 1, they would also show the NP-attachment preference in the on-line task, because the experimental sentences were almost the same as Experiment 1. On the other hand, if the results show VP attachment preference, that might be because learners also have VP-attachment bias similar to L1 readers. In that case, learners initially process PP-attachment sentences with VP-attachment preference, but they change their decision after second processing by some sort of factors. The final prediction suggested that the learners do not show any preference in on-line processing. According to the results of Experiment 1, as learners finally preferred the NP attachment sentences in the off-line task, they intended not to decide on the attachment site of PP in the on-line task, while they resolve the attachment in the off-line task without limitation of time or their own cognitive resources.

The previous studies used preference task with confidence rates in a pilot study, not a main study; thus, the results could not be compared with L1 and EFL readers. Based on the results of Experiment 1, the answer rates in the NP conditions were better than in VP conditions, while the previous studies supported the VP attachment condition.

RQ2-2 was related to the distance of attachment sites and PP. The interaction in Thornton et al. (2000) was based on the assumption of L1 readers' VP-attachment preference. On the other hand, EFL learners in Experiment 1 showed the dominance of NP attachment as discussed above. It means that the same interaction as previous studies was not predicted in Experiment 2, and even if the results show the significant interaction of distance, it was regarded as a different effect.

3.2.2 Methods of Experiment 2

3.2.2.1 Participants

A total of 33 Japanese EFL undergraduates performed the two tasks in Experiment 2. Their English experiences were the same as those of the participants in Experiment 1. Their major varied. None of them had participated in Experiment 1. The data of two participants were deleted since they could not complete the tasks. The final data sets were 31.

3.2.2.2 Material sentences

The experimental sentences in Experiment 2 were constructed to add spillover regions to the end of sentences in Experiment 1. The spillover regions allow us to observe the effects that occur after but at the target region; this effect was called the spillover effect (e.g., Rayner, Kambe, & Duffy, 2000). The spillover regions were two-to-four temporal expression phrases such as *after some hesitation*. To confirm sentence comprehension, all sentences were followed by the single yes/no comprehension question in Japanese. The questions included the contents without PP attachment region. The four conditions of 2 (Attachment: NP vs. VP) × 2 (Length: Short vs. Long) experimental design were also adopted in Experiment 2. The sentences that include ambiguities or unknown words reported from two or more participants in Experiment 1 were excluded or modified into simple words in Experiment 2. The final number of sentences was 36, which comprised nine sentences for each condition.

Some filler sentences were also added, since the experimental sentences were chosen more clearly than in Experiment 1. The filler sentences consisted of four types of sentence structures: (a) the ambiguous PP-attached sentences with or without the term *with*; (b) the ambiguous RC-attached sentences, which were usually in the style of "A of

B + relative"; (c) the high PP-attachment sentences (i.e., VP attachment condition) without the preposition *with*; (d) the high RC-attachment sentences; (e) the low PP-attachment sentences (i.e., NP attachment condition) without the preposition *with*; and (f) the low RC-attachment sentences. The biased RC attachment sentences were disambiguated by the number agreement between RC and antecedent NP. The learners read 36 experimental sentences and 50 filler sentences in total.

3.2.2.3 Procedure

This experiment was conducted individually. The participants performed the on-line sentence comprehension task using computer and response pad and then the off-line preference task with paper and pencil. The participants were randomly assigned to one of four counterbalanced groups; they read the assigned experimental sentences in the on-line task and also the identical sentences in the off-line task. At the beginning of the experiment, the general aim of the study was disclosed to the participants.

On-line sentence comprehension task.

First, they performed the sentence comprehension with a non-cumulative method like moving window. Each word in the sentence was presented in word-by-word style on a computer screen as shown in Figure 3.7. Participants were instructed to read a sentence at their own pace, and they pressed the key to read the next word. At the end of sentences, the comprehension question would appear in Japanese. The participants judged whether the comprehension question was appropriate to the sentence they had just read in English or not. Feedback on their response was not given. The comprehension questions were provided in Japanese so that participants' superficial memories of sentences did not affect performance. The experimental trials were preceded by four practice trials. The

presentation order of sentences was randomized in each participant. They repeated this procedure until the end of the session lasting approximately 30 minutes.

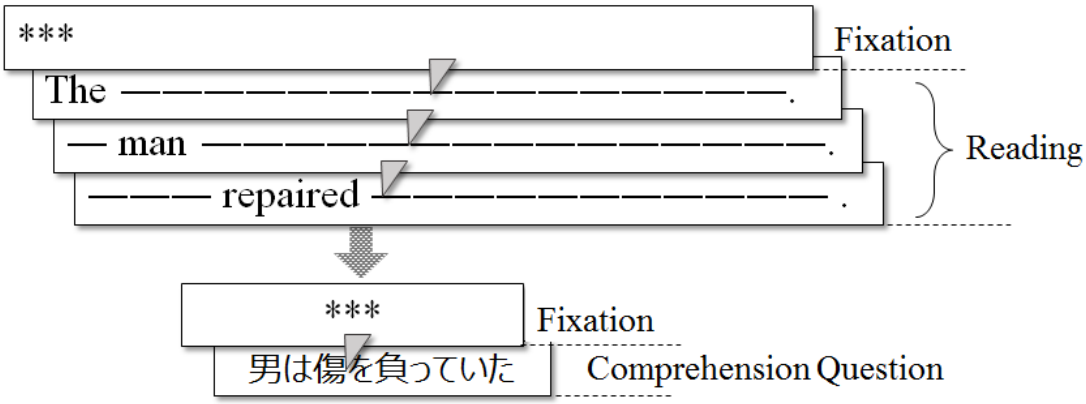

Figure 3.7. Procedure of reading sentences.

Off-line preference task.

After the reading session, the off-line preference task was carried out, but the experimental procedure was slightly different from that of Experiment 1: A list of sentences was presented to learners in order to judge the sentence validities. The preference task in Experiment 2 was conducted using the 7-point Likert scale (1: *The underlined phrase is absolutely high-attached*; 4: *I do not know where the underlined phrase is attached*; 7: *absolutely low-attached*), as shown in Figure 3.8. The sentences in the list were the same as the on-line reading time task, but the presentation order was changed from the previous task.

冊子の英文を読んで、**下線部が(a), (b) のどちらに結びつくと思うかを 7 段階で評価し**、当てはまる数字に〇をつけてください。

例）

Grandmother didn't see any articles on Jupiter. (a) Grandmother (b) article

絶対に(a)だと思う	おそらく(a)	どちらかと言えば(a)	どちらともいえない	どちらかと言えば(b)	おそらく(b)	絶対に(b)だと思う

1 ---------- 2 ---------- 3 ---------- 4 ---------- 5 ---------- 6 ---------- 7

Figure 3.8. Sample instruction of the preference task with confidence rates.

3.2.2.4 Data analysis

The reading-time (RT) data were split into the following three regions for analysis and averaged in each region: Region 1 contained the preposition that begins with the ambiguous phrases (e.g., with); Region 2 contained the disambiguated regions (PP attachment) without preposition; and Region 3 was spillover region as shown in the sentence (20):

(20) The little girl cut the apple / $_{(1)}$ with / $_{(2)}$ the plastic knife / $_{(3)}$ after some hesitation.

In each region, the 2 (Attachment, NP and VP) × 2 (Length, Short and Long) two-way ANOVAs of repeated measures were conducted. The preference task data were converted into 4-point Likert scale to make possible to compare confidence rates between two attachment conditions: The most confident and correct point that participants rated was given 3 points (i.e., 1 in VP and 7 in NP-attachment sentence) and the correct but least confident point was given only 1 point (i.e., 3 in VP and 5 in NP-attachment sentence); no point was given to the answers indicating ambiguity (4: *I do not know where the PP is attached*) in each condition (i.e., 4 = Participants interpreted each condition appropriately, 1 = inappropriately as shown in Table 3.10). Another two-way ANOVA

with the same design as the RT data was conducted for the revised scores.

Table 3.10

Correspondence Relationships Between the Raw 7-point Likert Scale and 4-point Likert Scale in Data Analysis

The 7-point Likert scale in raw score						
(a) high-attachment (VP)						(b) low-attachment (NP)
1	2	3	4	5	6	7
The 4-point Likert scale in analysis						
3	2	1	0	1	2	3

3.2.3 Results of Experiment 2

3.2.3.1 Verification task

Overall, participants answered 83.69% (SD = 1.18) of the comprehension questions correctly. One participant's data was discarded, because she failed to comprehend more than 30% of the experimental sentences. No other participant had an error rate greater than 25%. The remaining 31 participants' data were analyzed in the further analyses as seen in Table 3.11. Since the overall accuracy rates were relatively low, compared with other EFL sentence studies, the two-way ANOVA was carried out in order to confirm the comprehension of each condition.

The ANOVA results found only significant main effect of length, $F(1, 30) = 10.44$, $p = .003$, $\eta^2 = .26$; and other measures turned out to be insignificant: the main effect of attachment, $F(1, 30) = 2.21$, $p = .147$, $\eta^2 = .07$; and interaction between attachment and length, $F(1, 30) = 1.00$, $p = .325$, $\eta^2 = .03$. It should be noted that only the main effect of length without interaction does not support the distance effect.

Table 3.11

Descriptive Statistics of Verification Task in Experiment 2 (%)

Attachment	Length	*M* (*SD*)	95%CI
VP	Short	84.95(11.31)	[2.16, 2.85]
	Long	79.93 (12.31)	[2.43, 3.23]
NP	Short	89.61 (10.71)	[5.74, 6.27]
	Long	80.29 (12.74)	[5.92, 6.30]

Note. *N* = 31.

3.2.3.2 Reading time

Before analyses, RTs were trimmed in the following ways: To compare reading times across participants, items, and regions, the RTs greater than means + 3SD were removed as outliers and replaced with means + 3SD in each participant and item. Because the number of syllables in each region was not equal, the RTs were adjusted by the number of syllables (see Table 3.12). Repeated ANOVAs were conducted for the adjusted RTs separately in each region. The RTs included all participants' data, regardless of their comprehensions:

Region 1 (With).

The reading times in the initial word of PP did not show any significant main effect nor interaction: $F(1, 30) = 0.01$, $p = .924$, $\eta^2 = .00$ for the main effect of attachment; $F(1, 30) = 0.43$, $p = .516$, $\eta^2 = .01$ for the main effect for length; $F(1, 30) = 1.64$, $p = .210$, $\eta^2 = .05$ for interaction.

Region 2 (PP).

In the reading time data of PP excepting the preposition, only the main effect of length was significant: $F(1, 30) = 6.94$, $p = .013$, $\eta^2 = .19$. The finding suggests that the participants read faster in the long than in the short condition. The other effects were still insignificant: $F(1, 30) = 1.72$, $p = .200$, $\eta^2 = .05$ for the main effect of attachment; $F(1, 30) = 0.11$, $p = .748$, $\eta^2 = .00$ for interaction.

Region 3 (post PP).

To examine spillover effect, the reading times of region after prepositional phrase were analyzed. The results did not find any significant effect: $F(1, 30) = 1.72$, $p = .200$, $\eta^2 = .05$ for the main effect of attachment and $F(1, 30) = 0.04$, $p = .850$, $\eta^2 = .00$ for the interaction, while the main effect of length showed moderately the same tendency as Region 2, $F(1, 30) = 2.99$, $p = .094$, $\eta^2 = .09$.

Table 3.12

*Means of Reading Time (milliseconds) With Standard Deviation in Experiment 2 (*N = *31)*

Length	Attachment	Region 1 (*with*)	Region 2 (PP)	Region 3 (Post)
Short	VP	646.61 (196.45)	430.22 (233.52)	442.88 (112.24)
	NP	688.76 (290.84)	429.68 (184.33)	458.63 (198.42)
Long	VP	710.11 (231.49)	369.30 (118.95)	423.54 (110.27)
	NP	673.58 (209.17)	357.77 (122.01)	434.54 (110.27)

Note. Reading Times (RTs) were adjusted by the number of syllables.

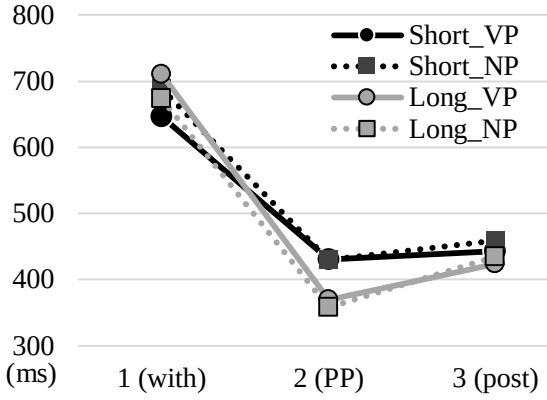

Figure 3.9a. The reading times in Experiment 2.

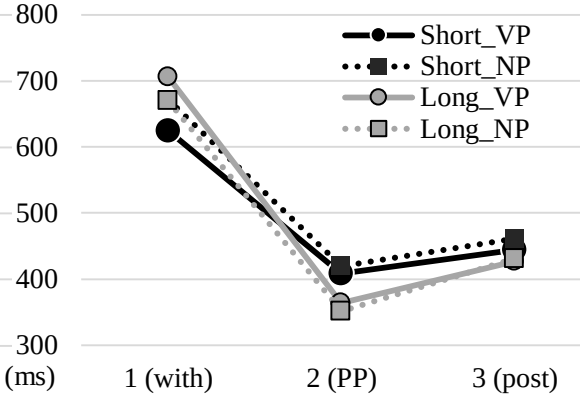

Figure 3.9b. The solution times in Experiment 2.

As shown in Figure 3.9a, the results of RT showed only one significant difference between length conditions in the disambiguated region (Region 2), indicating that there was no preference of the attachment sites in on-line processing, and the long conditions were processed faster than the short conditions. Although the result implied that readers felt fewer cognitive loads in the long condition than in the short condition, this is not likely to be the case. Nakanishi (2012) suggested that the longer distance did not cause higher memory cost for EFL learners, but it did not imply that readers read sentences with the long phrases faster than sentences with the short phrases. To confirm the length effect, the RTs were recalculated with incorrect answers removed in the next section.

3.2.3.3 Solution time

The reading times in the previous section included the sentences that the participants read without full comprehension. Where data was removed from sentences due to the comprehension questions being answered incorrectly, the solution time (which was limited to reading time with correct responses to comprehension questions, hereafter ST) was analyzed (see Table 3.13 and Figure 3.9b):

Region 1 (With).

The result of ST was different from RT data in Region 1. The significant main effect of attachment was found, $F(1, 30) = 4.29$, $p = .047$, $\eta^2 = .13$, though the main effect of length was not significant, $F(1, 30) = 0.48$, $p = .494$, $\eta^2 = .02$. The interaction was not significant, either, $F(1, 30) = 0.73$, $p = .398$, $\eta^2 = .02$.

Region 2 (PP).

The STs in Region 2 were similar to the RTs. Though the main effect of length did not reach the significant *p*-value, it still showed marginal significance: $F(1, 30) = 4.00$, $p = .054$, $\eta^2 = .12$. The significant main effect of attachment and interaction was not found as in reading times, $F(1, 30) = 0.29$, $p = .593$, $\eta^2 = .01$ for main effect of length; $F(1, 30) = 0.53$, $p = .470$, $\eta^2 = .02$ for interaction.

Region 3 (post PP).

The STs in Region 3 did not show any significant main effect or interaction: $F(1, 30) = 2.44$, $p = .129$, $\eta^2 = .08$ for the main effect of attachment; $F(1, 30) = 0.00$, $p = .988$, $\eta^2 = .00$ for the main effect of length; $F(1, 30) = 1.04$, $p = .316$, $\eta^2 = .03$ for the interaction.

Based on the results of RTs and STs, only significant results were found in the length effect in Regions 2 and 3: it indicated that learners' reading pattern did not differ in the correctness of comprehension questions.

Table 3.13

*Means of Solution Times (milliseconds) With Standard Deviation in Experiment 2 (*N = *31)*

Length	Attachment	Region 1 (*with*)	Region 2 (PP)	Region 3 (Post)
Short	VP	624.15 (226.88)	408.17 (221.73)	444.33 (113.71)
	NP	671.53 (273.23)	419.42 (161.08)	461.36 (102.48)
Long	VP	705.31 (243.04)	363.89 (118.90)	427.00 (116.58)
	NP	669.61 (193.76)	351.30 (122.78)	430.74 (105.30)

Note. The solution time was adjusted by the number of syllables.

3.2.3.4 Preference task with confidence rate (off-line data)

Table 3.14 shows both raw and revised 4-point scale scores of the preference task. The ANOVA for the 4-point scale scores showed the significant main effect of attachment, $F(1, 30) = 19.37$, $p < .001$, $\eta^2 = .39$; the insignificant main effect of length, $F(1, 30) = 1.91$, $p = .178$, $\eta^2 = .06$; and significant interaction between attachment and length; $F(1, 30) = 5.74$, $p = .023$, $\eta^2 = .16$; as shown in Table 3.15.

Table 3.14

The Descriptive Statistics of Raw and 4-point Scores of Preference Task in Experiment 2

		Raw score		Revised 4-point scale score	
Length	Attachment	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI
Short	VP	6.02 (0.77)	[2.16, 2.85]	2.64 (0.88)	[2.32, 2.96]
	NP	2.52 (1.02)	[5.74, 6.27]	3.05 (0.75)	[2.77, 3.32]
Long	VP	6.01 (0.61)	[2.43, 3.23]	2.20 (1.14)	[1.79, 2.62]
	NP	2.87 (1.12)	[5.92, 6.30]	3.12 (0.54)	[2.92, 3.32]

Note. The possible maximum score was 7 in raw, and 4 in 4-point scale.

The subsequent analysis for interaction indicated that interaction was due to the least confidence in VP-long condition, as shown in Figure 3.10: the significant simple main effect of Attachment in short condition, $F(1, 30) = 4.96$, $p = .034$, $\eta^2 = .14$; the significant simple main effect of Attachment in long condition, $F(1, 30) = 24.74$, $p < .001$, $\eta^2 = .45$; the insignificant simple main effect of Length in NP sentence, $F(1, 30) = 0.23$, $p = .636$, $\eta^2 = .01$; and the significant simple main effect of Length in VP sentence, $F(1, 30) = 5.57$, $p = .025$, $\eta^2 = .16$. The results also showed that EFL learners read the NP-attachment sentences with confidence, as seen in the preference task of confidence rate in Experiment 2.

To investigate whether learners' preference was generally affected by distance, the follow-up analysis was conducted for filler sentences in the following section.

Table 3.15

Summary Table for Two-way ANOVA of the Effects of Attachment and Length on the 4-point Scale Scores

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
Attachment (A)	1	13.59	13.59	19.37	< .001	.21
Length (L)	1	1.01	1.01	1.91	.178	.02
A x L	1	1.99	1.99	5.74	.023	.03
Error (A)	30	21.05	0.70			
Error (L)	30	15.96	0.53			
Error (A x L)	30	10.38	0.35			

3.2.3.5 Follow-up analysis 1: Filler sentence analysis

The filler sentences used in Experiment 2 were examined like Experiment 1. In this analysis, the solution time and preference task were examined separately. The constructions of filler sentences were the same as Experiment 1, but the number of sentences were different: (a) 13 ambiguous sentences using with-PP attachment, (b) 11 NP-attachment sentences using PP attachment (excepting *with*), (c) 11 VP-attachment sentences using PP attachment (excepting *with*), (d) eight ambiguous sentences using reduced relative clause attachment (RRC), and (e) five unambiguous sentences using RRC attachment. Before the analyses of preference tasks with confidence rates, correct answer rates of verification questions are shown in Table 3.16. The rates were angular transformed, because the number of sentences was not controlled.

Table 3.16

*The Correct Answer Rates of Verification Questions for Filler Sentences (*N *= 33)*

	Unambiguous sentence				Ambiguous sentence	
	High		Low			
Structure	*M*	*SD*	*M*	*SD*	*M*	*SD*
PP attachment	62.85	9.66	60.18	9.39	59.60	5.83
RC attachment	70.91	31.88	68.29	22.43	64.02	11.12

Note. PP = Prepositional phrase, and RC = Relative clause. Unambiguous PP attachment did not include the preposition of *with*, while ambiguous PP attachment included it.

***Analyses of preference tasks with confidence rates*.**

In order to compare the answers of NP attachment and VP attachment of the filler sentences, the confidence rates of unambiguous sentences were manipulated into 4-

point Likert scale as in the main analyses in Section 3.2.3.4. For the ambiguous sentences, the number of times that each participant chose high-attachment options (i.e., 1 to 3) or low-attachment options (i.e., 5 to 7) were counted, while responses that chose 4 (*I do not know where the underlined phrase is attached*) were regarded as an unresolved answer, and were ignored. The total of four paired *t* tests was used to compare each sentence construction.

Results of the preference task with confidence rates in unambiguous and ambiguous sentences.

As for the unambiguous sentences, the paired *t* tests indicated that confidence rates of both sentence structures had significant differences between high and low attachments (see Table 3.18). It also said that learners correctly and confidently answered in low-attachment sentences regardless of sentence constructions.

Table 3.17 also shows the answer times on ambiguous sentences. Note that how to count the attachment site was different from the answer rates of unambiguous sentences, because ambiguous sentences did not include the syntactically or semantically correct attachment site. The attachment biases of fully-ambiguous sentences showed the marginally significant differences in PP-attachment sentences, and significant differences in RC-attachment sentences: Learners chose the low attachment sites in RC ambiguous sentences, while they preferred to attach PP including the preposition *with* to VP attachment sites.

Table 3.17

*Results of Answer Rates of Attachment Site in Unambiguous Sentences and Ambiguous Sentences (*N = *33)*

	High		Low				
Structure	*M*	*SD*	*M*	*SD*	*t*(32)	*p*	Cohen's *d*
			Unambiguous sentences (4-point scale)				
PP attachment	1.11	1.05	2.46	0.85	6.21	< .001	1.41
RC attachment	0.62	2.14	1.59	1.38	2.34	.026	0.54
			Ambiguous sentences (frequency data)				
PP attachment	7.19	3.03	5.59	2.31	1.77	.086	0.59
RC attachment	2.88	1.90	4.47	1.67	-2.70	.011	0.89

Note. High = High attachment site (i.e., VP attachment in PP attachment sentences), Low = Low attachment site (i.e., NP attachment in PP attachment sentences), PP = Prepositional phrase, and RC = Relative clause. Unambiguous sentences in PP attachment did not include the preposition *with*.

***Analyses of RTs and STs of the filler ambiguous sentences*.**

The reading times of unambiguous PP-attachment sentences were measured to examine whether EFL learners process the PP-attachment sentences without the preposition *with* in the same way as the experimental sentences that included *with*-preposition sentences. Similar to the RT and ST analyses of the experimental sentences, the two analytic regions were set: (a) the preposition word region (Region 1), and (b) the disambiguated region (Region 2). Some filler sentences did not include post PP region (see Region 3 in Sections 3.2.2); thus, the region of post PP (Region 3 in Section 3.2.3.2) was eliminated from analyses. The number of syllables and outliers were adjusted as in

the experimental sentences. In this analysis, each attachment-site condition included 11 sentences, but the sentences were not counterbalanced, because the participants read both conditions. RTs included reading times of all sentences, while STs excluded reading time data where learners answered incorrectly the verification questions. The results were summarized in Table 3.18.

Table 3.18

*Means of RTs and STs (milliseconds) With Standard Deviation of Filler Unambiguous PP-attachment Sentences in Experiment 2 (*N *= 31)*

		Region 1		Region 2	
Measures	Structure	*M*	*SD*	*M*	*SD*
RT	High	543.53	138.09	875.01	273.09
	Low	608.44	206.17	826.74	308.11
ST	High	518.89	112.96	871.40	325.98
	Low	594.73	160.66	842.97	323.29

Note. RT = Reading time, ST = Solution time. RTs were adjusted by the number of syllables. STs were calculated with sentences in which the verification questions were correctly answered from reading data.

Results of RTs and STs of filler unambiguous PP-attachment sentences.

To compare reading times of two attachment sites, the two-way (Attachment: High and Low × Regions: 1 and 2) ANOVAs were conducted for RTs and STs. Both were the within-participant factors. The analyses especially focused on the differences of high and low attachment sentences. First, the ANOVA for RTs did not show the significant interaction between attachment and regions $F(1, 32) = 2.78$, $p = .105$, $\eta^2 = .08$; nor the

main effect of attachment, $F(1, 32) = 0.10$, $p = .759$, $\eta^2 = .00$. It showed only the significant main effect of regions, $F(1, 32) = 48.49$, $p < .001$, $\eta^2 = .60$. The difference between regions did not make sense in sentence processing; thus, RTs of filler sentences did not show the meaningful differences. Data of STs that excluded incorrectly answered sentences in the verification task showed the same tendency. Neither the significant interaction nor the main effect of attachment were not found out; $F(1, 32) = 2.39$, $p = .132$, $\eta^2 = .07$ for interaction; and $F(1, 32) = 0.66$, $p = .422$, $\eta^2 = .02$ for the main effect of attachment site. The main effect of regions was significant, $F(1, 32) = 44.52$, $p < .001$, $\eta^2 = .58$. These findings indicated that learners did not show any preference in high or low PP-attachment sentences in on-line processing. It should be noted that the comparisons of filler sentences without experimental sentence sets were not strictly appropriate, but the same tendency as the experimental sentences was gained.

In sum, the follow-up analysis of filler sentences showed that RC-attachment sentences were chosen as low attached sentences in off-line processing, regardless of the ambiguities of sentences. In contrast, the PP-attachment sentences showed complex results: In unambiguous PP-attachment sentences except for the preposition *with* (i.e., sentences that disambiguated by semantic information), learners performed better in forced disambiguated low-attachment sentences; in ambiguous sentences using the preposition *with*, learners chose VP-attachment sentences; and lastly, the on-line reading-time data of ambiguous sentences with or without incorrect comprehension showed the null preference of attachment sites.

3.2.3.6 Follow-up analysis 2: Analyses with verb types

In this section, only the off-line preference task with confidence rate was reanalyzed. The types of verbs for solution times could not be analyzed, because some of

the reading time data that the learners had answered correctly in each condition were missing because the number of action and perception verbs was not equivalent. The descriptive statistics for scores of the preference task are shown in Table 3.19 with each verb type separate. The 2 (Attachment: VP and NP) × 2 (Length: Short and Long) × 2 (Verb type: Action and Perception) three-way repeated ANOVA was conducted for the 4-point scale confidence rates. The results showed the significant one-way interaction between Attachment × Length, $F(1, 30) = 4.90$, $p = .035$, $\eta^2 = .01$; and between Attachment ×Verb, $F(1, 30) = 8.63$, $p = .006$, $\eta^2 = .02$. The main effects of attachment and verb were significant, $F(1, 30) = 13.39$, $p = .001$, $\eta^2 = .11$; and $F(1, 30) = 9.22$, $p = .005$, $\eta^2 = .03$. The other main effects and one-way and two-way interactions did not show significance. The subsequent test for only the interaction between verb and attachment was conducted, because the interaction between attachment and length is already reported in Section 3.2.3.4.

The subsequent analysis for the interaction between verb and attachment showed the significant simple main effects of verb types only in VP condition, $F(1, 30) = 12.35$, $p < .001$; but insignificant in NP condition, $F(1, 30) = 0.05$, $p = .817$; and also the significant simple main effects of attachment sites in action verbs, $F(1, 30) = 6.52$, $p = .016$; and in perception verbs, $F(1, 30) = 13.73$, $p < .001$. The results showed that learners answered more confidently regarding sentences with action verbs than perception verbs in VP-attached sentences, while there were no preferential differences of verb types in NP sentences. In addition, learners had more confidences in NP-condition sentences than VP-condition sentences, regardless of verb types. The interaction Verb type × Attachment was due to the low performance in VP-attached sentences with perception verbs, and overall biases towards the NP condition, as shown in Figure 3.11.

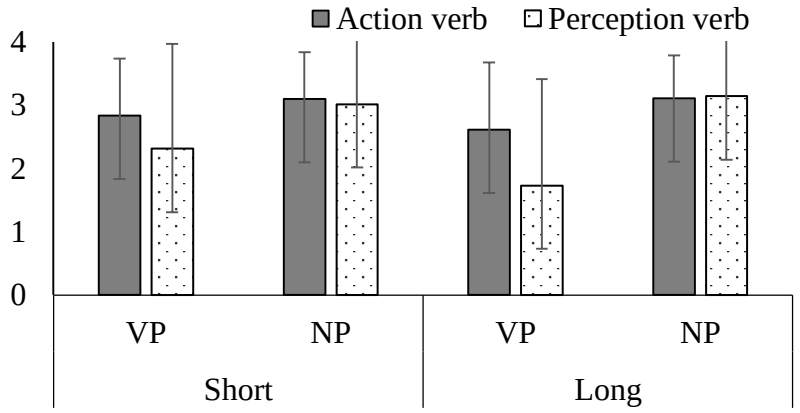

Figure 3.11. Results of 4-point scores of the preference task with verb types.

Table 3.19

*Descriptive Statistics of 4-point Scores of the Preference Task With Verb Types in Experiment 2 (*N *= 31)*

		Action verbs		Perception verbs	
Length	Attachment	*M* (*SD*)	95%CI	*M* (*SD*)	95%CI
Short	VP	2.84 (0.90)	[2.50, 3.16]	2.31 (1.65)	[1.70, 2.92]
	NP	3.10 (0.74)	[2.82, 3.37]	3.01 (1.17)	[2.58, 3.44]
Long	VP	2.61 (1.06)	[2.22, 3.00]	1.73 (1.68)	[1.11, 2.35]
	NP	3.11 (0.67)	[2.86, 3.36]	3.14 (1.03)	[2.76, 3.52]

Note. The possible maximum score was 4.

3.2.4 Discussion of Experiment 2

3.2.4.1 Attachment preference in on- and off-line tasks (RQ 2-1)

The preferences of PP attachment sites were not found in the on-line self-paced reading task. However, lack of bias in PP attachment site does not indicate that learners could not use semantic information: Because the overall accuracy rates of the comprehension questions after reading sentences were relatively high, learners should

have been able to analyze the semantic information. The on-line task's results that did not show any attachment preference were partially consistent with the shallow structure hypothesis, which predicted no preference in ambiguous PP attachment sites (Clahsen & Felser, 2006a, 2006b). Other reasons likely affected learners' performance of self-paced reading: One of the possible reasons was the lack of task demand, as seen in Experiment 1. The on-line task required immediate decision making, which placed an additional burden of processing on readers. The participants in this experiment, however, answered the comprehension questions, but they did not need to decide on attachment sites, which led them to use underspecification strategies. Such strategies are frequently used when readers do not necessarily have to resolve ambiguous sentences in situations where comprehension questions are not given (Swets et al., 2008). Learners who have suffered from an overload of the on-line task would not attempt to process additional tasks (i.e., attachment decision) in order to lessen their processing burden using the underspecification strategy. The result of unambiguous experimental sentences was consistent with that of the first follow-up analysis for filler ambiguous PP-attachment sentences using the preposition *with*: Learners had a preferred attachment site in off-line processing, but they did not resolve it in on-line processing.

The off-line task (i.e., the detailed preference task with confidence rates) indicated the overall NP attachment preference as discussed in Experiment 1. The result was unsurprising, because the experimental sentences used in Experiments 1 and 2 were almost the same. The differences off-line tasks between Experiments 1 and 2 were with or without confidence rates of preference decisions. In Experiment 1, participants needed to determine which site to attach PP to, regardless of their confidence. They were allowed to report some sentences as ambiguous by checking a box, but they also had to choose one of the options. The restrictions would led the participants to make a blind guess or,

to choose the nearest attachment site to reduce their processing load if they had a structural preference. The preference task with learners' confidence rates in Experiment 2 were affected by the degree that learners had confidence in their decision as well as the accuracy rates. When the verb types in sentences were analyzed separately, the results showed that learners had more confidence when they chose the NP-attachment option. In contrast to the preference task in Experiment 1, learners had confidence to answer NP-attachment sentences even when these sentences included action verbs. The relationships between verb types and attachment sites were previously discussed in Experiment 1. Therefore, learners were affected by lexical information, such as verb types to resolve the ambiguous attachment site. When learners could not use lexical information due to some sort of problem such as they did not know the meaning of a verb, their confidence ratings tended to bias the recent options.

3.2.4.2 Distance effect in on- and off-line tasks (RQ 2-2)

As for the effects of length manipulation, reading times in Region 2 (disambiguous region) were faster in the long condition than in the short condition, which was surprising. Some prior studies expected and indicated that readers integrated additional phrases into their syntactic construction in the long condition. This effect is called the *wrap-up effect*, and it usually occurs after sentences or phrases (Rayner et al., 2000). Both length conditions require one to decide which sites to attach PP to by reference to the verb and object NP, but readers spent more time to parse the PP region in the short conditions compared to the long conditions in contrast to the wrap-up effect. According to the self-reading paradigm, the shorter processing time indicated no processing difficulty or non-additional activity such as making inference, compared with the longer processing time. However, the results in the reading-time data did not imply

that the long-condition sentences were more easily read than the short-condition sentences. The results of off-line comprehension questions that were conducted immediately after each sentence reading supported the length effect: the readers made more mistakes in the long conditions than in the short conditions. Since the comprehension questions were constructed without an additional phrase in the long conditions, the lengthening of the object NP affected the parsing of entire sentences. The possible reason for this irregular anti-length effect in reading-time data was due to the chunking of long phrase parts into one, as discussed in Section 3.1.4.2. Learners determined that the long part was redundant information, and they attempted to leave out the redundant phrase as one chunk. In particular, learners might attempt to reduce time gap between words in the long part by pressing a key quickly, as the word-by-word sentence reading violated readers' perceptual syntactic boundaries. As a result, they could not pay enough attention to the subsequent disambiguate regions that occurred after the long phrase. Therefore, learners read the PP region faster in the long condition but they comprehension was worse than in the short condition.

The length effect was not consistent with the results of the EFL sentence processing study (Nakanishi, 2012), which did not find the length effect in the PP-attachment sentences. Nakanishi's study adapted phrase-by-phrase self-paced reading, while this study and Thornton's L1 self-paced reading study used the word-by-word style.

The length effects are not equal to the distance effects, however. The distance effects as mentioned in Thornton et al. (2000) occurred when the VP-attachment dominance was moderated because of long phrase that intervenes between attachment sites and PP. Thornton's study premised the VP-attachment preference of L1 readers in the short condition before it was found that distance in the long condition cancelled out the VP-attachment preference, but this study did not meet the preconditions as discussed

in Experiment 1. The result that longer STs occurred in the NP conditions as well as the VP conditions did not indicate the distance effect, and it remained the only length effect.

The off-line preference task with confidence rates showed quite different results from the self-paced reading task with regard to distance. The confidence rates clearly decreased in the long condition, and this occurred in only the VP condition, while there were no differences in the NP conditions whose distance between attachment sites and PP were the same in the two length conditions. Although the interaction pattern of the confidence rates was not exactly the same as in the prior L1 study (Thornton et al., 2000), the distance between verb and PP might have affected learners' performance in the VP sentences, considering the overall NP-attachment preference in Experiments 1 and 2.

3.2.5 Conclusion of Experiment 2

The purpose of Experiment 2 was to examine (a) whether learners show attachment preferences in on-line processing, and (b) whether learners sensitively respond to material distance in on-line processing.

First, learners' attachment preference for PP was not found in on-line processing, while the preference task with the confidence rates showed that learners answered the NP-attachment sentences more confidently (RQ2-1). Although their attachment decisions were not consistent, it was supposed that learners used lexical information of sentence verbs, and then they focused on the recency principle when subcategorization information was not available.

Second, learners did not show the distance effects in on-line processing, but their distance effects similar to L1 studies were shown in off-line processing (RQ 2-2). The finding of on-line processing did not mean that EFL learners were not affected by the additional phrase in the long condition, because differences in reading times between the

short and long conditions were found out. The off-line preference task with confidence rates showed that their performance fell in the VP-long condition. This pattern was not exactly the same result as in previous studies (Thornton et al., 2000), but it might show a distance effect when considering the NP preferences in Experiment 1.

A complement survey.

Although the primary aims of the present study were not to determine the preferred attachment sites of PP, the NP-attachment site preference was not consistent with the results of previous studies, and the reason for this is still unclear. L2/EFL studies showed results that were opposite to ones in this study, though investigations into Japanese EFL learners are limited. Studies with VP-attachment evidence claimed structural preference in both fully ambiguous and syntactically ambiguous sentences. The syntactically ambiguous sentences are resolved using semantic or pragmatic information. The basic question is whether Japanese EFL learners are familiar with the usage of the preposition *with*. If learners are not familiar with VP-attachment use (i.e., instrument), it is suggested that their comprehension of VP-attachment sentences is hindered when compared to learners. In order to investigate learners' usage of *with*, a simple production survey was conducted, based on EFL sentence production studies (e.g., Rah & Adone, 2010a; Spivey-Knowlton & Sedivy, 1995, Experiment 4; Ying, 1996, Experiment 3).

Eighteen Japanese EFL learners participated in the complement study. The participants were second-year undergraduates, and had studied English for more than eight years. Data of a student who could not complete the task was excluded; hence, the final count of the data set was 17. English proficiency of the participants was of the intermediate level (TOEIC scores were around 600). This survey was conducted in an English writing class, and the participants were unaware of the purpose of this survey. In

this simple survey, the participants were asked to construct a single sentence using *with*, *in*, and *at*. The procedure was repeated three times. Thus, there were three sentences that were independent to one another (i.e., no sentence was connected to another). The prepositions *in* and *at* were regarded as filler structures, and they were not analyzed. Though EFL sentence production study provided the subject and verb of the sentences to examine PP attachment sites (Rah & Adone, 2010a), this study did not give the particular verb or nouns in order to focus on learners' unforced use of *with*. The participants completed the sentence complement task without consulting a dictionary or browsing the Internet. The whole procedure took approximately 10 minutes.

A total of 51 sentence protocols were collected. The writing protocols were categorized into PP-attachment structures and other structures. Protocols that did not contain verbs or incomplete sentences were excluded (Ying, 1996, Experiment 3); the number of the remaining protocols was 49. Out of these, 26 (53.06%) included PP-attachment structures (i.e., [VP] + [NP] + [PP]). Many of them (23 out of 26 protocols, 88.46%) were regarded as a VP-attachment sentences, and the usages were quite biased: Most protocols of VP-attachment (18 protocols, 69.23 % of PP-attachment sentences) used the term *with* to indicate accompanying person as in *I played the piano with my sister* or *I often go to the library with my friends*. Others used the instrumental use of PP attachment (5 protocols, 19.23%): *She cut a meet* (as in the original) *with a knife*. The others denoted NP-attachment structures (only 3 protocols, 11.53%): *I saw a girl with a blond hair*. Other structures included phrasal verbs such as *go with*, *talk with*, and *play with*; and *with* in attendant circumstances such as *The girl are standing with her arms across her chest*.

The survey of sentence completion tasks demonstrated that learners preferentially produced the VP-attachment sentences using the preposition *with*, but

learners' impression of the preposition *with* was highly biased toward the role of the accompanying person. Furthermore, the experimental sentences of the VP-attachment condition rarely include such usage, and the most frequent usage of PP was instrument (e.g., shot + [NP] + with). Although receptive and productive processing was not identical, processing bias of learners was shown. This finding was also consistent with the previous studies (Rah & Adone, 2010a). These biases of grammar usage might occur because of instructions in the English classroom.

3.4 Summary of Study 1

Attachment preference.

Table 3.20 summarizes the overall findings of Experiments 1 and 2. The overall attachment preferences in the two experiments turned out to be low-positioned NPs when they have to choose the attachment sites. Focusing on the verb types of sentences, learners preferred the low attachment site in sentences with perception verbs, while they preferred the high attachment site with action verbs. In addition, learners also had semantic biases toward the usage of *with*, as shown in the complement survey. Therefore, learners' sentence processing was strongly affected by semantic and lexical cues of sentences.

While such cues changed learners' disambiguated decisions, the structural preferences were also found in the study. Overall tendencies showed that learners preferred to attach, or confidently attach, the current information to the most recently processed NP, which is consistent with recency or late closure (Schütze & Gibson, 1999), and did not support the minimal attachment principles (Fraizer & Fodor, 1978; Fraizer & Rayner, 1982). Though some of previous studies demonstrated the high-position preference (i.e., VP preference) for L1 readers and L2/EFL learners using a prepositional phrase sentence with similar construction to that of this study (Thornton et al., 2000). One

of the EFL sentence processing study (Nakanishi, 2012) used PP attachment sentences and concluded that Japanese EFL learners used the minimal attachment strategy. However, his materials were different from other studies in points that the experimental sentences included ambiguous and forced-disambiguated processes (e.g., *The woman visited the man early this morning* <u>*with a car instead of walking*</u> in minimal attachment and long conditions). At the point of *with a car*, the sentence was still ambiguous, and it was disambiguated after reading *walking*. As the materials included some different processing, this study was not inconsistent with his study. In addition, some EFL studies were not consistent with the high-attachment preference of PPs. For example, Yokokawa (1994) could not find the dominance of VP-attachment preference in self-paced reading.

Length and distance effects.

As for the length manipulation, Table 3.20 shows that the learners consistently experience comprehension difficulties in the long condition, especially in off-line tasks and with the exception of preference tasks. Therefore, length manipulation mostly affects learners' sentence comprehension when the added phrase is clearly a pre-modifier of nouns in a sentence. However, the results of the preference task showed null length effects; thus, the off-line preference task would mitigate length effect, because the task distracts readers' attentions to the attachment sites or modification relation. Experiment 1 also entailed a translation task, which showed a similar tendency to that of the on-line task. The reason for these different results in two kinds of off-line tasks might be the processing loads that learners experience during the tasks. When EFL learners processed sentences in on-line, they experienced difficulty in reading sentences featuring long phrases; similarly, they had difficulties in answering the off-line translation task, which requires greater cognitive capacities of learners.

Table 3.20

Summary of Findings in Study 1

Experiment	Task	Attachment	Length	Interaction
Experiment 1	Preference task	VP < NP	*n.s.*	*n.s.*
[Off-line]	action verbs	VP < NP		
	perception v	*n.s.*		
	Translation task	*n.s.*	Short ≧ Long	*n.s.*
	action verb			
	perception v			
Experiment 2 [Off-line]	Preference task with confidence rate	VP < NP	*n.s.*	The least in VP-long condition
	CQs after self-paced reading	*n.s.*	Short > Long	*n.s.*
[On-line]		Reading time		
	With	*n.s.*	*n.s.*	*n.s.*
	PP	*n.s.*	Short > Long	*n.s.*
	Post PP	*n.s.*	Short ≧ Long	*n.s.*
		Solution time		
	With	*n.s.*	*n.s.*	*n.s.*
	PP	*n.s.*	Short ≧ Long	*n.s.*
	Post PP	*n.s.*	*n.s.*	*n.s.*

Note. The underlined conditions were interpreted as a preference of EFL learners. The off-line preference and translation tasks regard higher correct rates as preference, while the on-line reading time data regard faster time as preferred options.

On the other hand, the distance effect was not found in the on-line study, although L1 study (Thornton et al., 2000) confirmed the effect in the on-line method. In

off-line processing, learners' confidence rates of attachment preference decreased in the VP-long condition in that the distance between attachment sites and PP was the longest in four conditions. Since the distance effect was regarded as a sensitive construction of syntactic structure, this study's finding implied that learners were not sensitive to the syntactic structure during on-line processing, and their sensitivities improved when they could refer to a whole sentence in an off-line task.

Study 1 discussed the learners' PP-attached sentence processing, and it found that learners highly depended on the semantic, pragmatic and lexical cues to resolve syntactically ambiguous attachment sites. However, the results are not always the same as they were for the participants in this experiment, and individual factors of each learner may affect the decisions of PP attachment sentences. Additionally, the strategy of underspecification occurs even when the experimental sentences are disambiguated, and not just in fully ambiguous sentences. Study 2 will observe learners' own reports on sentence comprehension to investigate individual strategy use with local and syntactically ambiguous sentences.

Chapter 4

Study 2: Learners' Use of Strategy in Sentence Processing

4.1 Pilot Study for Processing Unit Tests

4.1.1 Purpose of pilot study

Prior to Experiment 3, a pilot study was carried out to confirm relationships between learners' processing unit and reading proficient level. Based on Hijikata (2012), this study used the processing unit as a learner's individual factor. The individual processing unit was assessed with a timed reading: Learners read a sentences with a time limitation, and they recall sentences in English without referring ordinal sentences. Learners with large processing units would recall large span of sentences. This study used the condition with strict time limitation, although Hijikata (2012) had already examined this relation.

4.1.1.1 Pre-pilot Study

Before the pilot study, a pre-pilot study was carried out to confirm the procedure of the test of processing unit span (Hijikata, 2012) within five seconds for each sentence. For the processing unit span test in this study, 20 sentences were set according to Hijikata's method (2012). She manipulated these sentences' transitivity in four conditions. However, because this study did not focus on the transitivity, the experimental sentences were chosen equally in each condition of Hijikata's set (see Appendix D for the whole list). Each sentence includes three or four chunks, and each chunk includes 3.48 words on average.

In this pre-pilot study, two participants performed the processing unit task to

confirm procedure and time limitation. As they majored in English education, their reading proficiency was relatively higher than other undergraduate participants. The two participants pointed out that the processing unit span test with five seconds was so easy that the ceiling effects would occur (i.e., when almost all participants achieve high scores, the differences of scores between participants do not appear statistically). Therefore, this study added the processing unit span test with a two-second condition, which is the same time as the shortest rehearsal time in working memory (Baddeley, 1986, 2000; Kadota, 2007). This two-second condition made the processing unit span more distinctive than the five- second condition, thus giving better and more accurate results. This study used both five seconds and two seconds to divide the participants into two groups according to their size of processing unit span.

4.1.2 Methods of pilot study

4.1.2.1 Participants

A total of 59 Japanese EFL undergraduate and graduate students participated on a paid volunteer basis. They majored in many different fields such as, humanities, Japanese language and culture, psychology, international studies, social science, engineering science, biological science, and informatics. None of them had been to English-speaking countries for some months or years. Two participants' sets of data were excluded from the following analyses, because they could not complete all of the tasks. An additional participant who had read four passages of the reading proficiency test before this pilot study was also excluded from analyses. Therefore, the final number of data sets was 56.

4.1.2.2 Materials

In this pilot study, two types of materials were prepared: reading proficiency test and processing unit test. Sentences for the processing unit span test were the same as in Pilot Study 2. The test consisted of 20 sentences. Ten of these sentences were used in the five-second condition, and the other 10 sentences were used in the two-second condition. The reading proficiency test questions were prepared from the STEP test (STEP, 2008a, 2008b, 2009a). The reading proficiency test consisted of six passages and 23 questions (for the instruction and example passage, see Appendix E). The criteria for choosing passages were that the readability and length of passages were average and representative of each grade of the STEP test.

4.1.2.3 Procedure

A whole procedure was conducted individually or in a group of a few participants. This study includes two sessions: the processing unit span test (Session 1), and a reading proficiency test (Session 2).

First, the participants performed a processing unit test to measure their processing span of texts. Sentences of this test were shown on the screen in front of each participant. The participants were told to write down what they remembered of the sentence in English, as soon as the sentence disappeared from the screen. After a practice session, the participants performed the test under the five-second condition. All participants had finished recalling within one minute on average. After the five-second condition, the processing unit span test under the two-second condition was carried out. The participants were told that the presentation time would be shorter. The recall task of each sentence took approximately 30 seconds on average. Each task had a practice session before the test began, in order for the participants to become accustomed to performing the tasks.

They read 10 sentences in five seconds and then read the other 10 sentences in two seconds. Although the sets of materials were counterbalanced within learners, the two-second condition always preceded the five-second condition, and thus the order of conditions was not counterbalanced. Figure 4.1 shows the procedure of processing unit tests.

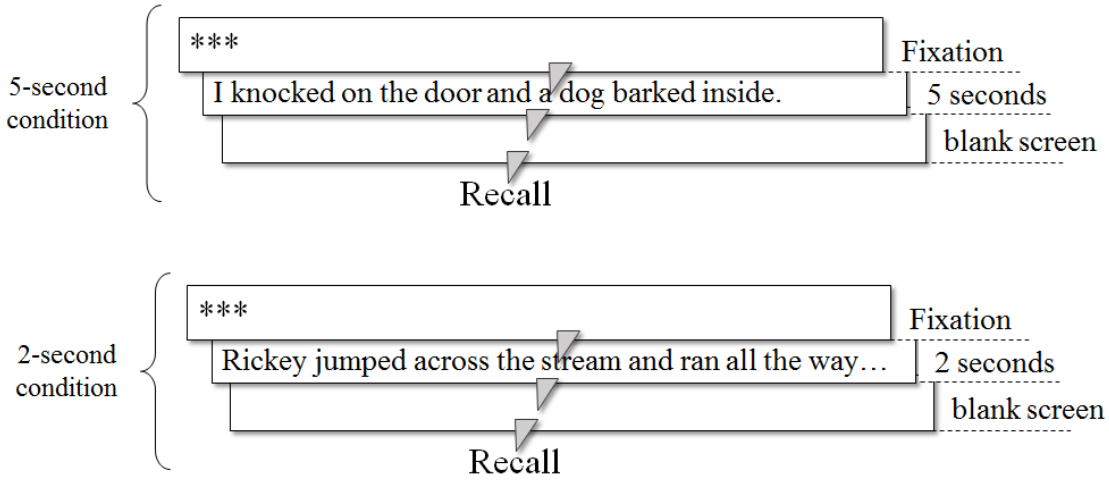

Figure 4.1. Procedure of processing unit test of 5- and 2-second conditions.

Finally, in Session 2, a reading proficiency test was conducted. The participants were told that they had to complete the test within 30 minutes, and that the test began with the easiest items first and increased in difficulty as the test progressed. After finishing, they were asked whether they had read each passage in the reading proficiency test. The whole procedure took approximately one hour to complete.

4.1.2.4 Scoring and data analyses

The following two types of data were scored for analysis: (a) the reading proficiency test and (b) the processing unit test.

Processing unit test.

The processing unit test was scored two ways based on Hijikata (2012): (a) the number of the recalled words, and (b) the number of the recalled phrase-based chunks. As for the former scoring, the content words that were defined by Fries (1952) were counted. Chunk division of the experimental sentences was based on Hijikata. The final scoring of Hijikata focused on the verb types, but the study omit this scoring procedure. Based on Hijikata (2012, Experiment 4), participants' protocols were scored for each phrase-based chunk. First, two raters assessed production of each word in experimental sentences. Following the criteria of function words revised by Hijikata (original criteria was from Fries, 1952), all function words in each chunk was excluded from scoring beforehand. When learners' written protocol included all the remaining content words in a chunk, the raters scored one point (the inter-rater agreement = 97.99%). No score was given in the following cases: (a) recalled chunk including serious spelling errors, (b) any word inconsistent with original word in terms of both word meaning and part of speech, and (c) changing word order beyond a chunk boundary.

Reading proficiency test.

The reading proficiency test contained reading sessions of three-level of grades in Eiken-test: the pre-second (k = 8), second (k = 9), and pre-first grade (k = 6). Each grade included two passages. To compare reading proficiency level with processing unit test, accuracy rates in each grade level were calculated as well as the rates of the whole proficiency test. These scores of reading proficiency test and processing unit test were analyzed by correlation analysis.

4.1.3 Results of pilot study

4.1.3.1 Reading proficiency test

Tables 4.1 and 4.2 show the descriptive analysis and results of each grade of the reading proficiency test. The participants were divided into two groups according to their reading proficiency test scores, and a *t* test confirmed the validation of the groups, $t(54) = 9.04$, $p < .001$. Table 4.2 indicate that the upper group could comprehend pre-second and second grade-level texts easily, and they answered correctly one-third of pre-first grade-level texts, which meant that their reading level was assessed as second grade or above on the STEP test. On the other hand, the lower group failed to comprehend pre-first grade-level texts, and some of them could not comprehend well even at a second grade level. Therefore, the EFL learners in the lower group could be labeled as below second grade-level-readers of the STEP test. The scores in Table 5.2 were normalized by making the angular transformation since each grade included a different number of items. The reliability scores for the whole items and each grade were relatively high excepting the pre-second grade (Cronbach α = .79 for the whole items, .63 for the pre-second grade, .70 for the second grade, .76 for the pre-first grade).

Table 4.1

Descriptive Statistics of Reading Proficiency Test in Experiment 4

Proficiency	*n*	*M*	*Min*	*Max*	*SD*	95% CI
Upper	26	17.42	14.00	23.00	3.06	[16.19, 18.66]
Lower	30	11.13	6.00	13.00	2.11	[10.34, 11.92]
Total	56	14.05	6.00	23.00	4.08	[12.96, 15.15]

Note. Maximum possible score = 23. The cutoff point between the Upper and Lower groups was 14.

Table 4.2

Details of Reading Proficiency Test in Experiment 3

	n	Total	Pre-2	2nd	Pre-1
Upper	26	73.59 (12.40)	93.75 (9.56)	84.38 (13.27)	36.02 (30.76)
Lower	30	48.67 (9.15)	75.81 (18.80)	51.25 (19.18)	8.60 (13.52)
Total	56	62.68 (17.69)	84.81 (16.90)	68.55 (24.02)	24.35 (28.88)

Note. The accuracy rates were angular transformed.

4.1.3.2 Processing unit test

Table 4.3 shows the results of processing unit test in 5000 and 2000 ms conditions. To examine the relationships between learners' proficiency and processing unit test, a 2 (Proficiency: Upper vs. Lower) × 2 (Time condition: 5000 vs. 2000 ms) two-way mixed ANOVA was conducted. The results found out the insignificant interaction between proficiency and time condition, $F(1, 54) = 0.25$, $p = .623$, $\eta^2 = .01$; the marginally significant proficiency, $F(1, 54) = 3.78$, $p = .057$, $\eta^2 = .07$; and the significant second condition, $F(1, 54) = 162.57$, $p < .001$, $\eta^2 = .75$; as shown in Table 4.4.

The differences between time conditions were apparent, and also reading proficiency was related to the processing unit test. The results supported the findings in Hijikata (2012). The processing unit test that requires comprehenders to encode some chunks at one view with strict time limitation can be efficiently worked as one of the reading proficiency indices. The following section will analyze the relation between two tasks more specifically.

Table 4.3

Means With Confidence Intervals and Standard Deviations of Processing Unit Test (%)

		5000 ms		2000 ms	
	n	*M (SD)*	95%CI	*M (SD)*	95%CI
Upper	26	73.38 (24.65)	[63.43, 83.34]	54.08 (20.20)	[45.92, 62.24]
Lower	30	63.87 (21.79)	[55.73, 72.00]	43.00 (15.32)	[37.28, 48.72]
Total	56	68.29 (23.44)	[62.01, 74.56]	48.14 (18.45)	[43.20, 53.08]

Note. Upper and lower division was based on the reading proficiency group.

4.1.3.3 Relationship between reading proficiency test and processing unit test

In order to investigate the detailed relationship between the proficiency test and the processing unit test, the correlation analysis was conducted as seen in Table 4.5. In particular, this analysis focused on the processing unit and each grade level of the proficiency test, when the result of ANOVA above could not be figured out.

Table 4.4

Summary Table for Two-way ANOVA of the Effect of Reading Proficiency and Time Condition

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
			Within-participants			
Time condition (T)	1	1.12	1.12	162.57	<.001	.75
T × Reading proficiency	1	0.00	0.00	0.25	.623	.01
Error (Time)	54	0.37	0.01			
			Between-participants			
Reading proficiency (P)	1	38.24	38.24	3.78	.057	.07
Error (Proficiency)	54	4.22	0.08			

Focusing on Items 5 and 6 (i.e., processing unit test of 5000 and 2000 ms conditions), the scores in the 2000 ms condition had the stronger and more significant correlation with reading proficiency test than those in the 5000 ms condition. The correlation scores of pre-second grade was low, which might be caused by lack of reliability of this item. The other correlation scores indicated the apparent advantage of the 2000 ms condition. That was not surprising, as high score in strict limited time means rapid processing, and processing time is a feature of proficient readers (Grabe, 2009). According to the results of ANOVA and correlation analysis, Study 2 will use the 2000 ms condition as a learners' individual factor as well as the other individual differences.

Table 4.5

*Correlation Between Reading Proficiency Test and Processing Unit Test (*N = *59)*

	1	2	3	4	5	6
1. Proficiency test (whole)	—					
2. Proficiency test (Pre-2nd)	.614**	—				
3. Proficiency test (2nd)	.882**	.335*	—			
4. Proficiency test (Pre-1st)	.799**	.260	.626**	—		
Processing unit span test						
5. 5000 ms condition	.151	.297*	.406**	.330*	—	
6. 2000 ms condition	.557**	.299*	.533**	.547**	.871**	—
M	52.64	69.94	55.90	22.88	68.29	44.93
SD	26.88	16.30	17.86	23.98	23.44	21.54

Note. **$p < .001$, *$p < .05$. The critical cells were highlighted.

4.2 Experiment 3: Learners' Use of Strategy in Local-ambiguous Sentence Processing

4.2.1 Purpose and research questions of Experiment 3

Experiment 3 will focus on the learners' individual strategy use and difference of strategy use according to their individual competence. The behavioral data show readers' online processing, but certain criteria are always required (e.g., when readers process disambiguated regions in VP condition faster than that in NP condition, readers may be interpreted as VP-attachment preference). In order to disclose readers' specified interpretation, other methods are desirable to complement the prior findings from the behavioral data. To resolve this problem, the present study utilizes the think-aloud method. This method allows us to observe comprehenders' gradual processing, detailed thinking or strategy use. Although the think-aloud method is frequently used for text reading or strategy studies, few studies applied this method to sentence comprehension which is processed quite automatically. If learners do not have any difficulties in comprehending experimental stimulus such as global ambiguous sentences, their oral report cannot reflect learners' automatic (i.e., speedy) processing. To emphasize processing difficulties of sentences, the study will focus on local ambiguous sentences including garden-path sentences instead of PP-attachment sentences (global ambiguities) as used in Study 1.

The first purpose of Experiment 3 is to examine whether learners process ambiguous/unambiguous sentence using individually different strategies. Prior studies have investigated various models or theories as shown, but other processing tendencies or strategies are possible to observe.

The second purpose is whether learners' individual factors affect each strategy use. As individual factors, reading proficiency test and processing unit as in the pilot study will be used.

The final purpose is to examine whether the think-aloud method is available for sentence processing studies.

RQ 3-1:	Do EFL learners with different processing units use different strategies in local ambiguous sentence resolution?
RQ3-2:	How does the think-aloud protocol differ according to the success and failure of sentence comprehension?
RQ 3-3:	Is the think-aloud method effective for sentence processing studies?

RQ3-1 predicts that learners with different processing units will parse sentences differently. The previous studies discussed that the different proficient readers decided different attachment sites due to their differences in cognitive resource (Rah, 2009; Rah & Adone, 2010a). The reading and grammatical proficiency of English are not equivalent to the processing unit, but are closely related to each other. The processing unit test requires readers to process and remember sentence or chunks at the same time in the limited times, thus learners with high proficiency or high processing speed can answer more units than ones with poor proficiency as discussed in the pilot study.

Regarding RQ3-2, prior studies have investigated general tendencies of learners in sentence processing such as shallow structure hypothesis. Learners do not always find out answers in the same way as native speaker, because learners give semantic information importance while native speakers could parse faster. However, some studies focused on readers' processing through behavioral data which did not allow us to show readers' specific parsing strategies. Nonetheless, the experimenter should infer what learners think during reading.

As for RQ3-3, the previous studies so far have rarely utilized the think-aloud method because some researchers thought that such low-level processing as sentence parsing did not suit the think-aloud method, since the lower-level processing is usually too automatized to describe what readers think about compared with text processing: This is one of the deficiencies of the think-aloud method. In addition, proficient learners whose parsing processes are highly automatized and rapid could only mention the final correct interpretation, and did not mention some previous steps in process or rejected interpretation even when they activated multiple interpretations and examined the plausibility of each interpretation. On the other hand, the advantage of the think-aloud method is to know readers' representation specifically during sentence processing. In addition, learners will show the processing strategies both explicitly and implicitly through the methods.

4.2.2 Methods of Experiment 3

4.2.2.1 Participants

Nine undergraduate students were recruited from the same population as in the pilot study. One of the participants failed to complete all the tasks, and three sets of data were excluded due to instrumental problems; thus, there were seven final data sets.

4.2.2.2 Materials

Two types of experimental sentences were prepared (see Appendix F): seven temporal ambiguous sentences with reduced relative clause and seven unambiguous sentences with unreduced relative clause (Kadota & Kuramoto, 2007; Rah, 2009). These short sentences consisted of approximately 10 words. Experimental sentences with reduced relative clauses (e.g., *The woman sent a doll was very excited*) caused a garden

path effect because the first verb *sent* can be interpreted as the main verb in initial parsing. On the other hand, the other sentences with unreduced relative clauses included no syntactic ambiguity (e.g., *The cook who was criticized by the customers used too much salt*).

A variety of 40 filler sentences were also added but were not treated as target sentences. The filler sentences included the objective sentences with reduced relative clause (RRC) and the sentences with unreduced relative clause (URC) and other types of ambiguous and unambiguous sentences (e.g., ambiguous conjunction sentences without commas: "*Before the mother ate the candies had disappeared.*" versus "*Before the mother ate, the candies had disappeared.*").

4.2.2.3 Pilot study for Experiment 3

Before the main experiment, the pilot study for Experiment 3 was carried out to verify the think-aloud procedure and revise the categories of protocols. Since the think-aloud method is rarely applied to studies on sentence processing, the present study revised the categories of Horiba (1996, 2013). Horiba originally set the following categories for lower processing: (a) *graphophonemic/graphomorphemic analysis*, (b) *word recognition*, and (c) *syntactic/ semantic analysis of clause or sentence*. According to the results of this pilot study, the categories might be deleted, added, or changed.

Procedure of the pilot study.

In the pilot study, the four participants read the same sentences as in the experiment: They read the sentences and performed the think-aloud procedure in which they orally reported what they were thinking. Before they started the procedure, they completed a practice session to get used to it. In the practice session, the participants

listened to a sample recording demonstrating the think-aloud technique using two filler sentences. Then they practiced the same two sentences and two new sentences using the think-aloud method. When they finished thinking aloud about a sentence, they pressed a key to read the following sentence at their own pace. After the participants understood the method, they proceeded to the think-aloud session. The sentences were presented randomly, with at least one filler sentence following every two target sentences. The participants were not allowed to remain silent for longer than five seconds. Instead, they were asked to read aloud the parts of sentences if they pronounced them in their minds or to explain what they found difficult to understand. When the participants expressed no words for approximately five seconds, the experimenter asked them to express their thinking during both practice and think-aloud sessions.

Results of the pilot study.

Throughout the pilot study, the categories of graphophonemic/graphomorphemic analysis in Horiba (1996) that never appear in learners' performance were deleted. In addition, the category of word recognition was integrated into semantic analysis because the participants often referred to the word meanings in Japanese after reading their English translation. The detailed categories and examples of the protocols are shown in Table 4.7.

4.2.2.4 Procedure

The experiment was divided into two sessions: In the first session, participants performed the reading task and then, they performed the remaining tasks in the second session.

In the first session, the participants performed the sentence reading task after they were instructed on general aims of the experiment. All of the sentences were presented in the center of a computer screen, sentence by sentence. The presentation order was randomized in each participant to avoid learning effects. The participants performed the think-aloud method while reading the sentences. They were asked to speak what they thought. After thinking each sentence aloud, they also spoke the Japanese translation of the sentence, as a final interpretation. The set of think-aloud and Japanese translation tasks was repeated for each sentence. Any feedback on the translation task was not provided to participants. All of their utterances were recorded. The first session took an hour and a half at longest.

In the second session, the participants performed the reading proficiency test with 30 minutes. Then, they took the processing unit test with the same procedure as the pilot study. Though no time limitation was set, they had finished the processing unit test for less than 30 minutes. The second session took approximately one hour.

4.2.2.5 Scoring and data analyses

In this experiment, three types of data were collected and scored: (a) assessment of reanalysis, (b) translation task, and (c) think-aloud protocols. As for the final indication, think-aloud protocols, this study attempted two versions of categorization or analyses.

Assessment of the reanalysis.

In order to investigate whether the participants re-analyzed the syntactic structure during the think-aloud protocols, two raters scored it independently. They discussed any discrepancy, and they defined the criteria for determining reanalysis as follows: (a) Reanalysis should contain two different interpretations at least. The

utterances of notice without any revised interpretation were not regarded as reanalysis (e.g., *Ah, my interpretation might be wrong*.); (b) Reanalysis should be related to critical structures of the experimental sentence. The references to other detailed structures or word meanings were not counted as reanalysis, even if those utterances led to revised interpretation.

Translation task.

Answers of the translation task were regarded as the participants' final interpretation of the syntax. These answers were rated from the perspective of syntactic accuracy: when the learners recognized the syntactic structure, they were awarded one point even if there was some trivial inconsistency in meaning; if they made any mistake in the syntactic structure of the sentence, they were awarded no points. These ratings were conducted independently by two raters, and any discrepancies were also discussed in order to resolve them, as shown in Table 4.7. The scores on the translation task were summed; the maximum possible score was seven.

Categorization of think-aloud protocols (1).

The participants' protocols were transcribed and then parsed into statements, which were almost equivalent to the clauses based on Horiba (1996, 2013). These statements were classified into the following six categories (see Table 4.6): (a) semantic analysis, (b) syntactic analysis, (c) context information, (d) metalinguistic comments, (e) reading aloud, and (f) others. The number of protocols in each category used by the participants showed their sentence-processing strategy.

Table 4.6

Categories of Think-aloud Protocols

Category	Definition
1: Semantic analysis	The reader attempts to analyze the semantic features of the linguistic item (e.g., meaning of words or phrase, and word meaning inferences). It also includes L1 translation and paraphrasing of the sentence.
2: Syntactic analysis	The reader attempts to analyze the formal features of the linguistic item (e.g., identifying the subject and verb, and referring to segmentation or modification of phrase)
3: Contextual analysis	The reader attempts to make the inferences beyond sentence or phrases.
4: Metalinguistic comments	The reader attempts to identify the grammatical term (e.g., reduced relative clause, RRC).
5: Reading aloud	The reader read aloud the whole sentence.
6: Others	The reader makes a comments to show their self-monitoring and strategic reaction.

Categorization of think-aloud protocols (2).

This analysis categorized the protocols into four different patterns using the following three points of view: (a) whether readers reach the proper interpretation in initial parsing; (b) among readers who make mistakes in initial parsing, whether they recognize inconsistency in their own interpretation; and (c) among readers who recognize inconsistency between their interpretation and the syntactic representation, whether they re-analyze their interpretation (including correct and incorrect interpretation of the

reanalysis). As shown in Table 4.7, Pattern 1 included any reanalysis and learners' initial interpretations were correct. In Pattern 2, learners performed reanalysis after they noticed their initial interpretation was not appropriate for the syntactic structure of the experimental sentences. Protocols in Pattern 3 were not correct, yet participants did not notice the ambiguity or their incorrect interpretation. Finally, in Pattern 4, learners recognized that their interpretations were somehow inconsistent with the sentence structure, but they could not resolve sentence ambiguity correctly.

Table 4.7

Patterns of think-aloud protocols

		Reanalysis	
		None	More than once
Final interpretation	Correct	Pattern 1	Pattern 2
	Incorrect	Pattern 3	Pattern 4

Data analysis.

To interpret the result of think-aloud protocols, two types of analysis were applied in this experiment: (a) quantitative analysis including correlation analysis, and (b) qualitative analysis (i.e., case study). In the first quantitative analyses, frequencies of each category in the protocol data were counted and compared with learners' individual factors in each sentence construction. The qualitative analyses will explore which sentence-features make participants resolve ambiguous construction and which ones remain unresolved.

4.2.3 Results of Experiment 3

Table 4.8 shows the example protocols of the example RRC sentence (*The man warned last night regretted what he did*). Prior to the main analyses, participants' factors were demonstrated in Table 4.9. Table 4.10 summarized the frequencies of use of each category in the protocol data and the correct response rate and reanalysis (see Figure 4.2).

Table 4.8

Example Protocols in the Experimental Sentence: The man warned last night regretted what he did (RRC condition)

Category	Examples of protocols
1: Semantic analysis	"*warned* means he warned." "I do not understand what this word implicates very well…"
2: Syntactic analysis	"Phrases till *last night* must be the subject of this sentence…" (subject and verb of sentence)
	"*The man warned*, the sentence was segmented at this point…" (segmentation of phrase)
	"*the man* is modified by *warned last night*." (modification relation)
3: Context Information	"He regretted what he did, though he was warned."
4: Metalinguistic comments	"The relative pronoun is reduced here."
5: Reading aloud	"*The man warned last night regretted what he did.* [The reader read aloud the whole sentence without any pause]"
6: Others	"I'm thinking now to translate this sentence."

Note. The words in italics refer to the original English sentences that helped to interpret the examples of the protocols. The other parts of the protocols were originally in Japanese.

Category 1 includes word inferences or comments on words as well as L1 translation of each word. Comments or self-monitoring on words were distinguished from comments on sentence structures. Category 2 consisted of three typical protocols that refer to the subject and verb of the sentence, modification relation, and phrase segmentation. Category 3 counted inferences of the situation, context or precondition of the sentence that are not explicitly explained in the sentence. The protocols in Category 3 were rarely observed especially in URC sentences. Category 5 usually occurred at the outset of utterances. It should be noted that protocols of phrasing sentence with adding L1 translation were counted as semantic analysis, while phrasing is one of the syntactic processing.

Table 4.9

*The Features of Participants in Experiment 3 (*N = *8)*

	M	*SD*	*Min.*	*Max.*	95%CI
Reading proficiency test	16.25	3.96	12.00	23.00	[12.94, 19.56]
Processing unit test					
5000 seconds (%)	64.64	14.13	48.00	88.00	[51.57, 77.71]
2000 seconds (%)	43.33	16.43	16.00	67.00	[28.14, 58.53]
Average (%)	53.99	13.52	34.00	72.00	[41.48, 66.49]

Note. The possible maximum score of reading proficiency test = 23.

Table 4.10

*Summary of Frequency Data in Each Sentence (*N = *8)*

	RRC		URC	
	M	*SD*	*M*	*SD*
C1: Semantic analysis	0.87	0.53	1.92	0.87
C2: Syntactic analysis	1.32	0.62	1.44	0.54
C3: Context	N.A.	N.A.	0.14	0.38
C4: Meta linguistic	0.10	0.06	0.71	0.49
C5: Oral reading	0.97	0.36	0.88	0.39
C6: Others	0.71	0.07	0.29	0.49
Correct response (%)	78.63	14.37	98.43	1.57
Reanalysis	0.06	0.03	N.A.	N.A.

Note. Category 3 in RRC and reanalysis in URC were not observed.

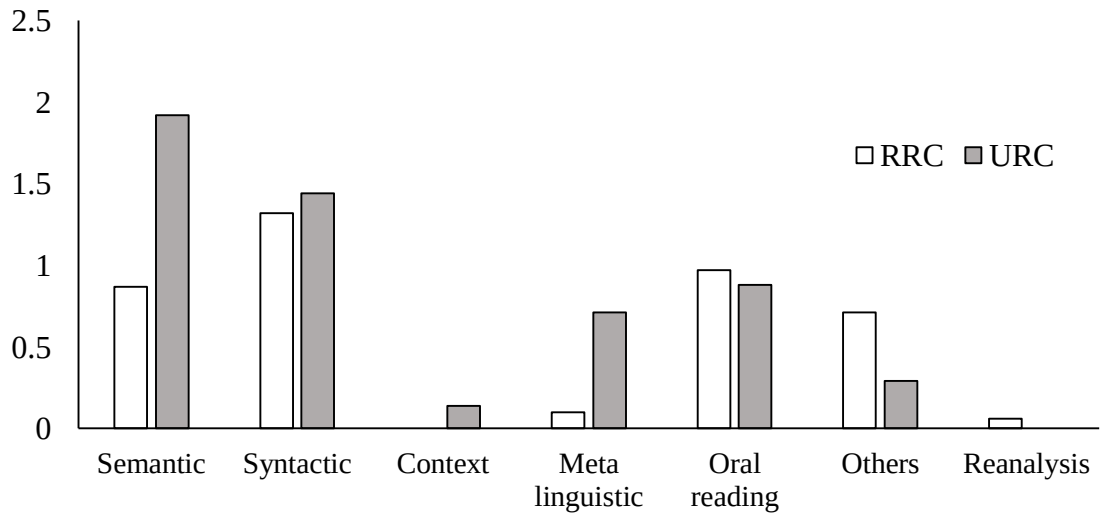

Figure 4.2. Average number of times that participant report each category in each sentence. RRC = Reduced relative clause, and URC = Unreduced relative clause.

4.2.3.1 Results of quantitative analyses

Correlation analysis.

To examine the relationships between individual factors and think-aloud protocols, correlation analysis was carried out. Since the frequencies of Categories 3 and 6 in the reduced condition were quite low, these two categories were excluded from correlation analysis. The results of the remaining four categories and individual factors were shown in Table 4.11a. In the unreduced condition, protocols of reanalysis and Category 6 were rarely observed, thus the two categories were excluded from analysis and Category 3 was also included in the analysis unlike the reduced condition as shown in Table 4.11b.

One of the salient findings in the RRC condition is that the number of reanalyses was negatively and strongly correlated with the learners' sentence processing unit (r = -.72, p = .046). The RRC sentences often requires comprehenders to reanalyze their interpretation, especially when the verb in relative clause could have been analyzed as the main verb: Namely, the sentence includes such ambiguities as *The man* <u>*warned*</u> *last night regretted what he did*. In these sentences, learners who have large processing unit rarely report the reanalysis. This is probably because learners with large processing unit could process sentences automatically, and revising their initial processing was included in earlier phase where the grammatical structure could not be explained.

The other significant correlations were not related to learners' individual differences: the strong and positive correlations between semantic and syntactic analyzing, and between semantic analysis and meta-analysis. The participants frequently read one of the phrase immediately after they translated into L1 words or phrases. The tendency should cause the correlations between semantic and grammatical processing.

As for the URC condition (i.e., unambiguous sentence), number of reanalysis

data were not enough to include the variable in analysis because readers need not reanalyze the sentence. The analysis in the URC condition found only two significant relationships: strong positive correlation between reading proficiency and semantic analyzing (r = .78, p = .031), and strong negative correlation between processing unit and syntactic analyzing (r = - .83, p = .021).

The correlation of the overall scores (i.e., accuracy rates of final interpretation in each condition) was found only with reading proficiency in the RRC condition. Non-correlation with other measures in protocols suggested that frequency data of the think-aloud protocols did not reflect whether learners comprehended sentences successfully or unsuccessfully.

Table 4.11a

Correlation Analysis of frequency data and learners' individual factors in RRC condition

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Proficiency test	—									
2. 5000 seconds (%)	.50	—								
3. 2000 seconds (%)	.44	.53	—							
4. chunk unit (%)	.54	.90**	.85**	—						
5. C1: Semantic	.44	-.29	-.09	-.23	—					
6. C2: Syntactic	.02	-.29	-.33	-.35	.81*	—				
7. C4: Meta	.69†	-.12	.20	.18	.74*	.52	—			
8. C5: Oral reading	-.44	-.18	-.30	-.27	-.23	.13	.10	—		
9. Reanalysis	-.22	-.69†	-.65†	-.72**	.50	.63	.30	.36	—	
10. Overall score	.69†	.52	.00	.33	-.06	-.37	.12	-.48	-.33	—
M	16.25	60.00	42.49	51.15	0.87	1.32	0.10	0.97	0.06	0.79
SD	3.96	18.54	15.49	14.88	0.53	0.62	0.18	0.36	0.03	0.14

Note. *$p < .05$, **$p < .01$, $^{\dagger}p < .01$. Maximum possible scores of the reading proficiency test = 23.

Table 4.11b

Correlation Analysis of frequency data and learners' individual factors in URC condition

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Proficiency test	—									
2. 5000 seconds	.50	—								
3. 2000 seconds	.44	.53	—							
4. chunk unit	.54	.90**	.85**	—						
5. C1: Semantic	.78*	-.13	.31	.26	—					
6. C2: Syntactic	-.29	-.20	-.83*	-.61	.01	—				
7. C3: Context	-.56	-.06	-.27	-.13	-.21	.46	—			
8. C4: Meta	.42	-.04	-.33	-.22	.46	.49	.25	—		
9. C5: Oral reading	-.19	.45	-.20	.11	-.27	.29	.14	-.34	—	
10. Overall score	.56	-.06	.27	.13	.21	-.46	-.29	-.14	.25	—
M	16.25	60.00	42.29	51.15	1.92	1.44	0.14	0.71	0.88	0.98
SD	3.96	18.54	15.49	14.88	0.87	0.54	0.38	0.49	0.39	0.04

Note. *$p < .05$, **$p < .01$, $^{\dagger}p < .01$. The highlighted cells were identical to Table 4.11a.

4.2.3.2 Results of qualitative analyses

Think-aloud protocols were divided into four types (see also Table 4.7) according to final interpretation (correct vs. incorrect) and reanalysis (absence vs. presence), and each type was discussed individually as a case study. Prior to the case study, the results of overall pattern frequencies were shown in Table 4.12. Because the targeted RC attachment sentences did not have severe garden-path effects, few participants made errors in final interpretations. The URC sentences also did not require reanalysis, because they did not include any ambiguity or processing difficulty. In order to examine each pattern, more severe garden-path sentences were added, as the responses to RC attachment structure did not correspond with Pattern 4.

Table 4.12

*Frequencies of Each Pattern (*N = *9)*

	Pattern 1 (none/correct)		Pattern 2 (none/incorrect)		Pattern 3 (with/correct)		Pattern 4 (with/incorrect)	
	k	%	*K*	%	*k*	%	*k*	%
RRC	84	70.00	30	25.00	6	0.05	0	0.00
URC	57	91.94	5	8.06	0	0.00	0	0.00
Total	141	77.47	35	19.23	6	3.30	0	0.00

Note. RRC = Reduced relative clause and URC = Unreduced relative clause. The targeted sentences were 14 RRC and 13 URC sentences. Some data were missing that participants could not complete.

Example protocols of RRC.

The most diverse patterns were observed in the following sentence: This

sentence requires reanalysis of readers' interpretation compared with other RRC sentences. Some protocols of this sentence were shown in each pattern. The original think-aloud protocols were in Japanese (see Appendix G1). The sentences in italics indicated a reference to the original English sentences, while the other sentences were in Japanese.

Pattern 1 (Correct final interpretation without reanalysis):

(22a) The newspaper reported the death of the film star sells quickly. [S 10]

(22b) *The newspaper reported..., the death of the film, film star. The death of the film star sells quickly*. The subject have lasted here, or the subject lasts until *sells*.

(22c) The newspaper reported the death of the film star sells quickly (reading the sentence aloud).

Protocol (22b) showed that the participant pointed out mainly syntactic parsing, while the participant in (22c) just read aloud the sentence before she translated the sentence into Japanese. She reached the correct interpretation of the ambiguous sentence (as shown in underlined sentence) immediately after reading it aloud, as seen in (22c), but this did not imply that the participant could construct the proper syntactic representation without any mistakes in mind. A possible reason for this is that her processing was certainly automatized. The prediction was supported by the results of her processing unit test, which was the best of all in this experiment. The other participants rarely generated this pattern of protocol in any sentence, which meant that this pattern is a specific strategy for relatively proficient learners. Note that both participants in these examples did not translate phrases into Japanese.

Pattern 2 (Incorrect final interpretation without reanalysis):

(22d) *The newspaper reported the death of the film star sells quickly.* This sentence can be divided up after *the newspaper reported*. The remaining part was *the death of the film star sells quickly*. *The death of the film star*, the death of *the film star sells*, q*uickly*? Hmm, I've reached just vague understanding, but I could not translate this sentence appropriately. I think this sentence means that <u>the newspaper quickly reported the death of the film star</u>.

Example (22d) showed that the participant felt that something was awkward in these sentences, but she referred only to the problems in Japanese translation or meaning, and she did not examine the detailed syntactic structure.

Pattern 3 (Correct final interpretation with reanalysis):

(22e) *The newspaper reported*, the newspaper reported, *the death of the film star*, the death of the film star. *sells quickly*. This *sells* means… The verb appears here again. Ah, *reported the death of the film star* modifies *newspaper*.

(22f) *The newspaper reported the death of the film star sells quickly. The newspaper reported* means the newspaper reported. T*he death of the film star, sells quickly. The death of the film star sells quickly…* Ah, *reported the death of the film star* modifies *newspaper*, and *reported* modifies *newspaper*. It (*reported*) might be used as an adjective. Well, *sell* means sell, and *quickly* means quick or quickly.

In the example (22e), the participant detected the anomaly of the syntactic structure when he noticed the two main verbs in one sentence. Then, he reanalyzed sentences and found out the proper modification relation. Similarly, Protocol (22f) also reached the semantically correct interpretation at last, and the cues that allowed him to notice his wrong interpretation might be same as for (22e). However, learners might not always construct detailed structures, as the participant in (22e) pointed out in "It (*reported*) might be used as an adjective," which was actually used as RRC.

Pattern 4 (Incorrect final interpretation with reanalysis):

There were no available protocols in this pattern.

The reason that this sentence frequently requires reanalysis might be the length of the relative clause. The RC in the other RRC sentences consisted of three to five words (e.g., *The woman* <u>*sent a doll*</u> *was very excited* or *The king* <u>*delighted by the gift*</u> *was in good spirits*. See Appendix F for the full list), and some of them included a by-phrase in RC, which made it simple to parse the RRC.

Examples of protocols of URC.

Because the URC structures did not include any processing difficulties, protocols other than Pattern 1 were rarely found. However, the protocols of URC were worthy to observe from the points of methodological view of the think-aloud method, because URC sentences included no ambiguities, and learners were more likely to encounter the URC structure than the complicated RRC structure. In Protocol (23b), the participant clearly faced difficulties in comprehending the verb *spoil*. Although his final interpretation was correct, the protocol showed that his sentence representation could easily have the

opposite meaning, if he had not known the meaning of the verb *spoil*.

(23a) The teacher spoiled the student who got the best grades.

(23b) *The teacher spoiled the student who got the best grades*. *Spoil*? What is "*spoil*"? The teacher did something with the students who got good grades. *Spoil*…I don't think the word *spoil* has a positive connotation. Does it mean dispraise? If so, the meaning of this sentence is that the teacher dispraised the student who got the best grades. I think that the meaning (of *spoil*) is not praise, but dispraise.

The example that failed to reach the correct interpretation is shown below:

(24a) The old lady that you're always talking about has just gone into hospital.

(24b) Phrase *the old lady that you're always talking about* was the subject of this sentence, and *that*-phrase modifies *old lady*. (The lady that) you're always talking about, well the old lady *has just gone into hospital*. <u>The meaning of *gone* is that she has gone (to somewhere), and *hospital* is located posterior (to *gone*). It said that (she) has gone at the hospital</u>. Namely, I imagined that *old lady*, has died. In sum, the old lady that you're always talking about has died at the hospital.

In the protocol (24b), the reason for her incorrect interpretation must be overreliance to the pragmatic information of *had gone* in the underlined protocols. School grammar in Japan usually points out the difference of usage of *have (had) been* and *have (had) gone*: the phrase *have been* implied that subject will not come back. Her interpretation was unexpected; the sentence did not include any attachment ambiguities, or other possible

processing difficulties, and the other participants did not employ such pragmatic information to process this sentence. Therefore, this particular problem was discovered even in unambiguous sentences through the think-aloud method.

Pattern 4 was not found out in the RRC or URC condition; the example of the filler sentence that has more serious garden-path effect was shown instead of the targeted condition. In the next example, Participant B recognized his mistake in the first interpretation, as shown in the italicized sentence. Then, he attempted to reanalyze the target sentence, and he finally made a different interpretation, which was still not the correct interpretation.

(25a) While the woman was eating the creamy soup went cold.

(25b) *While the woman was eating the.. creamy soup, went cold*. Eh, *While the woman was eating the creamy soup* this phrase include *while*, which means "during somethings." *While the woman was eating the creamy soup* indicates that *the woman* "the woman," *creamy soup*. The sentence means that "she eats creamy soup", "she ate it". So, the phrase indicates that *while she ate*. The next phrase is *went cold*, though I do not understand the subject of this phrase *went cold* well. The meaning of phrase *went cold* is kinds of getting cold. *Soup went*.. Soup modifies *creamy* in the phrase *creamy soup*, and perhaps soup is also the subject of phrase that beginning with *went* at the same time. If *creamy soup* is the subject of the phrase *went cold*, the meaning of the phrase might be that the soup became cold. In sum, while the woman was eating the creamy soup, the soup became cold.

It should be noted that learners rarely met the sentence with such local ambiguity, but Protocol (25b) suggests that learners do not attempt to integrate semantic and syntactic information as discussed in Lim & Christianson (2013).

Individual differences of protocols.

Finally, individual differences of protocols were shown, which was not reflected in the quantitative data. The participants could be roughly divided into two types: grammar based and semantic based.

The learners who read sentences on a grammar basis frequently pointed out the subject, verb, or object in the sentences, or the boundary of the phrase and chunk. After they confirmed the overall construction, they translated these sentences into Japanese. One of the typical protocols are shown as follows (Sentence 26a is original):

(26a) The man questioned by the lady couldn't answer.

(26b) *The man questioned, by the lady couldn't answer*. The subject of this sentence is *The man questioned by the lady*, and subsequently *the man questioned by the lady couldn't answer*. As the man verb of sentence is *couldn't answer*, the sentence means that the man who were questioned by the woman could not tell his answer.

Learners on a semantic basis translated each phrase into Japanese and then summed up these Japanese translations to comprehend sentences in protocols. Considering that chunking competence is a part of syntactic processing (Clahsen & Felser, 2006a; Hijikata, 2012), these learners did use both syntactic and sematic information. However, they could not revise their construction using syntactic information, as follows:

(26c) After *the man questioned*, the boundary of phrase separated the sentence; the subsequent phrase is *by the lady couldn't answer*. The subject of this sentence is this *man*. In sum, the man wandered why the woman could not answer.

In this protocol (26c), the participant seemingly integrated some words into one phrase; thus the shallow syntactic information was processed, but she did not even read sentence aloud at the beginning of protocols. It implied that her syntactic processing was insufficient both in initial and reanalysis processing. She attempted to resolve their comprehension problems by changing the comprehensible Japanese translation without noticing that her sentence representation included anomaly.

4.2.4 Discussion of Experiment 3

4.2.4.1 Effects of individual differences on strategy use in think-aloud protocols (RQ3-1)

First of all, the general strategies observed in Experiment 3 were discussed in this section. The RRC induced initially incorrect interpretations, and the participants had to modify their initial interpretations. The RRC sentences caused a local inconsistent effect in good-enough processing: local interpretation in ambiguous sentences disturbed readers' global interpretation. Most learners recovered their initial incorrect parsing after they encountered an inconsistent syntactic structure in their initial interpretation, as shown in Examples (22d) and (22e) (*The newspaper reported the death of the film star sells quickly*), but some learners kept their ambiguity unresolved, even though they reported that they noticed the anomaly of their interpretations.

This study also supported the idea that L2 or EFL learners easily utilize lexical and semantic cues rather than syntactic cues, but the participants did not always ignore syntactic information entirely. They sometimes focused on such syntactic information as structure and phrasing, as seen in Patterns 2 and 3. However, once they parsed a structure, they would not reanalyze sentence representations with other information. This finding indicated that learners have difficulties in integrating syntactic, semantic, and lexical information, as discussed in Lim and Christianson (2013).

As for the effects of processing unit, the negative correlation between processing unit and reanalysis in RRC condition suggested that learners with large unit could resolve the ambiguity before they spoke out on the initial and incorrect interpretation. Another possibility is that they did not need to reanalyze the ambiguous sentences, which is not unlikely, because such sentences are almost always initially wrongly analyzed by even L1 readers. The differences between learners who need and do not need reanalysis must be the automatization of syntactic processing. The syntactic processing of learners without reanalysis might be carried out too rapidly and effortlessly to report their parsing in real time. In addition, they might use implicit grammatical knowledge instead of explicit knowledge. In such a case, learners with implicit knowledge cannot verbalize their grammatical processing (Ellis, 2005).

On the other hand, the URC-sentence condition, which had more conventional sentences than the RRC, showed marginally different results. The reading proficiency test score correlated highly with semantic analysis in protocols.

4.2.4.2 The differences of protocols based on the success or failure of sentence comprehension and with or without reanalysis (RQ3-2)

According to the correlation analyses and categorization of protocols, the differences between successful and unsuccessful protocols were not apparent in the frequencies of the six categories among the same participants. Most unsuccessful protocols did not include noticing inconsistencies between learners' interpretation and the original sentence structure in protocols. Instead, there were differences of protocols among learners. The learners processed the sentences differently, and the differences were based on the degree of their processing automatization. The reasons for miscomprehended sentences varied: Some sentences were miscomprehended due to failure in syntactic parsing, while other protocols showed overreliance on semantic and pragmatic cues. The latter protocols might occur when they do not utilize syntactic information appropriately.

The structure condition did not include sufficient protocols with reanalyzing for comparison. Unless learners could not report reanalyzing of sentences due to their automatized processing, they rarely reanalyzed the sentence while reading. Therefore, the protocols did not show any apparent differences. The reason learners did not attempt to reanalyze the local ambiguous sentences might be due to the task features of task as discussed in Study 1. The local ambiguous sentence would confuse readers during their initial processing, and readers had to reanalyze sentence structures in order to modify their sentence representation. However, the translation task conducted after parsing sentences did not require the participants to disambiguate sentence structures. Such features of transition task led the participants to retain the sentences unresolved, even when they noticed anomalies in their sentence representation.

4.2.4.3 The effectiveness of the think-aloud method for sentence processing studies (RQ3-3)

In the think-aloud protocols, participants reported progress in their sentence constructing representation. The most effective factors of the think-aloud method were to identify where the processing difficulties lay, how learners recovered from the difficulties, and what information to utilize in order to recover from inconsistent representation, as seen in Protocol (25b). Although not every syntactic processing could be observed, it clarified the places where readers experienced serious processing difficulties (*I do not understand the subject of this phrase*) and discarded their initial interpretation before reanalysis (*If* creamy soup *is the subject of the phrase went* cold*, the meaning of the phrase might be…*). The think-aloud protocols also reported unexpected interpretations from learners, which was not determined from reading time data such as self-paced reading or eye-tracking studies. For example, (25c) arrived at the incorrect answer, which was unexpected, since the sentence (25a: The man questioned by the lady couldn't answer) included local ambiguities, the most likely error was *the man questioned*, while the correct structure is *the man couldn't answer*.

However, there were some deficits in the think-aloud method, as discussed in Section 4.2.4.1. The results of the correlation analysis indicated that learners with large processing units reported the syntactic and semantic information infrequently. One of the participants' protocols that reported the translations immediately after oral reading supported this result (Participant G). Her processing units were the largest among the participants in Experiment 3; thus, learners with automatized syntactic processing could not be completely observed. In addition, it also implied that the automatized processing of readers with relatively small processing unit could not be observed in think-aloud protocols. Therefore, the think-aloud method was effective especially for strategic

utterances.

4.2.5 Conclusion of Experiment 3

Experiment 3 aimed to investigate the specific processing when learners attempted to parse complex sentences of RRC structures. This experiment investigated (a) how local ambiguous sentences affect learners' sentence processing and induce parsing strategy, (b) the differences of think-aloud protocols when learners success or fail to achieve correct interpretation and when learners process sentences with or without re-analyses, and (c) to confirm the effectiveness of the think-aloud method for sentence processing studies.

Certain kinds of reader-based parsing strategy were observed, although the results should not be overly generalized because of the limited samples in this experiment. As for learners' general parsing strategies, learners with large processing units induced fewer protocols. They rarely reported reanalyzing temporally ambiguous sentences in their protocols, which were always correctly resolved after reanalysis. The differences in protocols, correctly and incorrectly answered, were not revealed in the frequencies of six categories in think-aloud protocols. As for the effectiveness of the think-aloud method in sentence processing studies, there were advantages, in that experimenters could observe detailed sentence representation and changing processes of representation, but the method did not reflect automatized or rapid parsing.

4.3 Experiment 4: Learners' Use of Strategy in Syntactically Ambiguous Sentence Processing

4.3.1 Purpose and research questions of Experiment 4

Experiment 3 investigated local ambiguous sentence or strict garden-path structures using the think-aloud method to know learners' strategy use in sentence processing. The results demonstrated that the think-aloud method was effective for investigating sentence processing with some exceptions. Experiment 4 focused on the global ambiguous sentences that do not have clear inconsistency or ambiguous points compared with local ambiguous sentences. As for the task after sentence reading, the comprehension questions in Experiment 3 did not require the ambiguous resolution, which led learners use underspecification strategy. To eliminate the influence of the test taking strategy, Experiment 4 adopted another style of reading that requires disambiguation of attachment sites; in addition, the feedback to their responses was given to confirm the correct answer of comprehension questions. In accordance with the purposes, the following three RQs were set:

RQ4-1 Do EFL learners show attachment preference in the think-aloud method?

RQ4-2 Do EFL learners sensitively respond to material distance in the think-aloud method?

RQ4-3: How does the think-aloud protocol differ according to the success and failure of sentence comprehension?

RQ4-4: Do EFL learners with different processing units use different strategies in syntactically ambiguous sentence resolution?

RQ4-5: How does the think-aloud protocol differ according to the task types?

Regarding RQ4-1, the syntactically ambiguous sentences in the present experiment did not include serious processing difficulty. Therefore, the explicit reanalyzing would not occur, but learners could explain their own strategies of PP-attachment sites, which might be effective for some questions of Study 1. RQ 4-2 examined whether the trigger of success or failure of sentence comprehension appears on learners' protocols; and how learners attempt to resolve the ambiguous attachment site when they cannot comprehend sentences. RQ 4-3 investigated differences between the protocols when they successfully and unsuccessfully comprehended sentences. The same questions in Experiment 3 did not work enough, but the questions would provide useful information regarding learners' representations. RQ 4-4 was related to the effects of individual differences on learners' sentence comprehension or protocol performances. This study used the processing unit test and grammar proficiency test as learners' individual differences. Additionally, participants' protocols were divided according to the types of protocols, and the features of the groups were investigated. Finally, RQ 4-5 compared think-aloud protocols when learners were given the translation task (Experiment 3) or verification tasks with corrective feedback. Study 1 showed that ambiguous regions were remained unresolved when learners did not have to disambiguate sentences. Similarly, it is predicted that learners could disambiguate sentences when they were given verification tasks with corrective feedback which requires learners to choose the attachment site of the sentences.

4.3.2 Methods of Experiment 4

4.3.2.1 Participants

A total of 21 ungraduated students participated in the Experiment 4. The participants were derived from the same pool of participants in Experiments 1, 2 and 3. No student

had participated in the other experiments in the dissertation studies. Their major varied. Data of two students were excluded due to recording difficulty; the final number of data sets was 19.

4.3.2.2 Materials

Processing unit span test.

A total of 10 sentences was prepared for each second-condition as in the pilot study and Experiment 3. In the same way as those experiments, the presentation order of sentences was counterbalanced, while the presentation of limitation times was not counterbalanced; the five-second condition was always conducted first, and the two-second condition followed next.

Sentences in think-aloud task.

All of the 20 experimental sentences were extracted from Experiment 2. The sentence design was the same as Study 1, namely four conditions of Attachment (VP and NP) × Length (short and long). Another 20 filler sentences including temporal ambiguity were extracted from Experiments 1 and 3. The number of sentences was decreased to avoid learners' exceeding cognitive loads. In addition, sentences with perception verbs were excluded in Experiment 4, because the verb types induced different results from the previous research in Study 1. Filler sentences consisted of RRC and URC sentences: ambiguous, high and low PP-attachment sentences of four for each structure. The PP-attachment sentences in Experiment 1 were used as the target sentences. All sentences, including experimental items and filler items contained a verification question that concerned the attachment site for ambiguous sentences in order to draw participants' attention to critical regions: For the most experimental sentences, verification sentences

like “The first/second NP *had* PP.” were constructed unless the whole meaning became implausible. Half of the questions had a correct answer “yes” and the other half “no”. For the ambiguous sentences in filler condition, corrective feedback was provided for all answers.

Grammar proficiency test.

Finally, a multiple-choice grammar proficiency test was prepared to measure the participants’ explicit grammar knowledge. This was derived from the Cambridge English online test, which the author revised to paper-based version and added items regarding questions of preposition. The original version included two parts: (a) five items showed a statement and required the readers to choose the best of three options to complete conversation, and (b) the other 20 items showed a sentence or conversation including gaps and required to choose the best of four options to complete it. The original instructions were written in English; these instructions were translated into Japanese to avoid ambiguity. The total score of (a) and (b) estimated English grammar proficiency level, such as Cambridge English: Preliminary (PET), First (FCE), and Advanced (CAE); B1, B2, and C1 level in CEFR, respectively. The additional five items queried the explicit knowledge of prepositional phrases. These items were derived from the Oxford Practice Grammar Intermediate Diagnostic Test; the author added the remaining two choices since the original items had only two options to choose from (see Appendix H).

4.3.2.3 Procedure

In this experiment, three types of task were conducted: (a) processing-unit span test, (b) think-aloud task, and (c) grammar proficiency test. The procedure of the processing-unit span test was identical to Experiment 3. In the think-aloud task, the dual-

task after reading was changed into verification task instead of translation task, based on the results of Experiment 4: The participants were provided one verification sentence after reading each sentence, and they were required to press the “yes” or “no” keys to give their answers. According to the answers, feedback was given to correct the participants’ comprehension except for ambiguous sentences in filler condition; either response was regard as correct in the ambiguous sentence, since both attachment sites were meaningful. The feedback would prevent them from sentence underspecification, (i.e., not attempting to resolve ambiguous region, as in Experiment 3). The answers to the question were recorded. In the grammar test, participants were required to choose only one of the options in each item. Though they performed this test without time limitation, most of them finished within 10 minutes. The whole session lasted for 30 minutes to one and a half hours.

4.3.2.4 Scoring and data analyses

Think-aloud protocols.

Before dividing protocols into six categories (see Table 4.9 again), one rater transcribed verbal protocols from the IC recorder and then parsed them into statements based on clauses (Horiba, 2013). Then two raters assessed 20% of the data (i.e., data of four participants) individually using the criteria of Experiment 4 (inter-rater agreement rate = 85.03%). Every disagreement was resolved through discussion. The additional criteria were as follows: (a) When participants read the whole sentence aloud, the protocol should be categorized as oral reading (Category 5), even if they hesitated while speaking; (b) The protocols of English phrase with Japanese translation were categorized as Semantic processing (Category 1), which was distinguished from the protocols including only English phrase (Syntactic processing, Category 2); and (c) meta comments at word

level were categorized as Semantic processing (Category 1), while Others (Category 6) included meta comments at the syntactic or structural level. For each condition, the number of production times of protocols was calculated in each category.

Processing unit test and grammar proficiency test.

The scoring procedure of processing unit test was the same as the pilot study for Experiment 3, and Experiment 3. One point was given to each correct answer in the grammar proficiency test.

In Experiment 4, three types of analyses were conducted. One of analysis was the same correlation analysis as in Experiment 3 (i.e., Analysis B). Analysis A in Experiment 3 was excluded, because the PP-attachment structure was so simple that reanalyzing sentences was unlikely to occur. Instead of categorizing protocols into four pattern of reanalysis × correctness, Experiment 4 will focus on correct or incorrect of the verification task. In addition, to investigate the learners' individual differences, the cluster analysis was conducted for participants' frequency data of five categories. Finally, a three-way mixed ANOVA (Attachment × Length × Grammar proficiency) was conducted for the correct rates of the verification task.

4.3.3 Results of Experiment 4

Table 4.13 shows the descriptive statistics of the grammar proficiency test. Prior to the main analyses, as preliminary analysis, *t* test was carried out to confirm the validity of group division in the grammar proficiency test. The result revealed that the score of the upper group was significantly higher than the lower group, $t(17) = 5.58$, $p < .001$. According to the Cambridge grammar test score, the higher group includes Preliminary

English Test (PET), First Certificate in English (FCE), and Certificate I Advanced English (CAE) level learners (i.e., B2 and C1 level in CEFR); the lower group includes only PET-level learners (i.e., B1 level in CEFR).

Table 4.13

Descriptive Statistics of Grammar Proficiency Test

Proficiency	*n*	*M*	*SD*	*Max*	*Min*	95% CI
Upper	8	17.00	0.67	20.00	16.00	[15.66, 18.34]
Lower	11	14.36	1.60	15.00	13.00	[13.91, 14.82]
Total	19	15.47	1.74	20.00	13.00	[14.63, 16.31]

Note. Maximum possible score = 25. Cut-off point = 16.

4.3.3.1 Quantitative analyses

Correlation between learners' individual factors and frequent data of protocols.

Materials in Experiment 4 included four conditions, and frequency data of protocols in each condition were calculated in the six categories. In order to decrease the number of variables in the correlation analyses, and remove a problem of the complexity in analyses, the frequent data of protocols were summarized in each category instead of analyzing data in the four conditions (e.g., Length × Attachment) and six categories.

Since the reliability of grammatical proficiency test was relatively low (Cronbach $\alpha = .36$), correlations with grammatical proficiency test and other indices were rarely identified. Compared with the correlation between the reading proficiency and processing unit tests in Experiment 3, the grammatical proficiency had a different tendency. The possible reason was whether tasks include time limitation or not. The processing unit test included the strict time limitation, while the grammatical proficiency

test assessed learners' grammatical knowledge without time limitation. Reading proficiency includes reading fluency such as reading rate and automaticity.

The results showed the same tendency as Experiment 3, which indicated that the learners with large processing units infrequently used syntactic processing in think-aloud protocols (see Table 4.14): Moderate and negative correlations with syntactic processing and five-second unit test (r = -.48, p = .027), two-second unit test (r = -.38, p = .087), and average chunk units (r = -.47, p = .032). The number of uses of semantic processing were also negatively correlated with unit tests: Moderate and negative correlations with semantic processing and five- second unit test (r = -.48, p = .029), two-second unit test (r = -.52, p = .016), and average chunk units (r = -.53, p = .013) were identified. These results were because learners with small processing units frequently phrase sentences into small pieces.

Other individual differences were observed in the strong and positive correlation between grammar proficiency and oral reading (r = .78, p < .001), which were not found in Experiment 3. Learners with high proficiency used the oral reading while they did not report the syntactic or semantic processing specifically. In their oral reading, what information they processed or attempted to process was unclear, as such processes could not be verbalized. Therefore, their sentence processing is automatized.

The moderate and positive correlations between semantic and syntactic processing (r = .58, p = .006) and semantic processing and contextual use (r = .44, p = .046) were also statistically significant, although these factors did not indicate learners' individual differences. The correlation between semantic processing and syntactic processing was high, because some learners (especially those with small processing units) translated each phrase into Japanese as in (27b). The original protocols were shown in Appendix G2. The italic sentences denote the original English phrase in the experimental

sentences; slashes indicate segmentations of protocols by the raters.

(27a) The sailor found the beautiful large new sea with the compass.

(27b) *The sailor found* well, sailor means..what?/ *found* he found / *beautiful large new sea* he found beautiful large new sea / *with the compass* he found (it) with compass. [Sentence 10]

In (27b), learners segmented the sentence into four chunks. Considering Hijikata (2012) wherein the sentence processing unit can be extended to the whole sentence, the chunking was very small. The chunking of (27b) and of other clearly denoted the phrase boundaries, using English phrases from original sentences and translations in L1.

Table 4.14

Correlation Analysis of Protocol Frequency Data and Learners' Individual Factors in Experiment 4

Measure	1	2	3	4	5	6	7	8	9
1. Grammar proficiency	—								
2. 5000 seconds	.25	—							
3. 2000 seconds	.23	.72**	—						
4. chunk unit	.26	.94**	.92**	—					
5. C1: Semantic	-.09	-.48*	-.52*	-.53*	—				
6. C2: Syntactic	.15	-.48*	-.38†	-.47*	.58**	—			
7. C3: Context	-.29	-.11	-.13	-.13	.44*	.30	—		
8. C4: Meta	.15	.28	.37†	.35	-.07	.35	.03	—	
9. C5: Oral reading	.78**	.05	.14	.10	-.30	.25	-.28	.20	—
M	14.00	0.58	0.40	0.49	1.93	0.47	0.05	0.17	0.83
SD	4.94	0.12	0.10	0.10	0.86	0.48	0.06	0.24	0.34

Note. $^{**}p < .01$, $^{*}p < .05$, †p < .10

Verification task.

The results of the verification task were presented in Table 4.15. The whole accuracy rate was 87.72%, which is quite high. In order to analyze differences of comprehension between conditions, the repeated three-factor ANOVA was conducted: 2 (Attachment site: NP vs. VP) × 2 (Length: Short vs. Long) × 2 (Grammar proficiency: Upper vs. Lower). The factors of attachment site and length were repeated measures, and the other grammar proficiency was a between-participants factor. The findings did not show the two-way interaction between Attachment, Length, and Grammar proficiency: $F(1, 17) = 2.92$, $p = .105$, $\eta^2 = .15$. The one-way interactions between Attachment × Length, and Length × Grammar were also insignificant: $F(1, 17) = 0.23$, $p = .640$, $\eta^2 = .01$, $F(1, 17) = 0.00$, $p = .975$, $\eta^2 = .00$. The only interaction between Attachment × Grammar turned out to be marginally significant, $F(1, 17) = 4.27$, $p = .054$, $\eta^2 = .20$. Any main effect was not significant: $F(1, 17) = 0.72$, $p = .407$, $\eta^2 = .04$ for main effect of Attachment; $F(1, 17) = 1.21$, $p = .287$, $\eta^2 = .07$ for Length; and main effect of Grammar, $F(1, 17) = 0.49$, $p = .493$, $\eta^2 = .03$.

Although the *p*-value of interaction between Attachment × Grammar was not smaller than .05, the effect size ($\eta^2 = .20$) was large enough to interpret this interaction. The subsequent test of this interaction indicated that significant simple main effect of Grammar proficiency was found only in VP condition, $F(1, 17) = 5.66$, $p = .029$, $\eta^2 = .25$, and not in NP condition, $F(1, 17) = 1.20$, $p = .289$, $\eta^2 = .07$. The high grammar proficient learners answered the VP-condition questions more correctly than the low proficient learners. The simple main significance of attachment also showed that the learners with larger grammar knowledge correctly answered the VP-condition sentences rather than the NP-condition sentences, $F(1, 17) = 3.67$, $p = .072$, $\eta^2 = .19$. The learners with small grammar knowledge did not show the tendency, $F(1, 17) = 0.88$, $p = .361$, $\eta^2 = .05$. These

findings showed that learners with large grammar knowledge preferred the VP-attachment, or they could modify to VP attachment when they initially interpret sentences as NP attachment.

It is surprising that the main effects of length were missing, because Study 1 reported the length effect even when the distant effect was not reported. The possible reason is that learners regarded three additional words in the long condition as one large chunk. Most of their protocols divided the phrase into two parts, the first two words and the remaining final word, and the final word was integrated into the phrase of objective nouns. In some protocols, learners ignored the additional phrase, because the phrase itself did not have meaningful information and also did not affect decision of attachment sites. The other possibility is related to the methodology limitation: learners could take longer time to read, comprehend, and analyze the sentences in the long condition. The additional long phrase were contrasted not to affect attachment site decisions.

Table 4.15

The Means and Standard Deviation of Verification Task in Experiment 3 (%)

		NS		NL		VS		VL	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Upper	8	76.67	28.50	89.38	11.48	97.50	7.07	92.50	14.88
Lower	11	90.00	11.62	88.64	14.16	80.00	20.37	88.64	10.98
Total	19	83.02	20.54	89.05	12.51	87.62	17.51	91.19	12.24

Note. NS = NP and Short condition, NL = NP and Long condition, VS = VP and Short condition, and VL = VP and Long condition

Think-aloud protocol data.

Think-aloud data was categorized into six groups with the identical procedure to Experiment 3 (see Table 4.8 again). First, as quantitative analyses, 2 (Attachment: NP vs VP) × 2 (Length: Short vs. Long) MANOVA was conducted for the number of production times of each category (see Table 4.16 and Figure 4.3). In addition, the frequency data were recalculated separately according to participants' performance of the verification task (i.e., correct response and incorrect one). As the verification task in Experiment 4 asked for the attachment sites, the correct responses in the verification task indicated the participants comprehended the attachment site correctly.

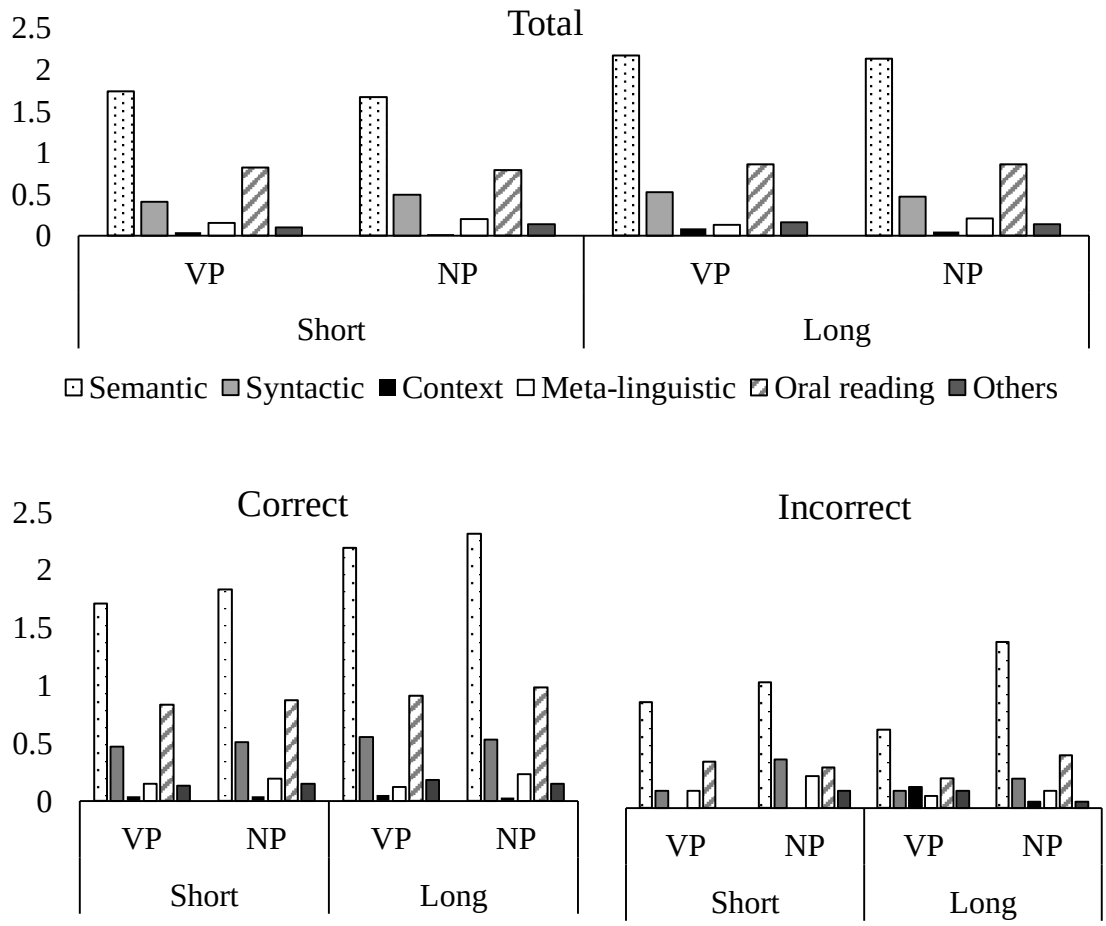

Figure 4.3. Frequency data of each category for total, correct, and incorrect responses.

Table 4.16

Frequency Data of Each Category for Total and Correct and Incorrect Responses

		C1: Semantic		C2: Syntactic		C3: Context		C4: Meta		C5: Oral		C6: Others	
Length		*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
							Total						
Short	VP	1.74	0.84	0.41	0.49	0.04	0.10	0.15	0.32	0.82	0.37	0.10	0.24
	NP	1.67	0.78	0.49	0.49	0.02	0.06	0.20	0.29	0.79	0.33	0.14	0.23
Long	VP	2.17	1.07	0.52	0.63	0.09	0.15	0.13	0.26	0.86	0.34	0.16	0.23
	NP	2.13	1.05	0.47	0.61	0.05	0.09	0.21	0.29	0.86	0.40	0.14	0.26
							Correct Answer						
Short	VP	1.71	0.84	0.47	0.52	0.04	0.10	0.15	0.30	0.83	0.38	0.13	0.25
	NP	1.83	0.68	0.51	0.55	0.04	0.11	0.19	0.31	0.87	0.36	0.15	0.26
Long	VP	2.19	0.99	0.55	0.66	0.05	0.14	0.12	0.26	0.91	0.36	0.18	0.26
	NP	2.31	1.10	0.53	0.64	0.03	0.08	0.23	0.30	0.98	0.42	0.15	0.25
							Incorrect Answer						
Short	VP	0.87	1.31	0.14	0.36			0.14	0.48	0.38	0.50		
	NP	1.03	1.41	0.40	0.85			0.26	0.54	0.33	0.46	0.14	0.36
Long	VP	0.64	1.33	0.14	0.48	0.17	0.37	0.10	0.30	0.24	0.44	0.14	0.48
	NP	1.36	1.74	0.24	0.70	0.05	0.22	0.14	0.36	0.43	0.60	0.05	0.22

Note. *N* = 19. Blank cells mean missing data.

Table 4.17

Results of MANOVA for Total Frequency Data

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
Category 1: Semantic analysis						
Attachment (A)	1	0.07	0.07	0.30	.591	.00
Length (L)	1	4.21	4.21	18.09	< .001	.25
A × L	1	0.01	0.01	0.04	.837	.00
Error (A)	20	4.59	0.23			
Error (L)	20	4.65	0.23			
Error (A × L)	20	3.49	0.18			
Category 2: Syntactic analysis						
Attachment	1	0.00	0.00	0.02	.888	.00
Length	1	0.05	0.05	0.33	.574	.01
A × L	1	0.09	0.09	1.18	.291	.01
Error (A)	20	1.88	0.09			
Error (L)	20	2.91	0.15			
Error (A × L)	20	1.59	0.08			
Category 3: Contextual information						
Attachment	1	0.02	0.02	1.88	.186	.03
Length	1	0.03	0.03	3.60	.072	.05
A × L	1	0.00	0.00	0.16	.693	.00
Error (A)	20	0.18	0.01			
Error (L)	20	0.17	0.01			
Error (A × L)	20	0.24	0.01			
Category 4: Meta-linguistic comments						
Attachment	1	0.08	0.08	3.93	.061	.04
Length	1	0.00	0.00	0.02	.902	.00
A × L	1	0.00	0.00	0.07	.789	.00
Error (A)	20	0.41	0.02			
Error (L)	20	0.61	0.03			
Error (A × L)	20	1.17	0.06			
Category 5: Oral reading						
Attachment	1	0.00	0.00	0.39	.540	.00
Length	1	0.06	0.06	3.69	.069	.05
A × L	1	0.00	0.00	0.16	.691	.00
Error (A)	20	0.29	0.01			
Error (L)	20	0.31	0.02			
Error (A × L)	20	0.53	0.03			

Note. N = 19.

The main findings from Table 4.17 for total data show the following results: There was no significant interaction between attachment and length in any measure. The only main effect of attachment was marginally significant in metalinguistic comments (NP ≧ VP), $F(1, 20) = 3.93$, $p = .061$, $\eta^2 = .04$. The main effects of length were significant in the semantic analysis (Long > Short), $F(1, 20) = 4.21$, $p < .001$, $\eta^2 = .25$, and marginally significant in the contextual information and oral reading (Long ≧ Short), $F(1, 20) = 3.60$, $p = .072$, $\eta^2 = .05$ for the contextual information; and $F(1, 20) = 3.69$, $p = .069$, $\eta^2 = .05$ for oral reading.

The results for correct response also included no significant interaction in any measure. The marginally significant main effects of attachment were found in meta-linguistic comments and oral reading (NP ≧ VP), $F(1, 20) = 3.09$, $p = .094$, $\eta^2 = .05$ for metalinguistic comments; and $F(1, 20) = 3.19$, $p = .089$, $\eta^2 = .04$ for oral reading. In addition, the main effects of length were significant in the semantic analysis and oral reading (Long > Short), $F(1, 20) = 16.37$, $p < .001$, $\eta^2 = .26$ for semantic analysis; and $F(1, 20) = 8.30$, $p = .009$, $\eta^2 = .10$ for oral reading (see Appendix I1, for statistics of the rest insignificant results of correct answers).

The results for incorrect response including the significant main effect of length were found out in the contextual information (Long > Short), $F(1, 20) = 5.87$, $p = .025$, $\eta^2 = .08$. Other main effect or interaction were insignificant (also see Appendix I2, for the overall statistics).

4.3.3.2 Individual differences of parsing strategies

The next analysis will focus on the learners' individual factors. To investigate the difference among individuals, the cluster analysis was conducted first. The variables were the sum of number of participants' produced protocols regardless of attachment or

length conditions. A cluster analysis using the Ward procedure was conducted to classify learners based on their protocol data. The distance among clusters was measured with squared Euclidian distance. As seen in Table 4.18 and Figure 4.4, participants in this study were divided into two clusters (see also Appendix J for dendrogram). The first cluster includes 17 participants, while the second includes only four participants (Participants 2, 3, 7, and 14). Since the number of participants in each cluster was not equal, parametric test was not appropriate for analyses, but the non-parametric test does not allow the two-way mixed design, which is why this study adopt the parametric test.

In order to label each cluster, a 2 (Cluster) × 6 (Categories) two-way mixed ANOVA was conducted for the frequency data. The interaction between cluster and categories, and the main effects of each measure were significant. The subsequent analyses show that the simple main effects of Categories 1, 2, 3, and 6 were significant (see Table 4.19 for descriptive statistics and results of ANOVA). In addition, the simple main effects of Cluster were significant in both Clusters 1 and 2 ($p < .001$ for Cluster 1; and $p < .001$ for Cluster 2). The multiple comparisons showed that the number of protocols of Categories 1 > 5 > 2 > 4, 6, 3; and also Category 2 was produced more than Categories 6 and 3 in Cluster 1. On the other hand, in Cluster 2, the significant differences were found only in Categories 1 vs. 3 and 4. On the basis of these scores, each cluster was labelled. The differences between Clusters 1 and 2 were the overall number of produced protocols; thus Cluster 1 was labeled as Group with Less Frequently Use of Protocols, while Cluster 2 was labeled as Group with More Frequently Use of Protocols. Next, to find out the individual effects on protocols, the grammatical proficiency and processing unit tests were compared between two different clusters.

Table 4.18

Descriptive Statistic of Frequency Data and Results of the Simple Main Effect of Category Between Two Cluster Groups

	C1: Less		C2: Frequent		
Measure	*M*	*SD*	*M*	*SD*	*P*
Category 1	6.56	2.14	12.90	2.64	< .001
Category 2	1.20	0.97	4.80	2.43	< .001
Category 3	0.13	0.17	0.45	0.34	.012
Category 4	0.65	0.99	0.90	0.89	.647
Category 5	3.44	1.24	2.85	1.93	.453
Category 6	0.35	0.48	1.40	1.39	.015

Note. *N* = 17 in Cluster 1, *n* = 4 in Cluster 2. Category 1 = Semantic analysis, Category 2 = Syntactic analysis, Category 3 = Contextual information, Category 4 = Meta-linguistic comments, Category 5 = Oral reading, and Category 6 = Others.

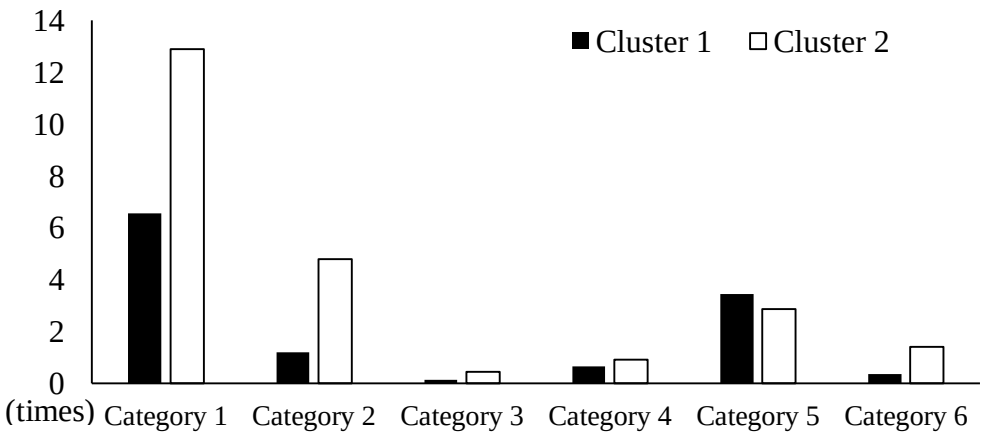

Figure 4.4. Frequency data of each category.

The *t* tests were conducted for processing unit test and grammatical proficiency

test scores to compare two cluster groups, as shown in Table 4.20. In addition, the correct answer rates of experimental sentences were calculated, because the frequency data was averaged out in the four conditions. The results showed that two groups had difference scores of processing unit test, while scores of grammatical proficiency test and correct answer rates of verification questions between two groups showed no differences. According to these results, the differences between two clusters were processing rate or efficiency of sentences. The participants in the second cluster had a significantly smaller processing unit than the first one. It is worthy to note that the two groups were not different in the amount of grammatical knowledge or explicit knowledge processing. Therefore, learners in Cluster 1 (Less Frequently Use of Protocols) had relatively slow processing rates, while ones in Cluster 2 (More Frequently Use of Protocols) process sentences rapidly. Cluster 2 includes data of a few participants; so they use distinctive strategies to complement their slow processing speed, which was quite effective for the learners in Cluster 2 got score in the verification questions as much as Cluster 1. There is another possibility that the learners in Cluster 2 always consider the detailed representation of sentence, and thus the performance of processing unit test was inferior to Cluster 1; though cause and effect remain unclear.

Cluster 2 produced more frequently utterances than Cluster 1 in most of the categories excepting Categories 4 (meta-linguistic information) and 5 (oral reading). The frequent reported group mainly and frequently pointed out the meaning information, while they did not explain the syntactic structures using their explicit grammatical knowledge.

Table 4.19

Descriptive Statistics of Individual Factors in Each Cluster and Results of t *Test*

	Less frequent		More frequent				
Measure	*M*	*SD*	*M*	*SD*	*t*(19)	*p*	*d*
Proficiency test	14.53	4.14	11.75	7.93	1.01	.324	0.56
Unit test (5s)	0.61	0.11	0.44	0.05	2.89	.009	1.61
Unit test (2s)	0.42	0.11	0.32	0.02	1.84	.082	1.02
Unit test (average)	0.51	0.10	0.38	0.03	2.61	.017	1.45
Answer rate	0.87	0.07	0.89	0.08	-0.43	.693	0.28

In Sentence 21, the most amounts of protocols were observed of all sentences, while some of them read in different condition as their allotted booklets were different. The former example was protocol in NP-long condition, and the latter one was in NP-short condition:

(28a) The man knocked down the strong female white tiger with the injury. [Sentence 21]

(28b) *The man knocked down the strong female…white tiger with the injury*. *Knock down* is the only verb in this sentence, so *knock down* is a verb. *The man* is the main subject. *The man* knock down. What (he knock down) is explained next. *The strong female white…* Ah, cool! He knock down the white tiger, even though the tiger is also strong and female. *With the injury* means that he get injured. He won (the tiger) with dying. Hang on. It might mean the injured tiger. Eh? Which does it mean that the man won the injured tiger or the man get injured? *With the injury* places near the tiger, so it might be *tiger* (that get injured).

The protocol included many comments of the participants. These comments might be helpful to deepen his sentence presentation, because he noticed the wrong interpretation immediately after he added the comments. He finally disambiguated attachment sites using the distant information according to the protocols. Another example showed the strategy of underspecification, even in the verification task with corrective feedback.

(28c) *The man knocked down* the man, *knocked* hit. *Knocked down…* Ah, as it said "knock down," he was knocked down? No, it means that he knocked down (something). *The tiger with the journey*, well, *injury*. *With the injury, tiger with the injury*. The (objective) tiger with injury, *with the injury* modifies *man* or *tiger*, and the former is more likely to modify. (The tiger) has already injured. Or if the phrase modifies *the man*, he has already injured. I cannot decide which to attach.

4.3.3.3 Follow-up analyses for filler sentences

To examine the attachment preference of learners, the accuracy rates of verification questions of the filler sentences (i.e., PP-attachment sentences) was analyzed as in Study 1, though the number of filler sentences was limited in this experiment. As for the unambiguous sentences (i.e., high and low attachment disambiguated sentences), the accuracy rates of each four sentences were calculated, and compared using paired *t* test. For the four ambiguous PP-attachment sentences, the same analysis could not be adapted, because these sentences did not have correct answers; the participants always took correct feedbacks (○) when they pressed either yes or no keys. The half of comprehension questions denoted high-attachment interpretation, and the other half denoted low-attachment. The participants' responses were converted into high or low

attachment: When participants responded *no* to the high-attachment verification sentence, their interpretations were regarded as low attachment, and vice versa. The number of times that the participants chose each attachment was counted, and also compared. The results were shown in Table 4.20.

Comparison of unambiguous sentences turned out to be significant, and learners answered better in the low attachment than in the high attachment sentences. The analysis of ambiguous sentences also found out significant differences, but it showed that the high attachment options were preferred, unlike in unambiguous sentences.

Table 4.20

*Accuracy Rates of Unambiguous Filler Sentences and Number of Times the Participants Choose Each Choice (*N = *19)*

	High		Low				Cohen's
Structure	*M*	*SD*	*M*	*SD*	*t*(20)	*p*	*d*
Unambiguous (%)	64.29	23.15	84.52	14.74	3.6	.002	1.04
Ambiguous	2.48	0.75	1.52	0.75	-2.9	.009	1.28

Note. The maximum possible scores of ambiguous sentences were four, which is the sum of high- and low-attachment respondents.

The sections so far have discussed the findings of PP attachment with think-aloud protocol data. The next session will answer to the other research question (RQ4-5) regarding comparison between different post-reading tasks.

4.3.4 Comparison between post-reading tasks (RQ4-5)

To compare the effects of two different types of post reading task on sentence-comprehension performance, data of Experiments 3 and 4 were analyzed again with the different sentences. The main focus of Experiment 4 was not on local ambiguous sentence processing unlike Experiment 3, but a few same sentences as Experiment 3 were extracted from the filler sentences, which did not ordinary intended to be analyzed. Finally, four sentences in both ambiguous and unambiguous sentences (i.e., total of 8 sentences) were set to compare two post-reading tasks. The reanalyzed sentences in RQ4-5 were chosen according to the following criteria: (a) The nouns in sentences were animate to control ambiguities, and (b) RRC/main clause ambiguity occurred. The verification sentences of the RRC and URC structures were constructed without the construction of "*NP had NP*" which were adapted in the four conditions of PP attachment sentences; instead, the verification sentences were constructed as whether the local interpretation or the global interpretation were satisfied. For example, "The woman sent a doll (the local interpretation)" or "The woman was very excited (the global interpretation)" for the RRC sentences "The woman sent a doll was very excited."

4.3.4.1 Data analysis

Data were retrieved from Experiments 3 and 4. These experiments will examine whether the final interpretation was correct or not, using different post-reading task: translation task in Experiment 3, and comprehension task with corrective feedbacks in Experiment 4. The accuracies of final interpretation in two experiments were analyzed; the scores of final interpretation (i.e., L1 translation) in Experiment 3 were rated by the experimenters, while the accuracy of Experiment 4 was regarded as the responses of verification questions. The data were summed in Table 4.22. In order to compare the

performances of experiments and RC structures, the Experiments (3 vs. 4) × Structures (RRC vs. URC) mixed two-factors ANOVAs were conducted for the number of times of renalyses and accuracy rates of final interpretation or comprehension questions.

4.3.4.2 Results

Before the main analyses, the homogeneities between two groups in Experiments 3 and 4 were examined. The identical test which participants in both two groups took (i.e., the processing unit span test) was used as a measurement of homogeneity. To confirm the homogeneity of participants in two groups, a two-way 2 (Experiments: 3 vs. 4) × 3 (Unit processing tests: 2000ms vs. 5000ms vs. average) mixed ANOVA was conducted (see Table 4.21 for descriptive analysis of the processing unit test). It indicated that the main effect of experiments was not significant, $F(1, 26) = 1.09$, $p = .305$, $\eta^2 = .04$. The main effects of tests was significant, while the interaction between experiments and tests were not significant, $F(2, 52) = 77.89$, $p < .001$, $\eta^2 = .75$; $F(2, 52) = 0.55$, $p = .582$, $\eta^2 = .02$. Thus, the homogeneity of two groups was confirmed.

Table 4.21

Summary of Processing Unit Test in Experiments 3 and 4 (%)

		unit of 5000 ms		Unit of 2000ms		Average unit	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experiment 3	7	64.64	14.13	43.33	16.43	53.99	13.52
Experiment 4	21	57.86	12.13	39.84	10.46	48.85	10.49
Total	28	59.55	12.74	40.71	11.98	50.13	11.28

Note. The data in this table were the same as Table 4.11 in Section 4.2.3.1 and Table 4.14 in Section 4.3.3.1.

Table 4.22 shows the descriptive statistics of the four RRC and URC sentences used in this analysis. The main analyses of two-factors mixed ANOVAs were conducted as shown in Table 4.23.

As for reanalysis counting, the interaction between Experiments and structures and the main effect of structures were not significant, $F(1, 26) = 1.66$, $p = .210$, $\eta^2 = .06$; $F(1, 26) = 1.66$, $p = .210$, $\eta^2 = .06$. The main effect of Experiments were significant, $F(1, 26) = 4.28$, $p = .049$, $\eta^2 = .14$. The first two statistical scores were identical each other, because the participants in Experiment 3 did not re-analyze the initial interpretation in the sentences chosen in this analysis. The number of times of reanalyzing were larger in Experiment 4 than Experiment 3.

Table 4.22

The Number of Times of Re-analyses and Accuracy Rates of Comprehension (%)

		Reanalyses				Accuracy rates (%)			
		RRC		URC		RRC		URC	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Exp. 3	7	0.00	0.00	0.00	0.00	47.62	35.91	100.00	0.00
Exp. 4	21	0.14	0.20	0.04	0.09	54.76	35.01	91.67	19.90
Total	28	0.11	0.19	0.03	0.08	52.98	34.71	93.75	17.51

Note. RRC = Reduced relative clause, URC = Unreduced relative clause

The results of the accuracy rates also showed the insignificant interaction between Experiments and structures, $F(1, 26) = 0.80$, $p = .380$, $\eta^2 = .03$, while the main effect of the structures turned out to be significant, $F(1, 26) = 26.53$, $p < .001$, $\eta^2 = .51$. The main effect of Experiments was not significant, $F(1, 26) = 0.01$, $p = .945$, $\eta^2 = .00$.

Table 4.23

Summary Table for Two-factors mixed ANOVA of the Effects of Experiments and Sentence Structures on Accuracy Rates and Reanalysis Counts

Source		*df*	*SS*	*MS*	*F*	*p*	η^2
				Within-participants			
Structures (S)	Reanalysis	1	0.03	0.03	1.66	.210	.06
	Interpretation	1	2.09	2.09	26.53	<.001	.51
S × E	Reanalysis	1	0.03	0.03	1.66	.210	.06
	Interpretation	1	0.06	0.06	0.80	.380	.03
Error (S)	Reanalysis	26	0.47	0.02			
	Interpretation	26	2.05	0.08			
				Between-participants			
Experiments	Reanalysis	1	0.08	0.08	4.28	.049	.14
	Interpretation	1	22.70	22.70	299.97	<.000	.92
Error (E)	Reanalysis	26	0.51	0.02			
	Interpretation	26	1.97	0.08			

The findings indicated that task differences affected the learners' number of times of reanalysis but not comprehension of the sentence itself, although the sentences in this analysis were quite limited in terms of both quantity and quality.

4.3.5 Discussion of Experiment 4

4.3.5.1 Attachment preference (RQ4-1)

In Study 1, the types of main verb affected learners' attachment preference, because the experiment sentences included both action and perception verbs: the prior

studies on PP-attachment sentences almost focused on only action verbs. To resolve this problem, the present study excluded the perception-verb sentences from experiment, and the results showed the different tendencies from Study 1. It also suggested that learners were sensitive to the verb subcategorization information (e.g., Boland & Blodgett, 2006). Note that the verification questions in Experiment 4 were different from the preferences tests in Study 1 in term of with or without correct or incorrect feedbacks to participants, while both tasks require to decide which site the PPs attached.

According to the results of verification questions, the overall participants in Experiment 4 did not show the attachment preferences in off-line processing, but learners with larger grammar knowledge performed better to comprehend VP conditions. The learners with smaller grammar knowledge did not show any differences in attachment preferences. The dominance of VP attachment were shown in L1 sentence processing studies and some of L2 studies, while NP attachment preferences have not been insisted in only L2/EFL studies. Furthermore, learners with smaller knowledge group even performed better than large group in the NP-short condition, although there was no statistical differences in NP conditions. This result suggested that the learners' grammatical proficiency will affect the attachment preferences: learners with good grammar proficiency prefer VP attachment, while learners with poor proficiency did not have biases. Considering these facts, learners do not use the sole strategy or processing model to resolve ambiguity, and the strategy and model was possible to be consistent and inconsistent with L1 processing. Therefore, the EFL sentence processing was not fundamentally different from L1 processing; instead, EFL parsing processing shifts to native-like processing gradually (Jackson, 2008; Juffs, 1998; Lim & Christianson, 2013; Rah & Adone, 2010a).

4.3.5.2 Distance effects (RQ4-2)

The definition of the distance effects was that the tendency of attachment preferences differ in the VP condition which had distances between attachment sites and PP. Experiment 4, however, did not show the distance effect in any measures.

Verification questions after self-paced reading task did not show even the length effect. The additional long phrase in object NP affected on-/off-line processing of EFL learners due to limited time or limited cognitive resource in Experiments 1 and 2. However, the think-aloud method in the present study did not have restricted time, which allowed learners to take a long time to process sentences, and their cognitive resources were not restrained. In addition, the fact that the verification questions did not ask the long additional phrases also affected comprehension. Learners could strategically exclude the long condition from syntactic parsing or sentence comprehension.

4.3.5.3 Differences of protocols according to success or failure of sentence comprehension (RQ4-3)

To compare the tendencies of correct responses with total and incorrect responses, the significant and marginally significant results are summarized in Table 4.24. From these results, even when participants correctly answered the verification question, they had to engage in the metalinguistic analyzing of NP-attachment sentences. This tendency was not due to the advantage of NP-attachment correct answer rates, because the frequency data for total score showed the same results. According to the scoring criteria, category of meta-analyzing denotes that learners mention to structure or interpretation of NP-attachment sentence; thus, they might have processing load of NP-attachment sentences.

Table 4.24

Summary of Frequency Data According to Accuracy of Questions

Measures	Total	Correct response	Incorrect response
Attachment	[Metalinguistic comments] NP ≧ VP	[Metalinguistic comments] NP ≧ VP [Oral reading] NP ≧ VP	
Length	[Semantic information] Long > Short [Context information] Long≧Short [Oral reading] Long≧Short	[Semantic information] Long > Short [Oral reading] Long > Short	 [Context information] Long > Short

Note. “>” indicates significant results, while “≧” reflects marginally significant results.

The possible reason is that PP using *with* is mainly interpreted as an *instrument* in VP-attachment sentences, while PP in NP-attachment sentences has various interpretations such as objects’ *possessions* or *properties*. Since PP in NP-attachment sentence contained more options to choose adequate interpretation than in VP-attachment sentences, learners had to spend considerable effort to process NP-attachment sentences. For example,

(29a) The hunter killed the lion with the big face.

(29b) *The hunter killed the lion with the big face*. The subject is *the hunter*; the verb is *killed*; the object is *the lion*, and *with the big face* modifies *the lion* in this structure, in this sentence structure. The meaning of sentences indicated that the hunter killed the lion with big face.

This result seemed to be inconsistent with Study 1 which concluded that learners preferred the NP attachment to VP attachment. The reason was that distance between attachment site and PP were not related to the interpretation of each condition.

Additionally, the main effect of length in semantic analysis and oral reading mean that learners produce more protocols in the long condition than in the short condition when they answer the comprehension question correctly. These results were not surprising. Because the long condition contains an extra three words, the oral production naturally increased in both semantic analysis and oral reading. The difference between the number of times semantic analysis of the length conditions was conducted was approximately 0.5 times in each attachment condition. In scoring, the protocols of whole sentence in English were counted in the category of oral reading. The results for total frequency data showed no main effect of length in oral reading, while correctly answered frequency data showed significant main effects.

When learners answered incorrectly, protocols about the context information were produced more in the long condition than that in the short condition. The produced raw amount of context information itself was quite low, but learners attempted to comprehend sentences using implicit context information in sentences in the long condition. The extra three words in the long condition might have been clues that learners other information, but this did not affect attachment decision. For example, the following protocol contains the context information (English phrases in the protocols were italicized):

(30a) The woman saw the coming dark difficult future with the cards.

(30b) *The woman saw the coming dark difficult future with the cards*. Phrase

with the cards is part of Structure *saw A with B*, I think. I don't understand this sentence well, but it might say that by using tarot card, types of the tarot card, the person might be foretold the coming dark, dark difficult future.

[after verification task and feedback] Ah, I made a mistake in parsing this sentence. The woman is a fortune-teller rather than one who gets her future told. (VP attachment in long condition of Sentence 30, Participant 9)

In Protocol (30b), the participant inferred context information about *cards*. Although Sentence 30 includes information for fortune-tellers, he might associate the words *future* and *card* with tarot cards and fortune-tellers. In the sentences, learners utilized pragmatic information of *future* and *cards* to comprehend sentences. The short condition also included the same words, but the additional three words worked as clues that helped participants describe the detailed situation of sentence representation.

4.3.5.4 Effects of learners' individual differences (RQ4-4)

The correlation analyses produced the following results in think-aloud protocols: (a) the larger processing unit learners had, the less frequently the learners described semantic and syntactic information of sentences; and (b) the more grammar knowledge learners had, the more frequently they used the oral reading of whole sentence in protocols. The former finding was the same as in Experiment 3, and their processing was frequently interrupted by phrase boundaries. The latter finding implied that proficient learners with broad grammatical knowledge parse syntactic or semantic structures of sentence while they read it aloud. Because oral-reading protocols of whole sentences were usually observed at the beginning of protocols, reading aloud might correspond to initial processing of sentences. In sum, learners with higher grammar proficiency constructed

sentence structures immediately or automatically after or during reading a whole sentence. The process was so automatized that they could not tell even when the time limitation was set.

Cluster analyses showed that a few learners who had small processing units thought differently from the other participants. It should be noted that the learners with small units did not always mean less proficient learners, because their performance of grammar proficiency test without time limitation were not inferior to the other group, so they have grammar knowledge as much as the other participants, while their processing unit span is small. Another probability was that they could not Differences between the groups of protocol features showed that learners with small processing units or poor processing speeds attempted to complement their processing speed using semantic or pragmatic information in order to comprehend sentences. This finding is consistent with the results of Experiment 3: (a) learners with small processing units split one sentence into small phrase groups, which resulted in many protocols regarding syntactic processing; (b) learners with small processing units did not spend time with such initial processing as oral reading; and (c) learners with small processing units frequently used semantic information, compared with learners with large processing units.

4.3.5.5 Effects of task types on sentence processing (RQ4-5)

Participants in two distinct experiments were required to perform think-aloud tasks, and they were also given different tasks after they had completed the protocols; the participants in Experiment 3 had to answer the L1 translation of sentences after completing the protocols, while the participants in Experiment 4 answered two-fold multiple-choice comprehension questions, which also provided feedback according to participants' responses. In other words, participants in Experiment 3 could not always

disambiguate the ambiguous parts when the sentences included ambiguities, while the comprehension tasks in Experiment 4 provided the correct structure to the participants, because it showed correct answers after every question.

The results indicated that the accuracy rates were not affected by task demands, while tasks that needed to disambiguate sentences facilitated the reanalyses of learners. Adding verification questions that needed disambiguation was possible to change learners' sentence processing and focus learners' attention on reanalyzing the anomaly points in sentences. However, learners who read sentences for translation tasks sometimes kept the structures of sentence unresolved to apply the strategy of underspecification. The results did not mean that their sentence comprehension was inferior to that of the learners in Experiment 4, or that they could not notice their misunderstandings, because the accuracy rates were the same as for learners in Experiments 4. Additionally, the learners sometimes mentioned the processing difficulties they experienced, but they mostly just left them unresolved. Therefore, the task constraints that were conducted after reading were not strong enough to promote learners' sentence comprehension but could change learners' sentence processing.

4.3.6 Conclusion of Experiment 4

Experiment 4 aimed to investigate the following five points using the think-aloud method: (a) the attachment preference in the PP attachment sentences, (b) the distance effects of PP attachment sentences, (c) the different performance of protocols according to success or failure of learners' sentence comprehension, (d) the different performance of protocols according to learners' processing units, and I different performance of protocols according to task demands.

First, EFL learners did not have overall bias regarding attachment sites of PPs

when the main verbs of sentence were limited to action verbs. The specific analyses based on their grammar proficiency level showed that the upper group preferred the VP attachment, while the lower group remained unbiased toward attachment sites. Therefore, the preferred attachment site was not unique but varied according to learners' proficiency. According to improvement of the grammar knowledge, learners' attachment site verged to the high-attachment; namely, native-like processing. Second, the distance from attachment sites to PP failed to bring about the length effect or distance effect on learners' protocol performance. This was because learners took a relatively long time to parse one sentence. Since their cognitive resources were not limited, they did not regard the additional three words as a cognitive load to process compared to word-by-word self-paced reading. Third, comparison between success or failure of comprehension and between with or without reanalysis of sentences shows that learners process the NP attachment condition with relatively high processing loads. Forth, regarding the effects of individual differences such as learners' processing units, the processing unit and the number of syntactic or semantic protocols were correlated, which is the same as Experiment 3. Additionally, some learners with a small processing unit attempted to parse sentences using more semantic, pragmatic, and contextual information. They finally achieved the same degree of sentence comprehension and explicit grammar knowledge as the other participants, which indicated that they were inferior to other learners regarding their processing unit or processing speed, but their strategies that used semantic information worked effectively, because their comprehension questions were the same as other learners. Fifth and finally, the post-reading task affected learners' sentence processing but it did not promote learners' sentence comprehension. In particular, the task with questions regarding sentence structures and corrective feedback made learners pay attention to the structures and increased their frequencies of reanalysis.

4.4 Summary of Study 2

The two experiments attempted to apply the think-aloud method to sentence processing study in order to investigate learners' explicit explanations during sentence comprehension using local and global ambiguous sentences. As for the relationships between think-aloud protocols and individual differences, learners with small processing unit reported more protocols, because their phrasal segmentation included many boundaries; they also often added semantic analyzing such as L1 translations after each segmentation.

Study 2 demonstrated again that learners did not bias the attachment sites intentionally when they were assigned translation tasks that did not require them to disambiguate sentences; while learners with high grammar proficiency answered correctly with regard to the VP conditions, when they were assigned the verification task with corrective feedback. Although the participants in Study 1 and 2 showed different preferences for attachment sites, the effects of post-reading tasks were consistent in Studies 1 and 2. Given the comprehension questions related to the ambiguous region, the answer rates of questions were affected, and they revealed the preferences. Additionally, the differences were caused by learners' allotment of cognitive resources. When learners had to disambiguate attachment sites, their reanalysis process was improved, though these verification questions did not improve their comprehension.

Chapter 5
General Discussion

In order to investigate learners' processing of PP attachment sentences, a total of four experiments and pilot studies were conducted. Experiments 1 and 2 in Study 1 focused on learners' behavioral data, while Experiments 3 and 4 investigated learners' though while they read sentences using the think-aloud method. A summarization of the present study was presented before the important points in the present dissertation were discussed.

5.1 Summary of findings in the present dissertation

Study 1 consisted of two experiments that examined learners' on- and off-line processing of PP-attachment sentences to answer the following five research questions:

RQ1-1: Do EFL learners show attachment preference in off-line processing?

RQ1-2: Do EFL learners sensitively respond to material distance in off-line processing?

RQ1-3: Do the question types affect EFL learners' off-line processing?

RQ2-1: Do EFL learners show attachment preference in on- and off-line processing?

RQ2-2: Do EFL learners sensitively respond to material distance in on- and off-line processing?

Experiment 1 examined EFL learners off-line processing of PP-attachment sentences. The experimental sentences were manipulated in length so that the VP conditions took into account the distance or lack thereof between the attachment site

(verb) and PP. Learners performed two-fold multiple choice tests as a preference task and translation task that did not require them to choose ambiguous attachment sites. EFL learners chose NP attachments rather than VP attachments when they had to choose one of two attachment sites. The results of the translation task were not biased. When the verb types were analyzed separately, learners showed null attachment preference with the action verbs, but their decisions in sentences with perception verbs biased towards NP-attachment preference. The results seem to contradict the minimal attachment of L1 studies, and the possibility arose that learners disambiguated attachment sites not by syntactic structure itself but by using distance from PP (RQ1-1). The distance effects were not found in either task, while the length effect was observed only in the translation task. In detailed analyses with different verb types, three-word phrases did not affect learners' decision of attachment sites in the tasks without time limitation (RQ1-2). The differences of preference task and translation task must be located in whether the task allows ambiguous attachment sites to remain ambiguous or not (RQ1-3). In addition, the translation task might sometimes draw learners' focus on how they convert English sentences into Japanese instead of what the sentences means or what structures the sentences have. Experiment 2 also investigated learners' on-line processing in PP-attachment sentences using word-by-word self-paced reading task. To modify the revised preference tasks, a 7-point scale was used instead of multiple-choice questions to examine the learners' detailed preferences and eliminate guessing by learners like in Experiment 1. Reading time data of the self-paced reading task did not show attachment preferences, while the off-line preference task with confidence rates showed the NP-attachment preference (RQ2-1). The reason for no effect in on-line task was that the comprehension questions after reading requires one to choose the attachment sites, rather than learners have to process them on-line. Furthermore, the self-paced reading task did not show the

distant effects. However, the preference task with confidence rates showed the distance effects (RQ2-2), though the reading patterns were not identical to those of L1 studies (Thornton et al., 2000).

From the results of Study 1, the new questions as to how learners processed English original sentences were aroused. Study 2 were conducted to resolve the following questions:

RQ 3-1: Do EFL learners with different processing unit use different strategies in local ambiguous sentence resolution?

RQ3-2: How does the think-aloud protocol differ according to the success and failure of sentence comprehension?

RQ 3-3: Is the think-aloud method effective for sentence processing studies?

RQ4-1: Do EFL learners show attachment preference in the think-aloud method?

RQ4-2: Do EFL learners sensitively respond to material distance in the think-aloud method?

RQ4-3: How does the think-aloud protocol differ according to the success and failure of sentence comprehension?

RQ4-4: Do EFL learners with different processing units use different strategies in syntactically ambiguous sentence resolution?

RQ4-5: How does the think-aloud protocol differ according to the task types?

Study 2 investigated learners' individual processing strategy specifically using the think-aloud method. The method was rarely utilized to investigate sentence processing in the previous studies, but this study aimed to observe learners' explicitly explained strategies.

Experiment 3 focused on the learners' recovery from the strict garden-path

sentences (RRC condition) and unambiguous sentences (URC condition) using the think-aloud method. The learners with small processing unit frequently reported syntactic information (RQ3-1), while learners with large processing unit did not frequently report protocols, and they rarely spoke out the reanalysis of sentences. These performances will relate to learners' processing speed and automatized syntactic processing as well as their processing units. When the protocols were divided into four types according to success or failure of sentence comprehension and with or without reanalysis of sentence, the number of each protocols were obviously biased. The reanalysis were rarely observed even in RRC sentences, which need to modify the initial processing (RQ3-2). The case studies confirmed that learners could over-rely on the semantic, pragmatic and lexical cues, which led learners to the most plausible or biased interpretations regardless of syntactic structures. Regarding think-aloud methods for sentence processing studies, some deficits which were already pointed out by previous studies (Olson et al., 1984) and others were disclosed. On the other hand, the effectiveness of the methods was to know points of processing difficulties (RQ3-3).

Experiment 4 aimed to figure out the strategies of the structural ambiguous sentences of PP-attachment to resolve the remaining questions from Study 1. The same materials of PP-attachment from Study 1 were reused to investigate learners' strategies regarding attachment decision or distances. Participants were asked to describe what they thought during sentence processing, and they were given comprehension questions regarding critical sentence structures and corrective feedback. To compare the results with other previous studies, the types of main verbs in sentences were controlled. The attachment preference in Experiment 4 was slightly different from Study 1. When sentences included only action verbs, learners still did not show the attachment preferences in the verification question task (RQ 4-1). The tendencies also differed

according to learners' grammar proficiencies: Learners with high grammar proficiency showed bias of high-attachment, which is a native-like processing pattern. Therefore, the attachment preference was not only one, but it changed into a native-like processing according to their proficiency. Regarding distance, learners were not affected by the additional three words intervening the main verb and the PP (RQ 4-2). RQ 4-3 investigated the differences between success or failure of sentence comprehension. It showed that learners reported the protocols in the NP conditions even when they successfully comprehended sentences. When they failed to comprehend sentences, they attempted to comprehend sentences using contextual or pragmatic information. RQ 4-4 confirmed that learners with large processing units infrequently reported syntactic and semantic analysis. Other specific strategy of excessive use of semantic and pragmatic information was reported. The group had relatively small processing units, thus their strategical use of semantic, pragmatic, or context information compensated for comprehension. Finally, RQ4-5 compared the effects of tasks in Experiments 3 and 4; the findings showed that learners' sentence processing was highly affected by the task given after reading. Compared with the translation task, the verification task with corrective feedback promoted reanalyzing processes, while it could not enhance sentence comprehension.

Based on these findings from the four experiments, the subsequent sections will discuss the following four points: (a) EFL learners' general use of semantic and syntactic cues, (b) effects of learners' individual factors on sentence comprehension or strategy use, (c) effects of tasks on sentence comprehension and processing, and (d) effectiveness of the think-aloud method on sentence processing. Additionally, the following relevant models will be discussed: (e) fundamental differences of sentence processing between L1

and EFL readers from shallow structure hypothesis and transitional positions are discussed in Section 5.2; (f) the models of attachment preferences (i.e., minimal attachment principle and recency) are also detailed in Section 5.2; and (g) serial or parallel processing.

5.2 Use of Semantic and Syntactic Cues

Effects of semantic cues: preferences of attachment sites.

The overall performances (i.e., accuracy rates) of each task in Experiment 1, 2 and 4 were relatively high. This meant that learners generally comprehended the experimental sentences, and they were able to disambiguate sentences using semantic information appropriately in each condition.

The present study gained the inconsistent data between the experiments on attachment preferences: NP-attachment preference with the perception verbs or null bias with the action verbs in Study 1, and null bias in Experiment 4. The former bias indicated that learners comprehended better with close attachment phrases than distant ones. The results were consistent with recency effects, rather than minimal attachment principles. Learners chose NP attachment sites more accurately than VP attachment sites when they had to decide the attachment site of PP (Experiments 1 and 2). In addition, other sentence structure such as relative clause attachment, indicated the recency principles. Experiment 4 in Study 2 showed slightly different results due to verb types. However, their decisions, especially the decisions of learners with minimal grammatical knowledge were not always biased towards VP attachment.

Considering the translation answers and think-aloud protocols in Experiments 1 and 4, the importance of verb was prominent. The error analysis of translation tasks showed that a sentence including an unknown verb for readers induced serious

comprehension problems. Interestingly, there were protocols wherein the attachment site of NP-attachment condition was determined based on the recency of PP in Experiment 4, though this was quite rare (e.g., Protocol example 28b in Experiment 4; *With the injury* is placed near the tiger, so it might be the *tiger* [that get injured]). On the other hand, no one reported the dominance or preference of VP-attachment sentences even when the verb or noun phrase of attachment sites included unknown words; namely, when the sentence became completely ambiguous for the participant. In addition, there was some biases regarding usages of *with* for Japanese EFL learners.

According to Experiment 4, some think-aloud protocols showed that the participants translated or converted English phrases into Japanese before they fully parsed the sentence structures. The serial processing models indicated that readers initially parse syntactic processing. Almost all protocols began with the oral reading of the whole sentence, and learners are likely to parse syntactic information and cues when reading the sentence aloud. However, in the case where learners process and translate every incoming phrase into Japanese, the sentence representations should also be constructed in Japanese. Since there is no structure identical to PP attachment in Japanese, such representation were not accorded with the minimal attachment principle.

The grammatically proficient learners of English in Experiment 4 answered more correctly in the VP-attachment sentences than in the NP-attachment ones, while less proficient learners did not show any preference. The differences between the two groups were due to the learners' amounts of knowledge. The sentences in Experiment 4 only included the action verbs that originally biased toward VP-attachment sentences (e.g., Kidd & Bavin, 2005). The proficient learners with enough grammar knowledge including verbs and subcategorization information of verbs could easily comprehend verbs and structures of "verb + *with*", and their comprehension was enhanced in VP-attachment

sentences; while less-proficient learners did not benefit from action verbs, because of lack of grammatical knowledge. As a result, their comprehension did not show any preference. Theoretically, the results in this study do not support the minimal attachment principle proposed by Fraizer and colleagues (Fraizer & Fodor, 1978; Fraizer & Rayner, 1982). In sum, learners were highly affected by lexical cues of main verbs and the preposition *with* when they processed the PP-attachment sentences. Learners with high grammatical proficiency could resolve the syntactically ambiguous VP-attachment sentences using their knowledge, rather than by adapting the minimal attachment principle. At the same time, learners' attachment resolutions were structurally affected by the recency principles instead of the minimal attachment principle.

Effects of syntactic cues: distance and length.

As for the effects of distance and length on PP-attachment sentence comprehension, the distance effect was identified only in preference task with confidence rates in Experiment 2. Although the interaction between attachment sites and length were not exactly the same as in an earlier study (Thornton et al., 2000), the learners in the study showed the distance effect in off-line tasks, and not in on-line tasks. The results support the transitional position indicating that the sentence processing skills of learners were inferior to L1 readers due to their syntactic processing speed, and that there were no fundamental differences between L1 and EFL/L2 readers (Lim & Christianson, 2013). Other indices did not show the distance effects, though some of them revealed length effects. The distance effects appeared only when the length effects appeared differently, according to the attachment sites; the distance between verb and attachment sites was manipulated in the length conditions, but the distance between object NP and attachment sites was always the same. The length effects showed that learners regarded three-word

phrases as obstruction when processing sentences. However, the long phrases did not always hinder sentence comprehension. The self-paced reading times showed that learners could read the disambiguated regions more quickly in the long than in the short condition.

5.3 Effects of Learners' Individual Factors on Syntactic Parsing Strategies

Study 2 includes two experiments with the think-aloud method and one pilot study. It aimed to investigate the effects of learners' individual differences on sentence comprehension. The two experiments regarded the grammar proficiency and processing unit span as learners' individual differences. The first test without time pressure targeted the learners' explicit grammatical knowledge, and the latter test with strict time pressure targeted their implicit knowledge (Ellis, 2005). This section discusses the effects of individual factors on sentence processing.

Effects of processing unit on comprehension and strategies.

The learners' sentence processing unit affected their sentence comprehension. In particular, it affected the number of protocols as shown in Study 2. Learners with large processing units reported less often than learners with small units. The results indicated that learners with large units could integrate information from some phrases into one meaningful chunk, while learners with small units regarded each phrase as a chunk. Hijikata (2012) reported that proficient learners sometimes perceived more than one sentence as a chunk, but with syntactically ambiguous sentences, few learners reported such large processing units in think-aloud protocols.

Other individual differences also affected sentence comprehension and think-aloud protocols. Learners with small processing units and with the same grammar

proficiency level as the other learners showed similar specific strategies during sentence processing. They reported a lot of the semantic and pragmatic information to comprehend sentences in the think-aloud protocols. Their detailed comments appeared in almost every sentence, rather than a specific sentence functioning as a cue that activated protocols. The learners in this group might consider referential situations, and activate contextual information even without detailed information regarding the original sentences. Considering that the learners had as much explicit knowledge as the other participants did, the differences of processes between them must have been caused by the processing speed of chunks. Their slow retrieval of syntactic information might have impeded sentence comprehension (Lim & Christianson, 2013). To compensate for the speed, they used contextual information even though the original sentences did not include such information. The learners in this group intentionally attempted to activate contextual or pragmatic information. Sentence processing models assume that readers activate and deactivate sentence interpretations according to input information that sentences convey (Fraizer & Rayner, 1982; MacDonald, 1994; Snedeker, & Trueswell, 2004; van Gompel et al., 2001). However, the findings of this group suggested that readers could activate certain information, not due to input information such as sentence structures, but due to their selective processing.

Effects of grammar proficiency on sentence comprehension.

The grammar proficiency test in Experiment 4 was conducted without a time restriction; learners could take enough time to answer each item; thus, the test targeted the learners' explicit knowledge (e.g., Ellis, 2005) in contrast to the processing unit test. Learners' explicit grammar knowledge affected the accuracy rates of PP-attachment decisions in off-line data, according to the results of Experiment 4. In particular, learners

with high grammatical proficiency processed the lexical information of verb more sensitively than learners with low grammatical proficiency.

Learners' grammatical proficiencies were correlated with the number of times of oral reading in think-aloud protocols, which was similar to the results of the processing unit. However, explicit grammar knowledge did not affect the frequencies of syntactic or semantic reports in think-aloud data, while the scores related to processing units were negatively correlated with it; explicit grammar knowledge affected the preference decision after reading the required specific lexical knowledge, but it did not affect their reports during reading. Therefore, explicit grammatical knowledge affected sentence performance after readers completed the whole parsing.

5.4 Task Demands and Differences

Studies 1 and 2 utilized two types of tasks: One of them required sentences to be disambiguated or the attachment site to be decided (e.g., multiple-choice preference task), and the other type did not require any of this (e.g., translation task). When learners were provided the former type of post-reading task, such as a preference task in Study 1, or comprehension questions regarding critical structures in Experiment 4, learners were biased toward one of the attachment sites. However, learners remained ambiguous (i.e., they were not biased) when they were given the latter type of task, such as a translation task. This strategic process is called underspecification in Swets et al. (2008). They used fully ambiguous sentences that readers could not disambiguate due to lack of information. Although the sentences in the current study could be disambiguated using semantic cues in each condition, the same strategy occurred in the forced-disambiguate sentences, including RC- and PP-attachment structures, in Studies 1 and 2. The results indicated that learners did not show any biases towards attachment sites strategically even when they

were provided enough information to do so.

In particular, tasks that required disambiguation of sentences promoted learners' reanalysis of sentence structures. Tasks providing a choice in attachment sites could draw readers' attention to sentence structures. However, the task instruction indicating that participants have to disambiguate sentences did not promote sentence comprehension per se.

Fundamental differences between L1 and EFL comprehenders (shallow structure hypothesis) vs. transitional positions.

The data of this study supports transitional positions based on the following factors: (a) attachment preferences/biases, and (b) the interaction between PP attachment sites and length. As for attachment preferences, some learners, especially those possessing greater grammar proficiency, showed the same tendencies as revealed in previous L1 studies, namely VP-attachment preferences, for instance, Experiment 4 in which experimental sentences were controlled. Therefore, the attachment preferences or biases are not always identical among EFL learners with different proficiency levels. The present study supported the transitional positions of learners' sentence processing; the proficiency level affected the sensitivities of verb subcategorization information based on the data from Experiment 4. It should be noted that according to this study, learners with greater grammatical proficiency adapted L1-like processing (i.e., performed well in VP-attachment sentences), not because they adapted the minimal attachment principle, but because they accessed the subcategorization information of verbs to resolve attachment sites.

In the interaction between attachment sites and length in on-line data (Experiment 2), learners displayed a different reading pattern from the L1 study

(Thornton et al., 2000). However, the distance effect appeared in off-line data. Taking the features of the experimental sentences into account, the distance effect might occur with delay in learners' sentence processing. The proficiency levels of the participants in this study were lower compared to other L2/ESL sentence processing studies. It is natural to discern the marginally different results from the previous studies.

Serial processing (garden-path model) vs. parallel processing (constraint-based model).

With regard to the problems in serial or parallel processing of multiple interpretations, the findings from think-aloud protocols in Study 2 suggest serial, rather than parallel processing. The constraint-based model assumes parallel processing, in that readers prepare multiple interpretations of sentences, and compare these interpretations parallelly using various pieces of information. Think-aloud protocols, however, did not give evidence of parallel processing. According to the parallel processing models such as the constraint-based model, readers selectively activate or de-activate some interpretations using various pieces of information. Applying the model to think-aloud protocols, it is assumed that learners mention one of the most plausible or suitable options from their interpretations until they encounter the processing difficulties: When they have to change their interpretation, they report the second alternative rapidly. However, some learners wondered as to how they reanalyzed or became silent after they noticed the incorrect interpretation. Their hesitation and silence clearly mean that they re-construct the alternative interpretation when they face difficulties, and that they did not prepare additional interpretations in advance. Additionally, the think-aloud protocols supported serial processing models such as garden-path model and race model. The models assumed that the readers initially engaged in syntactic parsing. Some learners pointed out the part of speech of each phrase in the protocols (e.g., "this phrase is the subject of this

sentence…”). Learners parsed syntactic information before they reached the ambiguous regions in the think-aloud protocol, although these comments were relatively surficial. These findings suggest that learners employ serial sentence processing in simple PP-attachment sentence.

Nevertheless, serial processing of EFL learners cannot be insisted upon based on the evidences from the protocols due to the following reasons: (a) The think-aloud method itself might contain some problems such as the fact that participants have to speak out linearly and hence, could not report multiple interpretations at a time; (b) some learners who parse sentences rapidly are possibly suited for parallel, rather than serial processing; and (c) even learners with less automated syntactic processing assume various initial interpretations which might be de-activated soon and did not occur in protocols.

In sum, the results of this study are confounding from a theoretical perspective. Both the garden-path model and the shallow structure hypothesis were partially supported and partially discarded, based on the findings of the research. As for the garden-path model, the assumption of serial processing was supported: Learners likely process one interpretation at a time, and they re-analyze and re-construct other interpretations after they encounter processing difficulties in the think-aloud protocol. Additionally, the initial processing might be syntactic, although semantic processing possibly occurred at the same time. On the other hand, the minimal attachment principle in the garden-path model did not suit the data: Learners’ parser did not prefer simple structures of sentence representation, but it showed recency effect to save their cognitive resources.

Regarding the shallow structure hypothesis, the assumption of shallow parsing of sentences was demonstrated in think-aloud protocols. However, the fundamental differences between learners and L1 readers could not be confirmed, because the off-line

processing of learners in Experiment 1 did not show the same reading pattern.

5.5 Effectiveness of the Think-Aloud Method in Sentence Processing Studies

Study 2 adapted the think-aloud method to investigate learners' sentence comprehension, which was rarely done in the previous research. Basically, think-aloud protocols could only reflect learners' explicit representation of sentences, because the other automated and implicit knowledge or processing could not be verbalized (Ellis, 2005). This study discovered the advantages and disadvantages of using think-aloud protocols. The advantages and disadvantages of the think-aloud method are discussed as follows:

Regarding the advantages of the method, it explicitly and qualitatively described the causes and reasons of the learners' failure at syntactic analysis; though reading-time or eye-tracking studies require researchers to inference the learners' difficulties and their detailed recovery. Using protocols allows to know where learners had processing difficulties and what kinds of difficulties (e.g., Protocol 24b showed the participant constructed the incorrect sentence representation because she activated the redundant pragmatic information). In addition, the protocols showed the changes in the interpretation, the learners' use of their grammatical knowledge, and how to recover from an incorrect interpretation. For example, though PP-attachment sentence has a simple structure for EFL learners to parse, some protocols were initially interpreted incorrect, and were required reanalyzing. Some studies of PP-attachment sentences have prepared various preposing contexts to investigate whether readers can use the information to decide ambiguous attachment sites (Ying, 1996; Yokokawa, 1994). The learners in the present study were affected by pragmatic information or context-like information that learners inferenced, even when they were not given any context information.

As for the disadvantages of the think-aloud method, there are three difficulties in Study 2: (a) learners with automatic syntactic processing, (b) automatic initial processing, (c) interference of learners' L1 (i.e., Japanese), and (d) the impacts of the think-aloud method on sentence comprehension.

First, the individual differences in terms of the automatization of syntactic parsing might have affected learners' performance. To analyze certain types of performances such as reading aloud, researchers had to infer how participants processed sentences. The shallow processing such as phrase segmentation or analyzing part of speech in each phrase (Clahsen & Felser, 2006a) did not appear in high-proficient learners, but the results did not suggest that they did not engage in syntactic processing. Second, similarly to the first point, the initial processing that could not be explained explicitly was difficult to observe in the think-aloud method. Learners usually read sentences aloud at the beginning of sentences, and their initial processing should be done at the time of reading aloud. The remaining protocols might be concerned with later processing to resolve ambiguous sentences. Third, the L1 language in think-aloud protocols might be influenced by observing learners' EFL sentence processing. When Japanese EFL learners attempted to mention word meanings, or semantic information, they had to explain English sentences in L1 Japanese. Although learners' parsers might activate some representation or information from sentence inputs, the representation could not be reflected in the protocols. Forth and finally, the think-aloud method itself affects readers' comprehension processing. As the self-paced reading task in word-by-word style possibly disturbs readers' chunk making and text comprehension, the think-aloud method can alter the normal comprehension process to be more introspective. In other words, the method itself enhances readers' self-monitoring.

Chapter 6
Conclusion

6.1 Overview of This Study

The present study investigated how Japanese EFL learners use syntactic, semantic, and pragmatic cues to resolve and interpret syntactically ambiguous sentences, especially when they have to recover from their initial incorrect interpretation, using PP-attachment sentences, and the effects of learners' individual differences on such sentence processing and strategy use during sentence processing. The main findings of the present study can be summarized as follows:

Study 1 investigated the on- and off-line PP-attachment sentence processing of EFL learners. Regardless of on- or off-line processing, learners' attachment preference of PP was biased according to the task requirement. The task that always requires the disambiguation of attachment sites, makes learners choose the closest options between two possible attachment sites. However, when tasks that clearly did not need the disambiguation of sentences were given, learners would intentionally leave attachment sites unresolved in order to save their cognitive resources. Because their attachment preferences were highly affected by subcategorization information of the main verbs in sentences, the minimal attachment principle is not likely to be adaptable for EFL Japanese students. Instead, learners might adapt the recency principles when the subcategorization information of the verb is not available. The distance effect from PP to attachment sites appeared only in the off-line confidence rating, and learners did not notice or respond sensitively to the distance in on-line processing. Their sentence comprehension suffered from the length itself.

Study 2 investigated the effects of learners' individual differences and post-reading tasks using the think-aloud method on sentence processing and strategy use. The findings from think-aloud protocols in Study 2 showed that learners with small processing units frequently reported syntactic and semantic processing in think-aloud protocols. The post-reading tasks with or without questions regarding the critical structures affected the learners' sentence reanalyzing processes. Some learners tended to translate each phrase into Japanese before they attempted to construct the syntactic structure, and they finally summed it up in Japanese to comprehend the sentences. Additionally, Experiment 4 with PP-attachment sentences demonstrated that the tendencies of attachment differed according to the learners' English skills. The effects of the translation task and comprehension task appeared in the learners' reanalyzing processes.

6.2 Limitation of the Present Study and Suggestions for Further Research

The findings of the study have several limitations from the perspective of experimental sentences and participants. This section discussed the limitations of the present research and suggestions for further research.

With regard to the experimental sentences, (a) the type of prepositions, (b) the length and distance in experimental sentences, (c) experimental design of sentences, (d) verb types of experimental sentences, and (e) the role assignment of prepositional phrases are discussed. First, the studies only utilized the prepositional phrase *with* to examine the learners' processing, because many previous studies focused on processing of *with*-PP attachment sentences which can avoid the influence of phrasal verb. Nevertheless, learners have difficulties in using or comprehending many other prepositions and other structures. Second, the length manipulation that adds three modification words just before PP as adopted in the previous studies (Thornton et al., 2000) made the experimental

sentences unnatural, because long phrases are sometimes used as a postmodification instead of premodification (Hemforth et al., 2013). Additionally, there was a problem of length manipulation. In this study, the length manipulation did not include phrase boundaries such as RC attachment with prepositions. In the RC attachment sentences, some studies included [NP1] of [NP2] [RC] structures. For example, Jun (2010) used the following sentences including syntactic boundaries (e.g., *John was excited to meet the niece of the actor who was recently starring in a very successful play*). However, learners could parse the additional three words as one chunk in the current study, and the long part sometimes did not work effectively to intervene among the constituents in sentences, particularly in the task without time pressure. Considering that the sentence that included distant attachment site evoked many think-aloud protocols in Experiment 3 (*The newspaper reported the death of the film star sells quickly* in Section 4.2.3.2), more variety in length should be examined.

Third, the experimental design that adds the long phrase to object NP is premised on readers' VP-attachment preference in PP sentences. As the study indicates the dominance of NP attachment sentences in EFL sentence processing, other designs which focus on NP attachment preference will determine the distance effect in a more detailed manner. For example, Nakanishi (2012) manipulated the length and distance by inserting adverbs immediately after the main verb. The study did not find length effects, but the manipulation might be useful to investigate the distance effect regardless of readers' attachment preference. Fourth, the experimental sentences did not control the subcategorization information of the main verb in sentences; it included both action and perception verbs. The detailed distinctions of verbs are needed. In order to avoid the effects of verb, other types of materials are also required to be chosen. Kim et al. (2013) chose the PP attachment sentences wherein the possible attachment sites were the noun

phrases, instead of verb and object NP (e.g., *The FBI agent noticed* <u>*the mirror*</u> *on* <u>*the wall*</u> *with crack*). Fifth and finally, the different role assignment of prepositional phrases should be investigated in further studies. Learners might have semantic bias in interpreting PP-attachment sentences as well as structural bias according to the complement survey, although it focused on learners' sentence production. Learners frequently associated the role of accompanying person with the preposition *with*. This tendency was due to their English education. In order to examine the learners' sentence processing, a textbook corpus will be useful, which is an important resource of input information for learners.

As for the participants, the experiments did not compare sentence comprehension of EFL learners with L1 speakers of English, though the comparison was normally conducted in previous studies. Since the off-line preference task was not conducted in the L1 sentence processing study (Thornton et al., 2000), L1 readers' results should be compared with those of the current study. Further studies need to focus on length effects of the phrase in more varied and natural sentences, and compare the performance of L1 readers with that of EFL learners.

6.3 Pedagogical Implications From the Present Study

The findings in the research suggest several pedagogical and methodological implications: From the pedagogical perspective, (a) use of translation task as an integration performance, (b) use of task after reading, (c) the suggestion for sentence-comprehension instructions and (d) the implication in sentence production were discussed; From the methodological perspective, (e) the effects of a long additional phrase and (f) adaption of the think-aloud method to sentence processing study are discussed.

Use of translation task as an integration performance.

The use of translation task will be discussed on the basis of results in Experiments 1 and 3. In English education in Japan, the grammar-translation method was used for a long time (e.g., Mochizuki, 2010). The method gradually has given way to the communicative one. However, many classes and admission tests still include translation tasks. The potential problems of the translation task are that learners translated English sentences into meaningful L1, but inconsistent with the originals; and that they translated English into the consistent structure in their L1, but the L1 products were meaningless. When translation task is assigned, some learners attempted to translate all words or phrases into L1 Japanese without thought of global structures of sentence to complete the task. Such learners tended to accumulate the Japanese phrases to comprehend final interpretation, putting aside syntactic structures or plausibility of the original English sentences. As a results of such wrong efforts, learners' translation became meaningless and inconsistent with the original sentences. When researchers and teachers intend to observe learners' integrated representations, these meaningless translations were quite unsuitable. To avoid failure, teachers should emphasize the importance of reanalysis in syntactic structure when learners notice the anomalies in sentence meaning or structures. Additionally, the scoring of translation tasks should focus on whether learners can construct the correct syntactic structures, and not whether they can translate English into natural Japanese sentences. Disclosing the scoring criteria of translation task to learners might also be useful to draw their attentions to global structures of sentences. The combination of the think-aloud method and translation tasks during sentence processing allows teachers to know how and where learners have difficulty in comprehending sentences. While this study used the think-aloud protocols as quantitative and qualitative data, teachers can utilize this method more easily and quickly to identify learners'

problems in sentence comprehension.

Effects of post reading task.

The post-reading tasks clearly affected learners' (un)ambiguous sentence processing throughout the whole study: The task that required readers to disambiguate sentences biased learners' decisions toward the heuristic processing and also facilitated reanalysis of sentences. In Japanese classes of English, globally or locally ambiguous sentences rarely appear in their textbooks or materials, but the findings of post-sentence-reading tasks offer general sentence comprehension task. Tasks that repeatedly make learners notice sentence structures in question have an effect on the allotment of cognitive resources. Without these questions, the strategy of underspecification might be adapted, and they may not attempt to disambiguate sentences. However, it should be noted that assigning such kinds of post-reading tasks did not enhance learners' sentence comprehension as discussed in RQ 4-5. Since learners have to save their limited cognitive resources to process subsequent information, it is impossible to parse the sentences without processing difficulties (Koda, 2005). In order to increase learners' syntactic processing skills or promote sentence comprehension, other instructions are needed. For example, participants with relatively slow processing speed in Experiment 4 intentionally activated contextual or pragmatic information from the experimental sentences, in order to compensate for the lack of information. The interpretations were sometimes unrestricted and inconsistent with the original sentence. Since such incongruent interpretations are not allowed in classrooms, teachers can provide referential information to help students' construct the sentence representation. A simple picture or a sentence can be an effective cue to disambiguate sentences and construct representations (e.g., Kidd & Bavin, 2005 for visual cues; O'Brien et al., 2014; Ying, 1996 for contextual cues).

The implication for grammatical instructions including PP attachment sentences.

The results from Experiments 1, 2, and 4 suggested that EFL learners' ambiguity resolutions were highly affected by the subcategorization information of verbs, and learners' usage of the preposition *with* were limited to a specific role assignment (i.e., to denote accompanying person such as "do something with one's family or friends"). In an English classroom and textbook, the frequently used verbs are fixed. After students have acquired the most frequent usage of the verb or preposition, teachers should present other usages as well as the original one. In particular, verbs have some complicated subcategorization information such as transitivity. Therefore, teachers should provide various input information including the target verb, in order to encourage students to use the appropriate lexical or pragmatic cues in a given context.

The implication for sentence production of PP attachment sentences.

The implication for sentence production is discussed. Although this study did not use production tasks such as sentence completion tasks in the main experiments, the findings showed learners' heuristic processing: learners preferred NP attachment when the sentence included perception verbs, and the long phrase that intervened between attachment sites and PPs disturbed the whole sentence comprehension. Based on the results, how to construct comprehensible and disambiguous sentences using *with* will be suggested based on: (a) the distance from attachment sites to PP, (b) additional phrases, and (c) the semantic relatedness of constituents. The primary point is that writers should put the attachment sites and phrases close to one another. Readers, especially EFL learners, chose closer sites to attach phrases in order to save their cognitive resources. The second point is that redundant information should be omitted from the sentences. Even phrases that did not include complex syntactic boundaries, such as relative clauses and preposition

phrases, constituents that intervene between attachment sites and attachment phrases basically interfered with sentence comprehension. The exceptions were in the case that the sentence includes perception verbs. According to Section 3.1.3.4, learners performed well in the long condition with perception verbs. Therefore, writers can intentionally add phrases after the verb that will make it clear that the subsequent PPs do not modify the verb; and that the attachment site is determined uniquely. Finally, the third point in sentence production is the semantic relatedness or collocation between attachment sites and phrases. The PP attachment sentences using *with* as in the experimental sentences in this study are syntactically ambiguous and confusing. Basically, readers have only semantic cues to disambiguate PP-attachment sentences, when the sentence does not provide referential information such as context or prosody. Therefore, writers should utilize the semantic information of attachment sites and phrases effectively. In the VP-attachment sentences, for example, the strong collocation sets of verbs and with-PP should be used. In the opposite way, the semantic relatedness between the object NP and PPs should be weakened in the VP-attachment sentences. When these writing strategies do not work well, writers can use other sentence structures, or change the word order to disambiguate sentences.

Effects of long additional phrases in sentence comprehension.

A long modifier phrase had a negative effect on the whole sentence comprehension, as well as the attachment decision, based on results of the study. According to the L1 study, native speakers were also affected by the three-word phrase in on-line processing (Thornton et al., 2000). It is natural that EFL learners whose working memory capacity is less than L1 readers are affected negatively even in off-line processing. To avoid the negative length effect, teachers could introduce chunking

instruction such as sense group reading with slash which allows learners to reduce the memory loads of redundant information (e.g., *large juicy green* in long condition). In the experimental situations, the long phrase in PP attachment sentence affected learners' whole sentence comprehension, even though it did not modify the VP/NP that each condition forced to disambiguate. The redundant modifiers to important constituents should be manipulated when experimental sentences are constructed.

Adaptation of the think-aloud method to sentence processing study.

This study adopted the think-aloud method, which is frequently used to reveal how EFL learners process passages during reading, or strategy use in vocabulary learning or writing passages. Some significant notes from this study as follows:

Only few sentence processing studies have utilized this method in order to observe readers' sentence representation or strategies, thus Study 2 applied and revised the categorization of think-aloud protocols on the basis of Horiba (1996, 2013). The original categories were constructed for the reading studies, which especially focused on higher-level processing such as making inferences in text. The present study chose to use the categories of semantic processing and syntactic parsing, but some protocols included both syntactic and semantic processing, and could not be distinguished. The present study's scoring criteria considered the protocols that reported phrase boundaries with Japanese translations as semantic processing (e.g., *The sailor found* well, sailor means..what? / *found* he found… in Example 27b), but the protocols also suggested that learners can shallowly parse sentences; thus the protocols included syntactic processing. In the use of the think-aloud method in sentence processing studies, additional criteria should be introduced. As far as the scoring of protocols is concerned, the number of protocols differed according to individuals. Learners who are familiar with such reporting

tasks can talk more, but other participants are not able to speak out their own comments immediately. Thus, solely quantitative analyses will not be sufficient. Instead, the contents of the protocol per se should be explored.

The data on the protocols needs to be obtained carefully, as the data are not always interpreted as a direct reflection of learners' processing. For example, this method is considered to be ineffective for automatized processing such as proficient learners' processing or initial processing, because verbalization was disturbed. The method should be targeted for non-native speakers and less-proficient learners.

Considering these notes, the think-aloud method was partially effective for investigating readers' specific processing or errors, but its analysis using other on-line or off-line data, or both, is essential to understand which situation is similar to that of the other on-line methods. For instance, the think-aloud method cannot fully observe automated processing of proficient readers or readers' initial processing that is too rapid to explain orally. However, at the same time, the eye-tracking method, which is frequently used in sentence processing studies, can measure precise reading-time data of initial processing and second processing separately, but does not provide readers' verbalized detail representations. To compensate for both deficits, the eye-tracking method can be adapted to experiments at the same time as the think-aloud method.

6.4 Concluding Remarks

Japanese EFL learners have difficulties in processing syntactic information, as discussed in many other theories and studies (Clahsen & Felser, 2006a, 2006b; Lim & Christianson, 2013; Rah & Adone, 2010a, 2010b). English education in Japan especially emphasized the importance of grammatical knowledge before communicative skills were focused on. However, teaching methods such as the grammar translation method that was

intended to make learners remember explicit grammatical knowledge were not suitable to enhance the use of implicit grammatical knowledge. The present study investigated the learners' syntactic processing skills and strategies.

The present study focused on the syntactic processing of learners, and one of the significances of the present study is to show learners' individual strategies to comprehend sentences. Unlike L1 readers, proficiency levels of EFL learners are varied, thus their individual differences such as processing speed, capacity of cognitive resources, and size of syntactic, semantic, and lexical knowledge affects their sentence comprehension or strategy use. Regardless of these differences, some EFL sentence processing models and studies focused only on the differences between L1 readers and EFL learners. To inspect the sentence processing of English learners, researchers should take into account the varieties of learners' individual differences, and investigate processing of each phase according to their proficiency levels. The present research is also significant to demonstrate learners' detailed representations of syntactic and semantic information using the translation and think-aloud methods. Particularly, the think-aloud method allows us to understand learners' incorrect sentence representation and their strategies, although the method has some deficits for sentence processing studies. Additional sentence processing studies using the think-aloud method will contribute toward developing sentence processing models.

References

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, *28*, 525–560. doi: 10.1080/01690965.2012.658072

Bartek, B., Lewis, R. L., Vasishth, S., & Smith, M. R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1178–1198. doi: 10.1037/a0024194

Baddeley, A. D. (1986). *Working memory*. Oxford University Press.

Baddeley, A. D. (2000). The episode buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423. doi: 10.1016/S1364-6613(00)01538-2

Bever, T. (1970). The cognitive basis for linguistic structures. In J.R. Hayes (Ed.), *Cognition and the Development of Language* (pp. 278-362). New York: Wiley.

Boland, J. E., & Blodgett, A. (2006). Argument status and PP-attachment. *Journal of Psycholinguist Research*, *35*, 385-403. doi: 10.1007/s10936-006-9021-z

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009), *Processing syntax and morphology: A neurocognitive perspective*. In R. D. Jr. van Valin (Series Ed.), *Oxford surveys in syntax and morphology:* Vol. 6. Oxford university press.

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition*, *33*, 247–271. doi: 10.1017/S0272263110000756

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good enough, underspecified, or shallow language processing. *The Quarterly Journal of Experimental Psychology*, *69*, 817–828.

Christianson, K., Hollingworth, A., Halliwell, J. F., Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368–407. doi: 10.1006/cogp.2001.0752

Clahsen, H., & Felser, C. (2006a). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*, 107–126. doi: 10.1017/S01427 16406060206

Clahsen, H., & Felser, C. (2006b). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3–42. doi: 10.1017/S0142716406060024

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466. doi: 10. 1016/S0022-5371(80)90312-6

Dekydtspotter, L., Donaldson, B., Edmonds, A. C., Fultz, A. L., & Petrush, R. A. (2008). Syntactic and prosodic computations in the resolution of relative clause attachment ambiguity by English-French learners. *Studies in Second Language Acquisition*, *30*, 453–480. doi: 10.1017/S0272263108080728

Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33–68). Cambridge University Press.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*, 141–172. doi: 10.1017/S0272263105050096

Felser, C., Roberts, L., Marins, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*, 453–489.

Ferreira, F., & Patson, N. D. (2007). The “good enough” approach to language comprehension. *Language and Linguistic Compass*, 71–83. doi: 10.1111/j.1749-

818x.2007.00007.x

Fraizer, L., & Clifton Jr., C. (1998). Sentence reanalysis, and visibility. In Fodor, J. D. and Ferreira, F. (Eds), *Reanalysis in sentence processing* (pp. 143–176). Kluwer Academic Publisher.

Fraizer, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*, 291–325. doi: 10.1016/0010-0277(78)90002-1

Fraizer, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178–210. doi: 10.1016/0010-0285(82)90008-1

Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second Language Research*, *21*, 175–198. doi: 10.1191/0267658305 sr257oa

Frenk-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *The Quarterly Journal of Experimental Psychology*, *50A*, 119–148. doi: 10.1080/027249897392251

Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods*, *37*, 581–590. doi: 10.3758/ BF03192728

Fries, C. C. (1952). *The structure of English: An introduction to the construction of English sentences*. New York, NY: Harcourt Brace.

Gabryś-Barker, D. (2008). *Syntactic processing in multilingual performance: A case study*. In D. Gabryś-Barker (Ed.), Morphosyntactic issues in second language studies (pp. 86–106). New York, NY: Multilingual Matters.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*,

68, 1–76. doi: 10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, Language, Brain: Papers from the first mind articulation project symposium* (pp. 95–126). Cambridge: The MIT Press.

Gibson, E., & Pearlmutter, N. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences*, *2*, 262–268. doi: 10.1016/S1364-6613(98)01187-5

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

Grabe, W., & Stoller, F. (2002). *Teaching and researching reading.* Harlow, UK: Pearson Education.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, *29*, 261–290.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*, 194–205. doi: 10.1162/089892999563328

Hemforth, B., Colonna, S., Petrone, C., & D'lmperio, M. (2013). Length matters: Informational load in ambiguity resolution. *Discours*, *12*, 3–21.

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.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, *22*, 369–397.

Horiba, Y. (1996). Comprehension processes in L2 reading: Language competence, textual coherence, and inferences. *Studies in Second Language Acquisition (SSLA),*

18, 433–473. doi: 10.1017/S0272263100015370

Horiba, Y. (2013). Task-induced strategic processing in L2 text comprehension. *Reading in a Foreign Language*, *25*, 98–125.

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. *SSLA*, *27*, 129–140.

Jackson, C. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning*, *58*, 875–909. doi: 10.1111/j.1467-9922.2008.00481.x

Japan Association of College English Teachers [JACET] Basic Word Revision Committee. (2003). *JACET list of 8000 basic words*. Tokyo: Author.

Japan Society of English Language Education (JASELE). (2014). *Trends and current issues in English language education in Japan: Integrating theory and practice* [Eigo kyouikugaku no ima: Riron to Jissen no togo]. Tokyo: Author.

Jegerski, J., & van Patten, B. (2014). *Research Methods in Second Language Psycholinguistics*. New York, NY: Routledge.

Juffs, A. (1998). Main verb versus reduced relative clause ambiguity resolution in L2 sentence processing. *Language Learning*, *48*, 107–147. doi: 10.1111/1467-9922.00034

Juffs, A., & Rodríguez, G.A. (2015). *Second language sentence processing*. New York, NY: Routledge.

Jun, S-A. (2003). Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research*, *32*, 219–249. doi: 10.1023/A:1022452408944

Jun, S-A. (2010). The implicit prosody hypothesis and overt prosody in English. *Language and Cognitive Processes*, *25*, 1201–1233. doi: 10.1080/01690965.2010.503658

Kaan, E., Ballantyne, J. C., & Wijnen, F. (2015). Effects of reading speed on second-language sentence processing. *Applied Psycholinguistics*, *36*, 799–830. doi: 10.1017/S0142716413000519

Kadota, S. (2007). *Shadowing to ondoku no kagaku* [Science of shadowing and oral reading]. Tokyo: Cosmopier.

Kadota, S., & Kuramoto, A. (2007). The process of comprehending GP sentences by Japanese EFL students: A psycholinguistic analysis of eye-movement data. *Journal of the Japan Society for Speech Sciences*, *8*, 19–43.

Kamide, Y., Mitchell, D. C., Fodor, J. D., & Inoue, A. (1998). Relative clause attachment ambiguity: Further evidence from Japanese. Poster presented at the 11th Annual CUNY Conference on Human Sentence Processing, Rutgers, NJ.

Keating, G. D., & Jegerski, J. (2014). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition*, 1–32. doi: 10.1017/S0272263114000187

Kidd, E., & Bavin, E. L. (2005). Lexical and referential cues to sentence interpretation: An investigation of children's interpretations of ambiguous sentences. *Journal of Child Language*, *32*, 855–876. doi: 10.1017/S0305000905007051

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*, 580–602. doi: 10. 1016/0749-596X(91)90027-H

Kim, C. S., Carbary, K. M., & Tanenhaus, M. K. (2013). Syntactic priming without lexical overlap in reading comprehension. *Language and Speech*, *57*, 181–195. doi: 10.1177/0023830913496052

Koda, K (2005). Information integration in sentence processing. In *Insights into second language reading: A cross-linguistic approach* (pp.95–122). Cambridge University Press.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lim, J. H., & Christianson, K. (2013). Second language sentence processing in reading for comprehension and translation. *Bilingualism: Language and Cognition*, *16*, 518–537. doi: 10.1017/S1366728912000351

Love, T., Haist, F., Nicol, J., & Swiney, D. (2003). A functional neuroimaging investigation of the roles of structural complexity and task-demand during auditory sentence processing. *Cortex*, *42*, 577–590. doi: 10.1016/S0010-9452(08)70396-4

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*, 121–136. doi: 10.1080/01690969408402115

Macizo, P., & Bajo, M. T. (2006). Reading for repetition and reading for translation: Do they involve the same processes? *Cognition*, *99*, 1–34. doi: 10.1016/j.cognition. 2004.09.012

Marnis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, *27*, 53–78. doi: 10.1017/S0272263105050035

McRae, K., & Matsuki, K. (2013). Constraint-based models of sentence processing. In T. A. Harley (Series Ed.), *Current issues in the psychology of language: Sentence processing* (pp. 51–75). New York, NY: Psychology Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97. doi: 10.1037/h0043158

Mochizuki, A. (ed.). (2010). *Teaching English as a foreign language based on the new course of study (revised version)* [Kaiteibann shingakushusidouyouryou ni motoduku eigoka kyouikuhou]. Tokyo: Taishukanshoten.

Nakanishi, H. (2012). The contribution of working memory capacity to processing sentences with prepositional phrases for Japanese EFL learners. *Annual review of English language education in Japan (ARELE), 23,* 281–296.

O'Brien, M. G., Jackson, C. N., & Gardner, C. E. (2014). Cross-linguistic differences in prosodic cues to syntactic disambiguation in German and English. *Applied Psycholinguistics, 35,* 27–70. doi: 10.1017/S0142716412000252

Olson, G. M., Duffy, S. A., & Mack, R. L. (1984). Thinking-out-loud as a method for studying real-time comprehension processes. In D. E. Kieras & M. A. Just (Eds), *New Methods in Reading Comprehension Research*. NJ: Lawrence Erlbaum Associates, Hillsdale, pp 253–286.

Osaka, M. (1998). Working memory. In N. Osaka (Ed.), *Yomi: Nou to kokoro no joho-shori* [Reading: Information processing of brain and mind] (pp. 239–262). Tokyo: Asakura shoten.

Pan, H.-Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua, 121,* 221–236. doi: 10.1016/j.lingua. 2010.08.003

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition, 25,* 501–528. doi: 10.1017/S0272263103000214

Rah, A. (2009). *Sentence processing in a second language: Ambiguity resolution in German learners of English* (Doctoral dissertation, University of Cologne). Retrieved from http://kups.ub.uni-koeln.de/4281/

Rah, A., & Adone, D. (2010a). L2 learners' processing of PP attachment ambiguities: A production study. In M. T. Prior, Y. Watanabe, & S.-K. Lee (Eds.), Selected Proceedings of the 2008 Second Language Research Forum (pp. 139–149).

Somerville, MA: Cascadilla Proceedings Project.

Rah, A., & Adone, D. (2010b). Processing of the reduced relative clause versus main verb ambiguity in L2 learners at different proficiency levels. *Studies in Second Language Acquisition*, *32*, 79–109. doi: 10.1017/S027226310999026X

Rayner, K., Carlson, M., Fraizer, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*, 358–374. doi: 10. 1016/S0022-5371(83)90236-0

Rayner, K., Kambe, G., & Duffy, S. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology*, *53*, 1061–1080. doi: 10.1080/713755934

Roberts, L. (2012). Individual differences in second language sentence processing. *Language Learning*, *62*, 172–188. doi: 10.1111/j.1467-9922.2012.00711.x

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition (SSLA)*, *14*, 357–385. doi: 10.1017/ S0272263100011189

Schütze, C. T., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language*, *40*, 409–431. doi: 10.1006/jmla.1998.2619

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*, 238–299. doi: 10.1016/j.cogpsych.2004.03. 001

Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*, 227–267. doi: 10.1016/0010-0277(94)00647-4

STEP: The Society for Testing English Proficiency. (2008a). *The EIKEN test in practical*

English proficiency grade pre 1. Tokyo: Obunsha.

STEP: The Society for Testing English Proficiency. (2008b). *The EIKEN test in practical English proficiency grade pre 2*. Tokyo: Obunsha.

STEP: The Society for Testing English Proficiency. (2009). *The EIKEN test in practical English proficiency grade 2*. Tokyo: Obunsha.

Swets, B., Desmet, T., Clifton C. Jr., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*, 201–216. doi: 10.3758/MC.36.1.201

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, *50*, 355–370. doi: 10.1016/j.jml.2004.01.001

Tanenhaus, M. K. (2004). On-line sentence processing: Past, present and future. In M. Carreiras & C. Jr. Clifton (Eds), *The On-line Study of Sentence Comprehension: Eyetracking, ERPs and beyond* (pp. 371–393). New York, NY: Psychology Press.

Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language*, *27*, 597–632. doi: 10.1016/0749-596X(88)90011-3

Thornton, R., MacDonald, M. C., & Arnold, J. E. (2000). The concomitant effects of phrase length and informational content in sentence comprehension. *Journal of Psycholinguistic Research*, *29*, 195–203. doi: 10.1023/A:1005197012421

Traxler, M. J., Pickering, M. J., & Clifton, C. Jr. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. Journal of Memory and Language, 39, 558–592. doi: 10.1006/jmla.1998.2600

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path

effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89–134. doi: 10.1016/S0010-0277(99)00032-3

van Gompel, R. P. G., & Pickering, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 621–648). Oxford: Elsevier. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.165.9576&rep=rep1&type=pdf

van Gompel, R. P. G., & Pickering, M. J. (2007). Syntactic parsing. In M. G. Gaskell (Ed.) *The Oxford handbook of psycholinguistics* (pp.455–503). Oxford University Press.

van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. Journal of Memory and Language, 45, 225–258. doi: 10.1006/jmla.2001.2773

Witzel, J., Witzel, N., & Nicol, J. (2012). Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics*, *33*, 419–456. doi: 10.1017/S0142716411000427

Yamashita, J., & Ichikawa, S. (2010). Examining reading fluency in a foreign language: Effects of text segmentation on L2 readers. *Reading in a Foreign Language*, *22*, 263–283.

Ying, H. G. (1996). Multiple constraints on processing ambiguous sentences: Evidence from adult L2 learners. *Language Learning*, *46*, 681–711. doi: 10.1111/j.1467-1770.1996.tb01356.x

Yokokawa, H. (1994). On-line processing of syntactically ambiguous sentences by Japanese university students. *ARELE*, *5*, 111–120.

Appendices

Appendix A

Experimental PP-attachment Sentences in Experiments 1, 2, and 4 and Results of Error Analysis of the Translation Task in Experiment 1.

No.		Experimental sentences	Error type			
			A	B	C	D
1	N/S	The policeman chased the criminal with the stolen money.				
	V/S	The policeman chased the criminal with the police car.	1			
	N/L	The policeman chased the terribly angry young criminal with the stolen money.			1	
	V/L	The policeman chased the terribly angry young criminal with the police car.				
2	N/S	The little girl cut the apple with the shiny coating.				
	V/S	The little girl cut the apple with the plastic knife.				
	N/L	The little girl cut the large juicy green apple with the shiny coating.				1
	V/L	The little girl cut the large juicy green apple with the plastic knife.				
3	N/S	The man painted the wall with the door.			4	
	V/S	The man painted the wall with the brush.				
	N/L	The man painted the very clean white wall with the door.	2		2	
	V/L	The man painted the very clean white wall with the brush.	1			
4	N/S	The couple admired the house with a garden.				
	V/S	The couple admired the house with a friend.				
	N/L	The couple admired the really big traditional house with a garden.				
	V/L	The couple admired the really big traditional house with a friend.				
5	N/S	The woman married the man with huge wealth.				
	V/S	The woman married the man with great pleasure.				
	N/L	The woman married the most famous brave man with huge wealth.				
	V/L	The woman married the most famous brave man with great pleasure.				
6	N/S	The doctor cured the patient with the fever.				1
	V/S	The doctor cured the patient with the medicine.				
	N/L	The doctor cured the pale weak hospital patient with the fever.				
	V/L	The doctor cured the pale weak hospital patient with the medicine.			1	
7	N/S	The hospital admitted the patient with cancer.	1			
	V/S	The hospital admitted the patient with urgency.	1			2
	N/L	The hospital admitted the seriously ill young patient with cancer.	1			
	V/L	The hospital admitted the seriously ill young patient with urgency.			2	2
8	N/S	The teenager ordered a pizza with tomatoes.				
	V/S	The teenager ordered a pizza with joy.				1
	N/L	The teenager ordered a grilled hot Italian pizza with tomatoes.				
	V/L	The teenager ordered a grilled hot Italian pizza with joy.				1
9	N/S	The farmer hurt the sheep with the black hair.				
	V/S	The farmer hurt the sheep with the staff.				2
	N/L	The farmer hurt the small unhappy young sheep with the black hair.	1			
	V/L	The farmer hurt the small unhappy young sheep with the staff.			1	4
10	N/S	The sailor found the sea with the whales.			2	

	V/S	The sailor found the sea with the compass.				
	N/L	The sailor found the beautiful large new sea with the whales.				1
	V/L	The sailor found the beautiful large new sea with the compass.				
11	N/S	The walker saw the deer with the disease.		1		
	V/S	The walker saw the deer with the camera.		1	1	
	N/L	The walker saw the poor old black deer with the disease.		1		
	V/L	The walker saw the poor old black deer with the camera.				
12	N/S	The man repaired the watch with the damage.				
	V/S	The man repaired the watch with the tool.				
	N/L	The man repaired the light cheap old watch with the damage.				
	V/L	The man repaired the light cheap old watch with the tool.	1			
13	N/S	The hunter killed the lion with the big face.				2
	V/S	The hunter killed the lion with the gun.				
	N/L	The hunter killed the horribly big dangerous lion with the big face.				
	V/L	The hunter killed the horribly big dangerous lion with the gun.	1			
14	N/S	The angry man hit the cat with the bells.			3	
	V/S	The angry man hit the cat with the umbrella.				
	N/L	The angry man hit the cute little white cat with the bells.			2	
	V/L	The angry man hit the cute little white cat with the umbrella.				
15	N/S	The criminal shot the dog with the wet nose.				
	V/S	The criminal shot the dog with the pistol.				
	N/L	The criminal shot the approaching innocent brown dog with the wet nose.				1
	V/L	The criminal shot the approaching innocent brown dog with the pistol.				
16	N/S	The nurse wet the skin with the burn.				1
	V/S	The nurse wet the skin with the water.				
	N/L	The nurse wet the painful bare red skin with the burn.			1	
	V/L	The nurse wet the painful bare red skin with the water.				
17	N/S	The bad boy damaged the wine glass with the cracks.				6
	V/S	The bad boy damaged the wine glass with the hammer.				
	N/L	The bad boy damaged the costly nice delicate wine glass with the cracks.			1	2
	V/L	The bad boy damaged the costly nice delicate wine glass with the hammer.	1			
18	N/S	The woman dried the wound with the blood.		1	1	
	V/S	The woman dried the wound with the towel.				
	N/L	The woman dried the wet deep ugly wound with the blood.		1		
	V/L	The woman dried the wet deep ugly wound with the towel.		1		
19	N/S	The lady warmed the soup with the garlic.		1		
	V/S	The lady warmed the soup with the heater.				
	N/L	The lady warmed the delicious healthy creamy soup with the garlic.				
	V/L	The lady warmed the delicious healthy creamy soup with the heater.		1		
20	N/S	The tourists followed the fox with the big ears.				
	V/S	The tourists followed the fox with the car.		1		
	N/L	The tourists followed the wild hungry yellow fox with the big ears.				
	V/L	The tourists followed the wild hungry yellow fox with the car.	2			
21	N/S	The man knocked down the tiger with the injury.			1	
	V/S	The man knocked down the tiger with the bat.				
	N/L	The man knocked down the strong female white tiger with the injury.				
	V/L	The man knocked down the strong female white tiger with the bat.				
22	N/S	A fisherman rescued the dolphin with the cut.			1	1
	V/S	A fisherman rescued the dolphin with the boat.				
	N/L	A fisherman rescued the big special wise dolphin with the cut.				1
	V/L	A fisherman rescued the big special wise dolphin with the boat.			1	
23	N/S	The banker knew the code with seven words.		1		
	V/S	The banker knew the code with full confidence.			1	3

	N/L	The banker knew the difficult long secret code with seven words.				
	V/L	The banker knew the difficult long secret code with full confidence.			3	1
24	N/S	The teachers scolded the students with the bad attitudes.				
	V/S	The teachers scolded the students with obvious hate.				
	N/L	The teachers scolded the three different busy students with the bad attitudes.				
	V/L	The teachers scolded the three different busy students with obvious hate.	1		1	
25	N/S	The musician heard the performance with thirty violins.				
	V/S	The musician heard the performance with great respect.			1	1
	N/L	The musician heard the nice long beautiful performance with 30 violins.				
	V/L	The musician heard the nice long beautiful performance with great respect.		1		
26	N/S	The manager demanded the report with sales figures.	2			
	V/S	The manager demanded the report with pride.	2		2	
	N/L	The manager demanded the correct short final report with sales figures.	1			1
	V/L	The manager demanded the correct short final report with pride.				1
27	N/S	The child tasted the chocolate with nuts.				
	V/S	The child tasted the chocolate with delight.				1
	N/L	The child tasted the extra-large dark chocolate with nuts.				1
	V/L	The child tasted the extra-large dark chocolate with delight.			1	1
28	N/S	The young girl loved the man with the smile.			1	
	V/S	The young girl loved the man with passion.				
	N/L	The young girl loved the popular tall handsome man with the smile.			1	
	V/L	The young girl loved the popular tall handsome man with passion.				
29	N/S	The deer smelled the trail with the grass.	1	1		
	V/S	The deer smelled the trail with careful attention.	1	2		
	N/L	The deer smelled the whole endless long trail with the grass.		2	1	
	V/L	The deer smelled the whole endless long trail with careful attention.				
30	N/S	The fortune-teller saw the future with the new problems.				
	V/S	The fortune-teller saw the future with the cards.				
	N/L	The fortune-teller saw the coming dark difficult future with the new problems.			1	
	V/L	The fortune-teller saw the coming dark difficult future with the cards.				
31	N/S	The professor taught the student with the red bag.				
	V/S	The professor taught the student with the blackboard.				
	N/L	The professor taught the polite quiet good student with the red bag.				
	V/L	The professor taught the polite quiet good student with the blackboard.				
32	N/S	The thief hit the man with no money.				
	V/S	The thief hit the man with a chair.				
	N/L	The thief hit the lonely shy English man with no money.				
	V/L	The thief hit the lonely shy English man with a chair.				
33	N/S	The actress kicked the cat with the long tail.				
	V/S	The actress kicked the cat with the foot.			1	
	N/L	The actress kicked the little poor white cat with the long tail.				
	V/L	The actress kicked the little poor white cat with the foot.			1	
34	N/S	The cook ate the cake with the cream cheese.				
	V/S	The cook ate the cake with the fork.				
	N/L	The cook ate the small sweet handmade cake with the cream cheese.				
	V/L	The cook ate the small sweet handmade cake with the fork.				
35	N/S	The king saw the crown with the large diamonds.	1	1		
	V/S	The king saw the crown with the glasses.		2		1
	N/L	The king saw the remarkably beautiful expensive crown with the large diamonds.				
	V/L	The king saw the remarkably beautiful expensive crown with the glasses.		1	1	1

36	N/S	The salesman glanced at the customer with dirty shirts.				1
	V/S	The salesman glanced at the customer with caution.	3			
	N/L	The salesman glanced at the extremely rude young customer with dirty shirts.				
	V/L	The salesman glanced at the extremely rude young customer with caution.				1
37	N/S	The troubled woman looked to the priest with experience.	2	2	1	2
	V/S	The troubled woman looked to the priest with hope.		1		
	N/L	The troubled woman looked to the basically kind popular priest with experience.	1	1	1	
	V/L	The troubled woman looked to the basically kind popular priest with hope.	1		2	
38	N/S	The old man listened to the opera with six acts.				3
	V/S	The old man listened to the opera with a hearing aid.				6
	N/L	The old man listened to the important funny German opera with six acts.				2
	V/L	The old man listened to the important funny German opera with a hearing aid.			1	1
39	N/S	The African hunter aimed at the cow with the wound.			1	1
	V/S	The African hunter aimed at the cow with the arrow.	1			1
	N/L	The African hunter aimed at the slow heavy old cow with the wound.				2
	V/L	The African hunter aimed at the slow heavy old cow with the arrow.	1			

Note. Each experimental sentence was constructed as follows: [S] [V] the ([additional long phrase]) [object NP] with [noun phrases in VP condition]/[noun phrases in NP condition]. N/S = NP-short, V/S = VP-short, N/L = NP-long, and V/L = VP-long condition. The highlighted cells were excluded from the analyses in Experiment 1.

Appendix B

Filler Sentences in Experiments 1, 2, and 4

Structure	Experimental sentences
Ambiguous	The journalist criticized the coach of the runner who was drinking too much.
	The woman knew the photographer of the singer who was reading a book.
	My friend met the assistant of the detective that was fired.
	Patricia saw the teachers of the students that were in class.
	The doctor called in the son of the pretty nurse who hurt herself.
	John spoke to the secretaries of the lawyer who was out of town.
	The doctor recognized the child with the nurse who was feeling very tired.
	The young girl favored the player with the driver who was talking to an old woman.
	Peter yelled at the tall protester with the loudspeaker.
	The teenager washed the soiled sink with the sponge.
	The old man pointed at the girl with the stick.
	The policeman touched the dirty beggar with the stick.
	The hooligan damaged the new shop with the fireworks.
Low attachment	The police detective conducted a search for a weapon.
	The new conductor noticed Arthur's absence from the rehearsal.
	Grandfather could read the numbers in bright colors.
	The financial administrator announced many cuts in the staff.
	The noisy fans gained some confidence in the coach
	I read the article in the magazine.
	Jane finally decided to read the books on the test.
	The kids played all the albums on the shelf
	The older campers questioned John's authority over the group
	The engineers designed the bridge over the river.
	The tourist learned the route through the mountains.
	The doctor contacted the nurses of the lawyer who was talking on the phone.
	The photographer liked the artist with the models who were smiling all the time.
High attachment	The comedian's friends showed their amusement at the party.
	The company lawyers considered employee demands for a month.
	The environmental agency allowed some exemptions from the start
	The board members discussed Mary's inclusion in the afternoon.
	The keen shopper expressed his interest in a hurry.
	The thieves stole all the paintings in the night.
	The scientist read the news reports on Sundays.
	The executive called people on the intercom
	The science teacher encouraged much excitement over the week
	The President suggested a solution to the people.
	The corporate executive considered the issues under pressure.
	The nurse trusted the doctors of the teacher who were preparing to go home.
	The little girl envied the princess with the maids who were eating chocolates.

Appendix C

Original Translation Answers in Japanese in Experiment 1

No.	Experimental sentence and translations
(11a)	The tourists followed the wild hungry yellow fox with the car. [the VP-long condition of Sentence 20]
(11b)	その旅行ガイドは野生のうえた黄色いキツネを車で連れて行った。
(12a)	The salesman glanced at the customer with caution. [the VP-short condition of Sentence 36]
(12b)	そのセールスマンはその客を注意でもって /glanced が分からない
(12c)	セールスマンは客に注意を促した。/glance の意味が分からなかった。
(13a), (17a)	The king saw the (remarkably beautiful expensive) crown with the glasses/large diamonds. [Sentence 35]
(13b)	王は巨大なダイヤモンドと王妃に見せた。
(17b)	王様は大変美しく高く、メガネがついている冠を見た。 [VL]
(14a)	The bad boy damaged the (costly nice delicate) wine glass with the cracks. [the NP conditions of Sentence 17]
(14b)	その悪い男の子はたたいてワイングラスを壊した。/Crack の意味分からず。
(14c)	その悪い少年は cracks でワイングラスを傷つけた。/Cracks は分からない
(14d)	悪い少年は高くてとてもデリケートなワインのグラスを店員のいるところで傷つけた。/The costly nice の意味が分からなかった。
(15a)	The old man listened to the important funny German opera with six acts/a hearing aid. [the long conditions of Sentence 38]
(15b)	第 6 感をつかって、老人はオペラをきいた。
(15c)	6 人の役者が出ているオペラをきいた。
(15d)	老人は聞く目的をもってオペラを聞いた。/Aid の意味が分からない
(15e)	年のとった男性は聴覚障害のあるドイツ人のオペラをきいた。/Funny
(16a)	The hunter killed the (horribly big dangerous) lion with the big face/gun. [Sentence 13]
(16b)	そのハンターは銃でおそろしい大きなこわいライオンに殺された。

Appendix D

Experimental Sentences of Processing unit span test in the pilot study for Experiment 3, Experiments 3 and 4 adopted by Hijikata (2012)

List A
(1) I knocked / on the door / and a dog barked inside. /
(2) It is not kind / to laugh at people / who make mistakes. /
(3) They danced / and drank juice / until two in the morning. /
(4) The twins turned / and smiled / at each other, / sharing a private joke. /
(5) Suddenly one of the gang / kicked him / in the stomach. /
(6) A couple of kids / started throwing stones / at my window. /
(7) One of the children / had cut her foot / on some glass. /
(8) He opened the door wide, / and gestured for me / to come in. /
(9) China's economic output / continues to grow / at a remarkable annual rate. /
(10) Every time I move / I get a pain / in my left shoulder. /
List B
(11) Rickey jumped / across the stream / and ran / all the way home. /
(12) My father worked / at the same place / all his life. /
(13) I was walking / along Main Street / when I met Ben. /
(14) They had to knock / the door down / to get in. /
(15) All drivers should really know / how to change / a flat tire. /
(16) Some college students / gathered wood / while others made a fire. /
(17) An audience of over 5,000 / had filled the hall / that night. /
(18) Their use decreased slightly / in September, / to 149 reported instances. /
(19) She could see / the lights of Hong Kong / shining in the distance. /
(20) If the egg floats / in a glass of water, / it's not fresh. /

Note. Slashes show the boundaries of chunk. Lists A and B were counterbalanced.

Appendix E

Instruction and Examples of Reading Proficiency Test in Pilot Study for Experiment 3

テスト冊子②

学類　氏名

■　表紙にお名前をお書きください。
■　これは長文読解問題です。
- 本文をよく読んで、問題文にもっともよくあてはまる選択肢を選んでください。
- 解答は冊子に直接〇をつけてください。
- それぞれの文章を、読んだことがある場合は、文章の下にあるチェック欄に✔をいれてください。

■　長文は６題で、制限時間は 30 分間です。
- 制限時間が短めなので、時間配分に注意して解くようにしてください。
- 問題は段々難易度が上がっていきます。

① A Famous Desk

One of the most famous pieces of furniture in the United States is a large desk known as the Resolute desk. This wooden desk belongs to the White House, and it has been used by many American presidents. The Resolute desk is over 120 years old and has a very interesting history. In 1845, an English explorer named Sir John Franklin set out from England with two ships to look for a new route to Asia through the Arctic Ocean. Franklin hoped that this would make travel between Europe and Asia faster. However, Franklin and his ships disappeared. In 1848, several ships, including a ship called the Resolute that belonged to the British navy, were sent to the Arctic to search for them.

The captain of the Resolute planned to spend two years searching for the missing ships. However, he failed to find them. Then the Resolute itself became trapped in ice. The captain waited for the ice to melt, but he eventually gave up and decided to return to Britain, leaving his ship behind. He and his crew walked over the ice and sailed home in the other ships that had been sent to the Arctic. Later, the Resolute was discovered by an American fishing boat and taken to a port in Connecticut.

At that time, the relationship between Britain and the United States was not very good, but the American government decided to repair and return the ship. In 1856, it arrived back in Britain. For the next 23 years, the Resolute was used by the British navy. After the navy stopped using the ship, Queen Victoria ordered a desk to be made from its wood. She sent the desk to President Rutherford B. Hayes to thank the Americans. Today, the Resolute desk continues to be an important symbol of the friendship between the United States and Britain.

□　この英文を読んだことがある場合は、左の□に✔を入れてください。

(1) Why did the Resolute leave Britain in 1848?
1　To search for a ship that belonged to the American navy.
2　To take a gift to the president of the United States.
3　To look for two ships that had become lost.
4　To find a faster route between Europe and Asia.

(2) The captain of the Resolute
1　took his crew back to Britain without his ship.
2　decided to leave his ship in a port in Connecticut.
3　found an American fishing boat that was trapped in the ice.

4 hoped to improve Britain's relationship with the United States.

(3) What happened to the Resolute in 1856?
1 It was broken up to make furniture.
2 It was found near the United States.
3 It was repaired by the British navy.
4 It was returned to the British.

(4) What is one thing we learn about the Resolute desk?
1 It was made using wood from an American fishing boat.
2 It was given to President Rutherford B. Hayes by Queen Victoria.
3 The American government bought it from the British navy.
4 The British still hope that the Americans will return it.

Note. This example is derived from Pre-2 grade.

Appendix F

Experimental RRC and URC Sentences and Filler Sentences in Experiment 3

	Sentences
1	The woman sent a doll was very excited.
2	The man warned last night regretted what he did.
3	The man questioned by the lady couldn't answer.
4	The money taken by the students was finally found.
5	The book written by the woman was hard to obtain.
6	The book found in the room was mine.
7	The child excited by the movie let out a loud cry and smiled broadly.
8	The king delighted by the gift was in good spirits.
9	The boat floated on the river finally sank.
10	The newspaper reported the death of the film star sells quickly.
11	The cook who was criticized by the customers used too much salt.
12	The player that was tested by the doctor was allowed to finish the game.
13	The teenager that was overlooked by the store clerk was very frustrated.
14	The parents advised the girl who went to college.
15	The pirates worried the captain who could not sleep.
16	The fan loved the singer who married a movie star.
17	The teacher spoiled the student who got the best grades.
18	The girl we just met is a famous model.
19	The new educational program the government suggested did not please the media.
20	The watch he lost was an expensive Rolex.
21	They never gave me back the CDs I lent them.
22	Sadly, the teacher that Jim was madly in love with got married last week.
23	The man who I met at the party last weekend didn't phone me.
24	The old lady that you're always talking about has just gone into hospital.
25	That bicycle whose wheels are missing is not worth very much.
26	The tennis player whose leg was broken couldn't play in the tournament.
27	The woman whose husband is in prison now wants a divorce.
28	Naomi asked the man she was arguing with to give examples.
29	The agent heard the actress would arrive.
30	The officer charged the driver was drunk.
31	The professor saw the student would succeed.
32	The banker forgot the accountant was stealing.
33	The soldier understood the decision was unfair.
34	Tom promised Meg that he's living with to keep the secret.
35	John taught Nancy that he had hard time with to speak up.
36	The student heard that the decision was wrong.
37	The pilot forgot that the flight was delayed.
38	The waitress saw that the customers would leave.

39	Before the mother ate the candies had disappeared.
40	After the teacher scolded the kids in class became quiet
41	As the woman edited the magazine amused all the reporters.
42	While the woman was eating the creamy soup went cold.

Appendix G1

Original Think-aloud Protocols in Japanese in Experiment 3

No.	Think-aloud protocols
(22a)	The newspaper reported the death of the film star sells quickly. [Sentence 10]
(22b)	The newspaper reported…で、the death of the film, film star. The death of the film star sells quickly. ここまでが主語、sells までが主語なので、映画スターが死んだことを伝えた新聞が、飛ぶように売れたと。
(22c)	The newspaper reported the death of the film star sells quickly. 映画スターの死を報じた新聞はすぐに売れる。
(22d)	The newspaper reported the death of the film star sells quickly. The newspaper reported で区切れるので、the death of the film star sells quickly. The death of the film star、film star の死が、sells? Quickly. うーん、なんとなく分かるけど、意味が訳として良い訳が出てこないので考えています。その newspaper はその映画のスターの死をいち早く報告した、とかだと思います。
(22e)	The newspaper reported その新聞は伝えた。The death of the film star 映画スターの死を。Sells quickly. この sells は… また動詞が出てきた… あ、reported the death of the film star が newspaper にかかってるから、映画スターの死を伝えた新聞は速く売れた。売れる、か。
(22f)	The newspaper reported the death of the film star sells quickly.The newspaper reported で、新聞は報道した。The death of the film star, sells quickly. The death of the film star sells quickly… あ、 reported the death of the film star が、newspaper にかかっていて、reported は newspaper にかかっている、形容詞として使うのかな。えーと、sell で売れたで、 quickly は速くとか速やかにという意味なので。その film star, 多分映画スターの死を報じた新聞は速やかに売れたということになります。
(23a)	The teacher spoiled the student who got the best grades.
(23b)	The teacher spoiled the student who got the best grades. Spoil? Spoil ってなんだっけ? 良い成績を取った生徒を先生が何かした。Spoil…spoil って良い意味の単語ではなかった気がする…けなした? けなした?で取ると、先生はもっとも良い成績を取った生徒をけなした。けなした…ほめた、ほめたではないと思う。
(24a)	The old lady that you're always talking about has just gone into hospital.
(24b)	The old lady that you're always talking about が主語で that 以下が old lady を修飾しています。あなたがいつも話している、その、んと、年老いた女の人は has just gone into hospital. Gone の意味が行ってしまった、だと思うんですけど、hospital が後ろにあるので、病院で行ってしまった、とすると、old lady, 年老いたというのは、亡くなったことを想像しました。あなたがいつもしゃべっていたその年老いた女の人は病院の中で亡くなった、亡くなりました、亡くなってしまいました、です。
(25a)	While the woman was eating the creamy soup went cold.
(25b)	While the woman was eating the.. creamy soup, went cold. えと、While the woman was eating the creamy soup で、これがえーと、While で何々の間。While the woman was eating the creamy soup で、 the woman その女性が、creamy soup クリーミーなスープを食べているという、

	食べていたということになりますので、食べていた間にという意味になります。Went cold、でこの went cold の主語が良く分からないんですが、冷たくなったとかそういう意味なので、soup went.. スープが冷たくなった。スープがこれ、creamy soup の creamy にかかると同時にスープが went の方の主語になっているかが良く分からないですが。Creamy soup がもしこれ went cold の主語だとしたら、スープが冷たくなったという意味かな。とすると、まとめると、女性がクリーミーなスープを食べている間にスープが冷たくなったという意味です。
(26a)	The man questioned by the lady couldn't answer.
(26b)	The man questioned、 by the lady couldn't answer. 主語が The man questioned by the lady、The man questioned by the lady couldn't answer. Couldn't answer が動詞なので、女の人に質問された男の人は答えを言うことができなかった。
(26c)	The man questioned で一回区切れて、by the lady couldn't answer. 主語はこの man で、男の人は女性が答えられないことを不思議がった。

Appendix G2

Original Think-aloud Protocols in Japanese in Experiment 4

Example No.	Think-aloud protocols
(27a)	The sailor found the beautiful large new sea with the compass. [Sentence 10]
(27b)	The sailor found えっと、セイラーってなんかなんだっけ。Found 見つけたよ。Beautiful large new sea 美しいきれいな海を見つけたよ。With the compass コンパスを使って見つけた。
(28a)	The man knocked down the strong female white tiger with the injury.
(28b)	The man knocked down the strong female white tiger with the injury. これは knock down しか動詞がないから knock down が動詞。The man が本主語。The man がノックダウンした。何を、が以下か。The strong female white…ああ、すげえ! 白いトラをしかも強くて女性のトラを倒しちゃった。With the injury、まあケガしながらってことか。死にそうで勝ったってことか。ん、待てよ。あ、ケガしてるトラか。ええ? どっちだこれ。ケガしてるトラに勝ったのか、男の方がケガしているのか。まあ位置的にトラの方が近いから with the injury tiger だよな。
(28c)	The man knocked down その男の人は knocked 叩いた。knocked down あ、ノックダウンなんで、倒された? あ倒したですね。the ティガー with the journey あ、injury. with the injury タイガー with the injury. トラを怪我を負った、with the injury が man にかかるのか、まあ多分 tiger にかかると思うんですけど。ケガをもともと負っていた。もしくは the man にかかった場合はケガを男の人が負っていたんですけど、ちょっとどっちだかよく分かんないです。
(29a)	The hunter killed the lion with the big face.
(29b)	The hunter killed the lion with the big face. 主語は The hunter 動詞は killed 目的語は the lion で with the big face は the lion を修飾しているという形、文の形で、意味は、そのハンターは大きな顔のライオンを殺しただと思います。
(30a)	The woman saw the coming dark difficult future with the cards.
(30b)	The woman saw the coming dark difficult future with the cards. with the cards が、saw A with B という感じなんですかね。よく分からないけど、まあ、タロット占い的なカードによって、タロット占い的なカードから、差し迫った闇の、闇深い難しい将来を、その人は垣間見たということですかね。 【理解問題後】ああ、今のは違いますね。その女のひとが霊媒師だったんですね、される側じゃなくて。

Appendix H

Grammar Proficiency Test in Experiment 4

文法テスト　　氏名: _____

I. 以下の質問に対して会話が成立する選択を選んで記号に〇をつけてください。

1. When can we meet again?

(a) When are you free?　(b) It was two days ago.　(c) Can you help me?

2. My aunt is going to stay with me.

(a) How do you do?　(b) How long for?　(c) How was it?

3. When do you study?

(a) at school　(b) in the evenings　(c) in the library

4. Would you prefer lemonade or orange juice?

(a) Have you got anything else?　(b) If you like.　(c) Are you sure about that?

5. Let's have dinner now.

(a) You aren't eating.　(b) There aren't any.　(c) Tom isn't here yet.

II. 次の下線部に当てはまるものを選んで記号に〇をつけてください。

6. The snow was ..... heavily when I left the house.

(a) dropping　(b) landing　(c) falling　(d) descending

7. I can't find my keys anywhere - I ..... have left them at work.

(a) can　(b) must　(c) ought　(d) would

8. I'll pick you up ..... to the conference.

(a) by the way　(b) in the way　(c) from the way　(d) on the way

9. When a car pulled out in front of her, Jane did well not to …... control of her bicycle.

(a) miss　(b) lose　(c) fail　(d) drop

10. According to Richard's ..... the train leaves at 7 o'clock.

(a) opinion (b) advice (c) knowledge (d) information

11. When you stay in a country for some time you get used to the people's ..... of life.

(a) habit (b) custom (c) way (d) system

12. The builders are ..... good progress with the new house.

(a) getting (b) doing (c) making (d) taking

13. She is now taking a more positive ..... to her studies and should do well.

(a) attitude (b) behaviour (c) manner (d) style

14. My father ..... his new car for two weeks now.

(a) has had (b) has (c) is having (d) had

15. What differences are there ..... the English spoken in the UK and the English spoken in the US?

(a) among (b) between (c) beside (d) with

16. At 6 p.m. I started to get angry with him because he was late .....

(a) as usual. (b) in general. (c) typically. (d) usually.

17. ..... you get your father's permission, I'll take you skiing next weekend.

(a) Although (b) Provided (c) As (d) Unless

18. A local company has agreed to ..... the school team with football shirts.

(a) contribute (b) supply (c) give (d) produce

19. I started working for this company two years .....

(a) earlier. (b) ago. (c) before. (d) after.

20. I really enjoy stories that are ..... in the distant future.

(a) found (b) set (c) put (d) placed

21. That old saucepan will come in ..... when we go camping.

(a) convenient (b) fitting (c) handy (d) suitable

22. Anyone ..... after the start of the play is not allowed in until the interval.

(a) arrives (b) has arrived (c) arriving (d) arrived

23. The house was burgled ..... the night.

(a) while (b) between (c) during (d) for

24. I didn't ..... driving home in the storm so I stayed overnight in a hotel.

(a) fancy (b) desire (c) prefer (d) want

25. The judge said that those prepared to..... in crime must be ready to suffer the consequences.

(a) involve (b) engage (c) undertake (d) enlist

26. This is an example ..... a good essay.

(a) in (b) with (c) on (d) of

27. Marianne seemed to take ..... at my comments on her work.

(a) annoyance (b) insult (c) offence (d) indignation

28. You should not have a dog if you are not ..... to look after it.

(a) prepared (b) adapted (c) arranged (d) decided

29. The farmhouse was so isolated that they had to generate their own electricity .....

(a) current. (b) supply. (c) grid. (d) power.

30. I'm really worried ..... my best friend.

(a) about (b) at (c) of (d) for

Note. Items 8, 19, 23, 26, and 30 were related to the knowledge concerning prepositions and were added by the experimenter.

Appendix I1

Results of MANOVA for Frequency Data of Correct Answers

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
Category 1: Semantic analysis						
Attachment (A)	1	0.35	0.35	1.35	.259	.02
Length (L)	1	4.85	4.85	16.37	.001	.26
A × L	1	0.00	0.00	0.00	.988	.00
Error (A)	20	5.11	0.26			
Error (L)	20	5.93	0.30			
Error (A × L)	20	2.63	0.13			
Category 2: Syntactic analysis						
Attachment	1	0.00	0.00	0.02	.905	.00
Length	1	0.05	0.05	0.28	.601	.01
A × L	1	0.02	0.02	0.15	.700	.00
Error (A)	20	2.19	0.11			
Error (L)	20	3.27	0.16			
Error (A × L)	20	2.24	0.11			
Category 3: Contextual information						
Attachment	1	0.00	0.00	0.03	.862	.00
Length	1	0.00	0.00	0.00	.986	.00
A × L	1	0.00	0.00	0.21	.651	.00
Error (A)	20	0.26	0.01			
Error (L)	20	0.23	0.01			
Error (A × L)	20	0.20	0.01			
Category 4: Metalinguistic analysis						
Attachment	1	0.11	0.11	3.09	.094	.05
Length	1	0.00	0.00	0.01	.945	.00
A × L	1	0.02	0.02	0.49	.493	.01
Error (A)	20	0.74	0.04			
Error (L)	20	0.48	0.02			
Error (A × L)	20	0.83	0.04			
Category 5: Oral reading						
Attachment	1	0.07	0.07	3.19	.089	.04
Length	1	0.17	0.17	8.30	.009	.10
A × L	1	0.00	0.00	0.09	.768	.00
Error (A)	20	0.43	0.02			
Error (L)	20	0.41	0.02			
Error (A × L)	20	0.58	0.03			

Note. $N = 19$.

Appendix I2

Results of MANOVA for Frequency Data of Incorrect Answers

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
Category 1: Semantic analysis						
Attachment (A)	1	4.07	4.07	2.03	.170	.03
Length (L)	1	0.06	0.06	0.06	.812	.00
A × L	1	1.57	1.57	0.44	.516	.01
Error (A)	20	40.21	2.01			
Error (L)	20	19.23	0.96			
Error (A × L)	20	72.04	3.60			
Category 2: Syntactic analysis						
Attachment	1	0.64	0.64	1.20	.287	.03
Length	1	0.13	0.13	0.63	.436	.01
A × L	1	0.13	0.13	0.45	.509	.01
Error (A)	20	10.68	0.53			
Error (L)	20	4.19	0.21			
Error (A × L)	20	5.85	0.29			
Category 3: Contextual information						
Attachment	1	0.07	0.07	1.51	.234	.02
Length	1	0.24	0.24	5.87	.025	.08
A × L	1	0.07	0.07	1.51	.234	.02
Error (A)	20	0.99	0.05			
Error (L)	20	0.82	0.04			
Error (A × L)	20	0.99	0.05			
Category 4: Metalinguistic analysis						
Attachment	1	0.15	0.15	1.35	.260	.02
Length	1	0.15	0.15	1.35	.260	.02
A × L	1	0.03	0.03	1.51	.234	.00
Error (A)	20	2.17	0.11			
Error (L)	20	2.17	0.11			
Error (A × L)	20	5.29	0.26			
Category 5: Oral reading						
Attachment	1	0.11	0.11	0.50	.487	.01
Length	1	0.01	0.01	0.08	.776	.00
A × L	1	0.30	0.30	0.79	.386	.00
Error (A)	20	4.27	0.21			
Error (L)	20	2.86	0.14			
Error (A × L)	20	7.58	0.38			

Note. $N = 19$.

Appendix J

Dendrogram of Cluster Analysis in Experiment 4