Generation of Knowledge-Based Inferences in Japanese EFL Learners' Reading Comprehension

A Dissertation

Submitted to the University of Tsukuba

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Linguistics

Mayuko KOBAYASHI

2017

Abstract of the Dissertation

Generation of Knowledge-Based Inferences

in Japanese EFL Learners' Reading Comprehension

by

Mayuko KOBAYASHI

Understanding text is constructing a coherent mental representation of the information in a text (van den Broek & Gustafson, 1999). When information that is not explicitly stated in the text is activated, inferences are generated. Deeper comprehension is achieved when readers activate and apply their appropriate prior knowledge in order to generate inferences regarding what the text is about. The process of inference generation is a key component of fluent reading, and as such, several aspects of the process are of interest, including the types of inferences readers make and the factors that determine whether and when inferences are generated. In the English as a foreign language (EFL) context in Japan, learners are required not only to comprehend explicit textual information, but also to grasp implicit information (e.g., theme and the protagonist's goal) through inference generation process. Nevertheless, they sometimes fail to make use of their knowledge fully and generate inferences, resulting in a shallow understanding of the text.

The significance of inference generation is widely accepted not only in the field of first language (L1) reading comprehension, but also in second language (L2) reading comprehension. Empirical evidence from L1 and L2 reading research indicates that when readers comprehend

a text, they attempt to build a mental representation by making inferences to coherence with the explicit contents in the text (e.g., Graesser, Singer, & Trabasso, 1994; Horiba, 1996, 2000). However, since very little research has been done on inference generation in L2 context compared with that of L1 (e.g., Collins & Tajika, 1996; Horiba, 1996, 2000; Yoshida, 2003), little is known about the mechanism of L2/EFL readers' inference generation. The purpose of the current dissertation was to investigate the generation process of some types of knowledge-based inferences (e.g., thematic, predictive) by EFL readers and the effects of text and learner factors on these inferences. Addressing this issue, three main research questions (RQs): were set as follows:

RQ1: How do the text and learner factors affect EFL readers' knowledge-based inference generation?

RQ2: How does the presentation of analogy with either surface or structural similarities affect EFL readers' knowledge-based inference generation?

RQ3: How do the text and learner factors affect EFL readers' explicit text comprehension after reading?

First, Study 1 examined the generation of thematic inferences in Japanese EFL learners' sentence-level reading. In sentence-level reading, processes are necessary to form structures that represent the syntactic and conceptual relationships among words in phrases or clauses. These processes help in encoding propositions, abstract units that represent the meaning of a sentence. As the result of Experiment 1, there were no differences in proficient and less proficient EFL readers' ability to generate knowledge-based inferences. Both groups were sensitive to the thematic inference items. When the author examined their representations of simple sentences, there were remarkable similarities in their representations. Both groups constructed representations that were consistent with the context-appropriate senses of

inference target words.

Second, Study 2 examined the generation of predictive inferences in EFL learners' text-level reading and investigated the effect of the text factors and learner factors on the generation of predictive inferences in EFL reading. At text-level reading, processes are required to form connections among successive propositions in a text. The processes mentioned above involve accessing not only knowledge about the language (e.g., word meaning, syntax), but also knowledge about the world. In Experiment 2, the influence of contextual constraints and L2 reading proficiency on predictive inference generation was investigated. As a result, it was found that proficient EFL readers generated more predictive inferences in highly constrained contexts than less proficient readers. At the same time, however, it was indicated that the generated inference information was encoded less weakly to mental representations compared with explicit text information. In Experiment 3, the specificity of predictive inferences was examined using strategic inference question task. The inference question, "What will happen next?," was given intentionally so as to draw readers' predictive inferences. As a result, it was found that the specificity of the inferences that are generated changes according to the strength of contextual constraints.

Third, Study 3 investigated how the text features work as access cues to the readers' knowledge structure using analogical texts. Analogical texts as advance reading organizer are used so as to strategically activate readers' prior knowledge. Showing an analogy activates the readers' schema related to sentence information, and understanding is then promoted by assimilating new information into the schema (Taniguchi, 1988). However, it was still unclear what kind of text information triggered activation of knowledge-based inferences in EFL reading. Experiment 4 examined how the similarity features between analogical texts (i.e., similarity types and causality) and L2 reading proficiency affect EFL learners' reading

comprehension through analogical transfer. The results showed that analogical transfer occurred in EFL reading comprehension. However, each process in the series was affected by different factors. Accessing and mapping processes were driven by the causality between texts. Meanwhile, the process of encoding the analogical information to EFL learners' text representations was mainly driven by the surface similarity between the texts. Moreover, it revealed that the transfer of causal structure from the source analog to the target representation effectively enhances text comprehension, but it requires a certain level of L2 reading proficiency. Experiment 5 checked whether EFL learners were activating the between-text-level theme of the text. The results showed that EFL learners were activating thematic information which was shared with analogical texts. Also, it was shown that text information with predictive features facilitates information transfer across texts and increases the production rates of recalled text information where themes are similar.

There are two major findings of the current study. The first finding is that a two-phase process have been at work in the generation of knowledge-based inferences by Japanese EFL learners. In the first phase, various concepts or information are activated that share features in common with prior text information and the readers' prior knowledge during reading. In the second phase, specific inference information is selected based on contextual constraints, which are encoded into representations. This suggested the text comprehension process in which EFL learners read on while predicting multiple outcomes and eventually predict an outcome that is coherent, based on newly acquired information. The second finding is that providing analogous information prior to the reading of texts have different effects on text comprehension depending on readers' L2 reading proficiency level. Proficient EFL learners can recall much of the text information in post-reading task by relying on information about structural similarities that lacks individual surface similarity features. In contrast, less proficient EFL readers find it to be

difficult to do the same, due to such factors as their limited cognitive resources. Nevertheless, it has been shown that the recall of text information can be enhanced by directing students' attention to structural causal information with the provision of inference questions or addition of predictive information to the text.

Lastly, this dissertation ends by providing some pedagogical implications for EFL classroom instruction and the future research in reading comprehension.

Acknowledgements

It is my great pleasure to acknowledge the many people who have given me guidance, assistance and encouragement.

First and foremost, I would like to express my deepest and sincere appreciation to my supervisor, Professor Yuji Ushiro at the University of Tsukuba. He consistently encouraged me to continue my research and taught me how to pursue academic research with enthusiasm. He also gave me many opportunities to learn the theory and practice of teaching reading skills in a foreign language.

I would also like to thank the other committee members: Professor Hirosada Iwasaki at the University of Tsukuba for his constructive advice on pedagogical implications of this research and academic writing; Professor Akira Kubota at the University of Tsukuba for his insightful questions, which have led me to think more deeply about many issues of my research. I am grateful to Professor Negishi Masashi of Tokyo University of Foreign Studies, who is an external member of the dissertation committee. His generous support and incisive comments gave me the opportunity to rethink the composition of the thesis. This dissertation could not have been finished without their guidance.

I am pleased to acknowledge the support of the Japan Society for the Promotion of Science (JSPS) for giving me a generous grant that made it possible to carry out this research.

My special appreciation goes to my senior associate members, especially Dr. Haruka Shimizu and Dr. Akari Kai. They consistently provided invaluable and detailed comments and helped me to proceed with my research. They pointed me in the direction of worthy academic research. Many thanks also to my junior associate members, Dr. Shingo Nahatame, Dr. Yusuke Hasegawa, Dr. Akira Hamada, Dr. Yukino Kimura, Dr. Natsumi Tanaka, Mr. Yoshinobu Mori,

and Mr. Masaya Hosoda, for their helpful comments and kind encouragements. I wish to express my heartfelt thanks to my colleague and friend, Dr. Shuichi Takaki who has inspired me strive towards my goal. He kindly cooperated in my data analysis and constantly supported me throughout my postgraduate years.

Finally, I would like to thank my family. Words cannot express how grateful I am to my mother-in law and father-in-law for their constant support. I am also grateful to my husband and son for their support, understanding, and love. Lastly, I would like to thank my parents who always believed in me and encouraged me throughout my life.

Mayuko KOBAYASHI

December, 2017

Table of Contents

Page

Page

Page

Page

List of Tables

Table Page

Chapter 6

List of Figures

List of Appendices

Major Abbreviations and Acronyms

ANOVA	Analysis of Variance
CC	Contextual Constraint
EFL	English as a Foreign Language
FOR	First-Order Relation
HOR	Higher-Order Relation
IU	Idea Unit
L1	First Language
L2	Second Language
STEP	Society for Testing English Proficiency
TAU	Thematic Abstraction Unit

Chapter 1

Introduction

1.1 Context of the Problem

In recent years, Japanese foreign language education has been shifting so as to cultivate the practical communicative ability of learners in actual language usage settings. Accordingly, the latest Course of Study for foreign languages published by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) states that not only the comprehension of the explicit information of the text, but also the comprehension of implicit information such as the themes or the author's intent are equally important. For the understanding of such implicit information, learners require a procedure that connects text information and their own knowledge. In other words, for the deeper understanding of the text, *learning from the text* (i.e., the acquisition of information and knowledge included in the text) is more important than the mere *learning of the text* (i.e., understanding or memorization of the text itself) (Kintsch, 1994). A similar statement is mentioned in the Common European Framework of Reference for language (CEFR), the new global standard for the evaluation of foreign language education. In the CEFR Can-do statements for reading comprehension, the following statement is written: "Can understand a wide range of long and complex texts, appreciating subtle distinctions of style and implicit as well as explicit meaning." That is, in addition to the comprehension of the text itself, to understand the implied meaning of the text is a vital component for successful reading. However, the above statement is categorized in the C2 level of the CEFR-J Can-do descriptor list[1], which is labeled as mastery level. CEFR-J is an adaptation of the CEFR for specific

[1] The CEFR-J Can-do descriptor list (see www.tufs.ac.jp/ts/personal/tonolab/cefr-j) consists of L2 levels of proficiency which describe communicative competencies in listening, reading, spoken production, spoken interaction, and writing.

English language teaching contexts in Japan. (TUFS Tonolab, 2012). It was shown that over 80% of Japanese EFL learners fell within the A1 and A2 level of the CEFR-J (also known as the Basic User level). Taking the following facts into consideration, we can presume that it is quite difficult for Japanese EFL learners to reach the level of comprehension beyond the text itself.

Reading to learn from texts particularly needs the process of inferencing because in order to acquire new knowledge, readers need to generate appropriate inferences from the textual information and from their own knowledge until they reach a deeper understanding of the text. The significance of inferencing is widely accepted not only in the field of L1 reading comprehension, but also in L2 reading comprehension. Empirical evidence from L1 and L2 reading research indicates that when readers comprehend a text, they attempt to build a mental representation by making inferences to coherence with the explicit contents in the text (e.g., Graesser, Singer, & Trabasso, 1994; Horiba, 1996, 2000). According to Campion and Rossi (1999) the term *inference* in reading comprehension is defined as: Inferences are all pieces of information that lead the reader to connect two events during text comprehension, or all information added by readers to information explicitly stated in a text. Graesser and Kreuz (1993) also explained inferences as follows: "Inferences are produced during text comprehension when world knowledge structures are activated, and the content of these structures is incorporated into the constructed meaning of the text" (p.146). Namely, since the words on a printed page are often not completely explicit, a variety of inferences are generated to go beyond the information given from the text and fill in the gaps of information. For example, the goals and plans information of the text that motivate characters' actions, traits, emotions, the causes of events, the properties of objects, spatial relationships among entities in the story, expectations about future episodes in the plots, referents of readers, and so on. In this way,

successful reading is achieved when a mental world is constructed with the interactive relationships between text information and readers.

As noted above, for the inference generation process, the activation of readers' knowledge structure acts a strong factor. Although, when we read the text in L1, various inferences based on our knowledge and experience are generated unconsciously, it becomes difficult in L2 text reading. In the classroom, many students tend to translate English into Japanese word by word. Due to this translation method, students sometimes fail to make use of their knowledge fully, resulting in a shallow understanding of the text. The following famous paragraph adapted from Bransford and Johnson (1972) illustrates the importance of knowledge activation in reading comprehension.

> The procedure is actually quite simple. First you arrange things into different groups. Of course, one pile may be sufficient depending on how much there is to do. If you have to go somewhere else due to lack of facilities that is the next step, otherwise you are pretty well set. It is important not to overdo things. That is, it is better to do too few things at once than too many. In the short run this may not seem important but complications can easily arise. A mistake can be expensive as well. At first the whole procedure will seem complicated. Soon, however, it will become just another facet of life. It is difficult to foresee any end to the necessity for this task in the immediate future, but then one never can tell. After the procedure is completed one arranges the materials into different groups again. Then they can be put into their appropriate places. Eventually they will be used once more and the whole cycle will then have to be repeated. However, that is part of life (p. 722).

If we read it word by word, it is difficult for us to comprehend what is written. However, it makes it very easy for us to understand this paragraph if we notice that it has something to do with *doing laundry*. As can be seen from the above example, readers activate their relevant knowledge or concepts through information on *what the story is about,* not by confirming word

by word. By doing top-down processing, which selects the necessary information and makes a prediction through the text, readers build a concrete mental representation.

Thus, readers' background knowledge is critical for inference generation and text comprehension. However, the knowledge needed to make inference is not equally accessible in all contexts. Knowledge information is more or less accessible depending on how quickly it can be retrieved (Barnes, Denis, & Haefele-Kalvatis, 1996). At the same time, less accessible knowledge is less likely to be used during text comprehension because such information takes longer to retrieve in focal text processing, and more accessible knowledge is more likely to be used to generate inference during text comprehension. Therefore, the accessibility of knowledge is related to some types of inferences generated based on readers' knowledge (i.e., *knowledge-based inferences*) in general. Mainly, there are two broad categories of inferences that can be drawn during reading: *necessary inferences* and *elaborative inferences*. Necessary inferences maintain a coherent story line by adding unstated but important information to explicit text. They form a causal link between knowledge and text that helps infer *why* an event occurred. Elaborative inferences embellish story content and amplify its context, even though they are not central to textual cohesion. They elaborate the image of the text so that *what* an event is like maybe inferred. Elaborative inferences are thought to strengthen long-term memory for text and, by making concepts more concrete, may facilitate the integration of subsequent propositions.

Considering the role of these inferences in text comprehension, necessary inferences may be minimally affected by knowledge accessibility (Barnes, Denis, & Haefele-Kalvatis, 1996). Since necessary inference is essential for readers to understand the causal relationship of the separate information in the text and make a coherent mental representation, even when the knowledge needed to make necessary inference is not easily accessible, the appropriate

knowledge information is retrieved more strategically regardless of the accessibility. On the other hand, since elaborative inferences are not necessary for maintaining text coherence, knowledge accessibility plays a greater role in determining whether elaborative inferences are generated. In a related matter, McKoon and Ratcliff (1992) stated that elaborative inferences are automatically generated only when the information is easily accessible.

From the reasons stated above, it is assumed that elaborative inferences are more strongly influenced by various text factors (e.g., contextual constraint, text type) and learner factors (e.g., L2 reading proficiency, working memory: WM). In fact, in the L2 research context, Yoshida (2003) has pointed out that learners with large WM capacity generate more inferences regardless of inference type and elaborative inferences are likely to be generated with higher level L2 reading proficiency learners. However, since very little research has been done on elaborative inference generation in L2 context compared with that of L1 (e.g., Collins & Tajika, 1996; Horiba, 1996, 2000; Yoshida, 2003), little is known about the mechanism of L2 readers' inference generation.

The main purpose of the present study is to clarify the generation process of inference which is generated based on the knowledge of readers. In particular, the focus will be on elaborative inferences (e.g., predictive inferences, thematic inferences) which are likely to be influenced by knowledge accessibility. This study also aims to shed light on how some text and learner factors affect the inference generation process.

1.2 Organization of the Thesis

This dissertation is organized with the present chapter and five other chapters. Chapter 1, Introduction, refers to the relationship of readers' knowledge structure and inference generation in reading comprehension, and the purpose of the current study.

Chapter 2, Review of Related Literature, presents an overview of the previous L1 and L2 research that is related to the present study. It summarizes the components of reading processes, inferences for constructing a mental representation, reading models related to inference generation, functions of readers' knowledge structure, and the effect of some text and learner factors on EFL learners' reading.

Chapter 3, Study 1 examines the generation of thematic inferences in Japanese EFL learners' sentence-level reading. In sentence-level reading, processes are necessary to form structures that represent the syntactic and conceptual relationships among words in phrases or clauses. These processes help in encoding propositions, abstract units that represent the meaning of a sentence.

Chapter 4, Study 2 investigates the generation of predictive inferences in EFL learners' text-level reading and the effect of the text factor and learner factor on the generation of predictive inferences in EFL reading. At text-level reading, processes are required to form connections among successive propositions in a text. The processes mentioned above involve accessing not only knowledge about the language (e.g., word meaning, syntax), but also knowledge about the world. Study 2 includes two experiments which examine the generation of predictive inferences. Experiment 2 uses a probe recognition task for the activation and encoding of predictive inferences. Corrected data are analyzed in terms of the strength of contextual constraint and L2 reading proficiency. Experiment 3 is conducted to clarify the influence of contextual constraint over the content of readers' predictive inferences. In this experiment, the strategy instruction "*What will happen next?,"* which promotes the generation of predictive inferences was given to readers, and a qualitative analysis was conducted on the open-ended answers according to their categories.

Chapter 5, Study 3 looks into how the text information works as access cues to the readers'

knowledge structure. Experiment 4 examined how the similarity features between texts (i.e., similarity types and causality) and L2 reading proficiency affect EFL learners' reading comprehension through analogical transfer. Experiment 5 checked whether EFL learners were activating the between-text-level theme of the text.

Chapter 6, General Discussion and Conclusion, integrates the results of the three studies reported in the preceding chapters and discusses the process of knowledge-based inference generation in EFL reading comprehension. This chapter also summarizes the overall findings, provides some pedagogical implications, and gives suggestions for further research. Figure 1.1 shows the organization of the three studies in this dissertation, and Figure 1.2 illustrates the relationship of each text and learner factors affecting EFL reading processes.

【Study 1】 Inference Generation in Sentence-Level Reading	Factor	Measurement
Exp 1: Effects of Contextual Constraint on Thematic Inference Generation	a) L2 Reading Proficiency b) Contextual Constraint	Lexical Decision Task Cued Recall Task

【Study 2】 Inference Generation in Text-Level Reading	Factor	Measurement
Exp 2: Effects of Contextual Constraint on Predictive Inference Generation	a) L2 Reading Proficiency b) Contextual Constraint	Probe Recognition Task Cued Recall Task
Exp 3: Effects of Contextual Constraint on the Specificity of Predictive Inferences	Contextual Constraint	Inference Generation Task Recall Task

【Study 3】 Inference Generation in Between-Texts-Level Reading	Factor	Measurement
Exp 4: Effects of the Different Types of Similarities Between Analogical Texts on Analogical Transfer	a) L2 Reading Proficiency b) Similarity Types c) Causality	Probe Recognition Task Cued Recall Task
Exp 5: Effects of Structural Similarity Between Analogical Texts on Thematic Inference Generation	a) L2 Reading Proficiency b) Similarity Types	Reading Time for Target Cued Recall Task

Figure 1.1. Organization of the current dissertation.

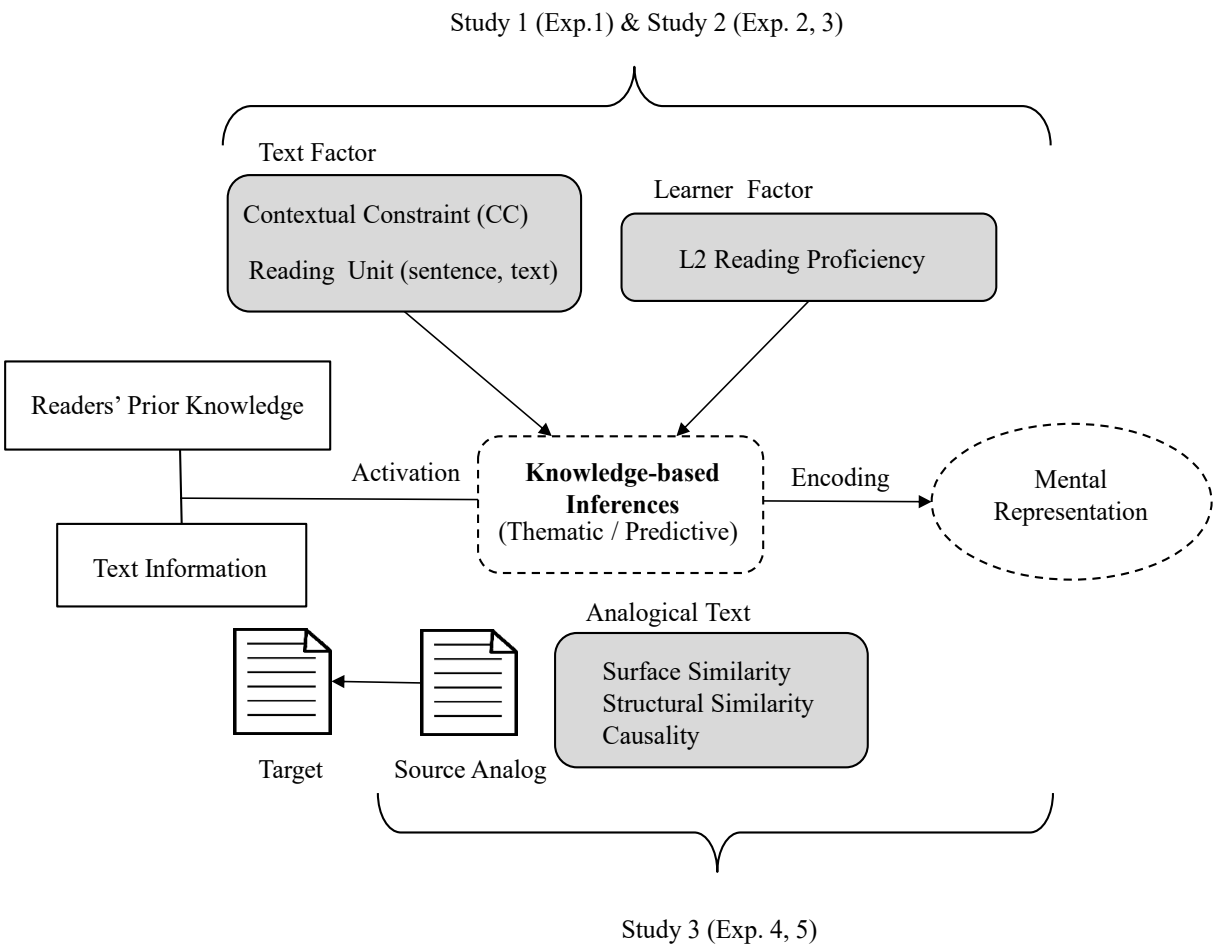


Figure 1.2. The relationship of each text and learner factors affecting reading processes.

Chapter 2

Review of Related Literature

This chapter presents an overview of past research related to the present study, and is mainly organized into three sections of reviews. The first section, Section 2.1, describes how readers understand texts in terms of the multiple levels of mental representations and components of reading processes. The second section 2.2, focus on the on the inference processes. In this section, the definition, categories, and theories of inferences are provided in detail. The third section 2.3 explains about several knowledge-based inferences. Even among the inferences, what types of knowledge-based inferences are generated during reading will be reviewed based on the inference categorization from some famous reading models (e.g., the constructionist theory , the minimalist hypothesis).

2.1 Components of Reading Process

When comprehending a text, it is generally assumed that readers employ a variety of cognitive processes simultaneously. Readers recognize words from the letters in the text, and understand sentences from the association of those words. Moreover, they must integrate the relationships between sentences and grasp the global meaning of the entire text. To put it another way in terms of psycholinguistic reading research, these components of the reading process are explained as follows. Readers recognize words automatically, form meaning propositions quickly, integrate propositional information into the text model, and restructure the text model to reflect the main ideas of the text being read. It is also clear that reading requires both sufficient knowledge of language and knowledge of the world as the basic supporting foundation on which to build comprehension (Grabe, 2000; van Dijk and Kintsch, 1983).

According to Grabe (2000), these reading components can be divided into either l*ower-level processing* or *higher-level processing.* Processes such as word recognition, lexical access, syntactic parsing, and propositional integration are included in lower-level processing. Particularly, word recognition fluency is critical for reading because readers need to see word forms and access the appropriate meanings rapidly and accurately. On the other hand, higher-level processing is defined as "working with larger units of information and information contributed by the reader" (p.230). Such identification of the main and supporting ideas and interpretation of information, which are primarily concerned with constructing the global meaning of texts, are commonly referred to as higher-level processes (Grabe, 2000; Grabe & Stroller, 2002; Koda, 2005). In addition, most L1 and L2 researchers believe that some types of inferences are generated while reading, and that syntactic and discourse signaling in texts is used to restructure the text network. By such higher-level processing, readers construct a situation model, which reflects text information as well as readers' background knowledge, individual responses, and affective responses. However, higher-level processing generates considerable disagreement over the specific processing mechanisms involved in text comprehension. Grabe (2000) pointed out this problem as follows: "The role of inferencing, contextual information, reader background knowledge, discourse structuring knowledge, and reading strategies (executive processing) have generated a range of alternative positions" (p.234).

The main interest of the current thesis is to focus on inference generation that is part of higher-level processing, and to clarify how contextual information and readers' knowledge structure relates to their inference process.

2.1.1 The Theory of Levels of Mental Representation

The primary goal of reading is to construct a *mental representation* of meaning from a text (Grabe & Stroller, 2002). This section describes the elements that compose the mental representation constructed while reading. A varied model of text comprehension should not only explain how information is extracted from the text itself, but also how this information is interpreted in light of the reader's knowledge. That is, to construct a meaningful representation requires that the reader employ their background knowledge to interpret the textual content. The distinction between memory for the surface structure and the meaning of a text was first introduced by Bartlett (1932) and, reformulated by Kintsch and van Dijk (1978) and van Dijk and Kintsch (1983). According to them, three levels of representations of the text are constructed in readers' minds: *surface memory*, *propositional textbase*, and *situation model*. The first level of text comprehension is the surface memory, which refers to the linguistic relationship between the exact words and phrases that readers capture from text. The syntaxes of clauses are decoded and recognized at this level. The second level of text comprehension is the propositional textbase, which is a semantic network of textual propositional units. It reflects the microstructure and macrostructure (i.e., hierarchical structure of a text). The third level of text comprehension is the situation model, which reflects the product of the interaction between textual information and world knowledge of readers. Thus, understanding a text requires a representation of the meaning of the text, which adds to the literal meaning of the text by incorporating relevant world knowledge. Among these representations, the propositional textbase and situation model play an important role in building the meaning of a text. Fincher-Kiefer (1993, 1996) claimed that inferences are not processed in the surface memory but are derived based on the propositional textbase and situation model.

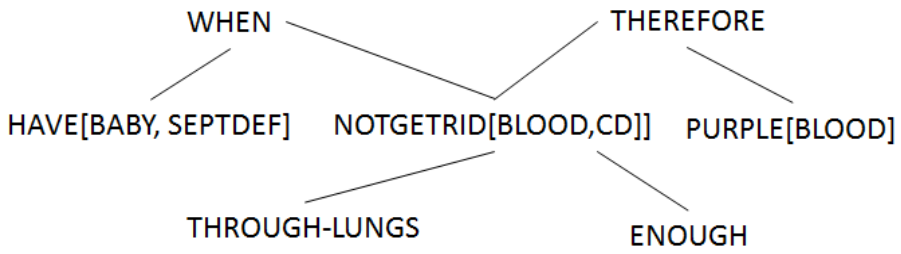
Figure 2.1 shows two sentences taken from a text on heart disease, which their textbase

and situation model. The textbase consists of three propositions (i.e., HAVE [BABY, SEPTAL-DEFECT], NOTGETRID [BLOOD, CARBON DIOXIDE], and PURPLE [BLOOD]) linked by two sentence connectives (i.e., WHEN and THEREFORE). A proposition is defined as the smallest idea unit which can be determined to be true or false.

Text:

When a baby has a septal defect, the blood cannot get rid of enough carbon dioxide through the lungs. Therefore, it looks purple.

Propositional Textbase:



Situation Model:

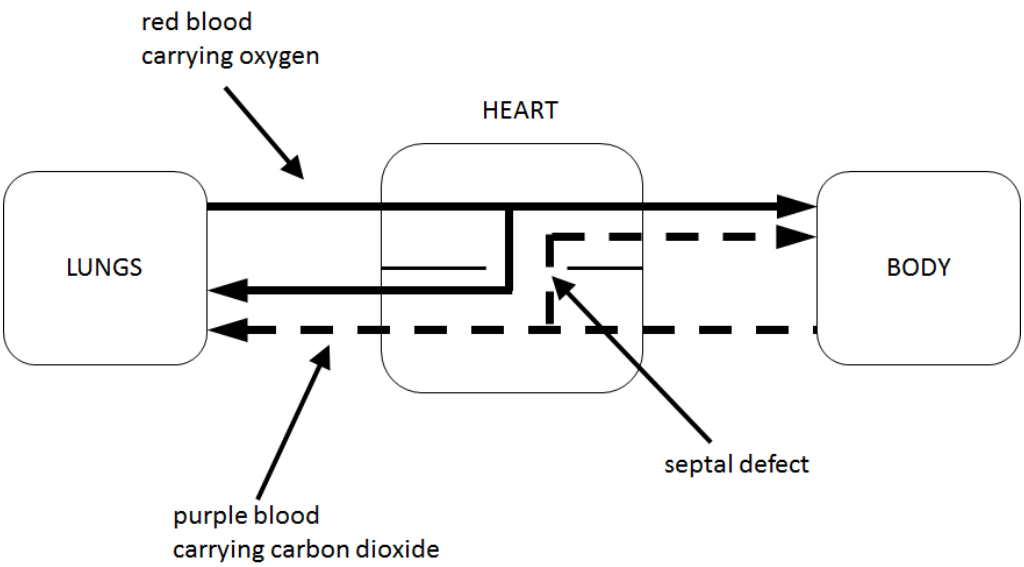


Figure 2.1. The propositional textbase and situation model (adopted from Kintsch, 1994).

While the textbase is regarded as the text itself (i.e., meaning representation of a text), the situation model represents what the text is (i.e., the reader's interpretation of the text) which includes a variety of inferences based on the reader's background knowledge. The situation model in Figure 2.1 is represented as a diagram. It shows that, because of the septal defect, red blood carrying oxygen is mixed together with purple blood carrying carbon dioxide. Therefore, some of the purple blood is therefore recirculated back into the body without picking up oxygen in the lungs. It is a crucial point that much of the information which is not mentioned in the text (e.g., "red blood carries oxygen" or "purple blood carries carbon dioxide") is derived from the reader's knowledge of the circulatory system, rather than derived directly from the text itself. Hence the deeper level of text understanding represented by the situation model.

Furthermore, Kintsch (1988) suggested that the construction of these mental representations is achieved through two major processing phases in construction-integration (CI) model. One of them is the construction phase, which forms the network of the propositional textbase. During this phase, a new text segment (e.g., clause) is processed, and additional knowledge units that are readily available or the relevant new clause are activated. The other is the integration phase, which forms the situation model that reflects the context, readers' world knowledge, and inferential information that can be derived from the textbase. This model posits that large amounts of information are activated automatically in reading and inference processing is deeply connected with the mental representation construction.

2.1.2 Coherence and Text Comprehension

In order to capture the process of text comprehension, *coherence* is also one of the key concepts. In these 20 years, ample empirical evidence has shown that text coherence plays an important role in text comprehension (McNamara, 1995; Linderholm, 2002; O'Reilly &

McNamara, 2007). The notion of text coherence can be defined as “the degree to which text propositions are interconnected in the reader's mental representation of the text” (McNamara & Kintsch, 1996, p. 254). That is, a text that helps the reader to form a coherent mental representation is called a coherent text.

As explained above section, the situation model (as the final product of reading comprehension) is constructed on the basis of integration of textual information and readers' background knowledge. A coherent text can relatively provide a simple and easy structure that allows readers to identify the relationships between ideas in the texts and make necessary connections among them. With a coherent text, readers may not need to make a lot of inferences in order to capture the relationship among ideas. On the other hand, a non-coherent text has many coherence breaks, and readers require background knowledge to fill in the gaps. In order to comprehend such a text, readers need to make additional efforts and to make various types of inferences to repair the breaks.

The coherence relationships of a text are mainly divided into two levels: *local* and *global* coherence. Local coherence is defined for pairwise relationships between sentences of a textual sequence. This level of coherence is achieved when a sentence can be connected to previous sentences in readers’ working memory (i.e., a person’s ability to store some kind of phonetic or textual information in a small amount of time). Global coherence is defined in terms of whole sets of sentences (e.g., the discourse as a whole). Global coherence is also known, in more familiar terms, as the theme, main idea, upshot or gist of a discourse or a passage of the discourse. This level of coherence is achieved when a sentence can be connected to main ideas of other paragraphs, that is, to an idea that is no longer maintained in working memory.

According to the constructionist theory, readers usually attempt to sustain coherence on both the local and the global level during reading (see Graesser, Millis, & Zwaan 1997; Graesser,

Singer, & Trabasso, 1994). Proficient readers will detect the inconsistencies in the text when there is a control at the local and global level. For example, a passage that starts out by describing ways to prevent global warming but ends up promoting increasing gasoline consumption is contradictory. Normally, a proficient reader would not be able to achieve global coherence of this text because of the control. Similarly, readers will notice incoherence at the local level. For instance, “The waitress dropped a cup. The girl's finger bled badly." Because there is no explicit statement that the girl mentioned in the second sentence is the waitress and that the girl's finger was hurt by the broken cup, the reader has to generate referential and causal inferences to bridge the local coherence break. Thus, it is essential that writers provide sufficient devices in the text to build up the coherence relationships, especially in texts designed for L2 or EFL learners with limited reading ability.

In summation, a coherent text is better organized than a less coherent text, and it is easier to comprehend. Readers are more likely to construct a meaningful and useful mental representation of such a text and achieve better comprehension of both the textbase and the situation model.

2.2 Inference Generation Processes

Through these processes, readers construct a coherent mental representation of the text. However, as stated above, their mental representation is constructed not only from explicitly written information in the text, but also from inference information generated by readers. They use their general knowledge to generate inferences that fill in gaps between what has been explicitly stated and what the fully filled-in message was intended to convey. That is, readers must often infer implicit or unstated information on the basis of their knowledge.

That readers engage in this inferential process is a basic assumption of many reading

comprehension models (Graesser, Singer, & Trabasso, 1994; Keefe & McDaniel, 1993; van Dijk & Kintsch, 1983). In reading processes, if inferences are generated, they promote an understanding of the text that goes beyond the surface meaning, and thus contribute to the construction of a concrete mental representation. Over the past 30 years, many first language (L1) studies have investigated the effect of inference generation on reading comprehension. For example, Cain, Oakhill, Barnes, and Bryant (2001) conducted a study with young native-speaker children and showed a strong relationship between reading comprehension skill and the ability to generate inferences. Primarily, reading comprehension problems have been attributed to deficiencies in their lower-level cognitive skills such as phonological skills, word-decoding skills, and vocabulary knowledge. Grabe (2000) pointed out that "word recognition fluency is critical for reading because readers need to see word forms and access the appropriate meaning both rapidly and accurately." However, Cain et al. found that even the children who had sufficient lower-level cognitive skills had difficulties in text comprehension if they lacked an inference generation facility as a higher-level cognitive skill. That is because readers who had difficulty in reading could integrate individual information in a locally coherent way, but they could not use inferences to grasp the meaning of the entire text as a globally coherent representation. Thus, it was implied that the generation of inferences was the fundamental element in proficient reading, and the better readers are, the more active their use of inference in reading.

Moreover, in the context of research on reading in a second language (L2) or English as a foreign language (EFL), the importance of inference is also highlighted from the perspective of comprehension strategy because the use of appropriate inference contributes to a more effective understanding of the text (e.g., Horiba, 1996, 2000; Yoshida, 2003). However, inference is a high-level cognitive activity; hence, it is a difficult skill for EFL learners to attain,

who are more likely to devote their cognitive resources to lower-level cognitive processing compared with L1 readers.

2.2.1 Classification of Inference Types

Most of these inferences during reading can be classified into two broad categories: *necessary* and *elaborative inferences* (Graesser et al., 1994). Firstly, necessary inferences, also known as *bridging* or *backward inferences*, are those inferences that are generated to maintain consistency in local coherence level (i.e., between adjacent units) and global coherence level (i.e., between nonadjacent units). Both cognitive theories of reading comprehension, constructionist theory (Graesser et al. 1994) and minimalist hypothesis (McKoon & Ratcliff, 1992) claimed that necessary inferences are always generated in reading because they are required to achieve full comprehension and related to the process of constructing meaning networks. Also necessary inferences include the following types of inferences: referential, case structure role assignment, causal antecedent, superordinate goal, and character's emotional reaction (Graesser et al., 1994).

Secondly, elaborative inferences, also known as *forward inferences*, which are generated to add a new concept and detailed information into text mental representation and embellish what has been explicitly stated in texts. Elaborative inferences help in creating a more enhanced image of the sentence by integrating it to the textbase as a mental representation (Fincher-Kiefer, 1996). In contrast to necessary inferences, it is stated that elaborative inferences are not always generated during reading because the information generated by the inference is not needed to construct a mental representation. Elaborative inferences include the following the types of inferences: causal consequence, instantiation of noun category, instrument, subordinate goal-action and state (Grasser et al., 1994).

Inferences in global coherence level are made when readers attempt to organize text information into higher order chunks. These inferences include thematic inferences, although it has not been found yet whether it is generated online. Elaborative inferences belong to the second category. Because their generation allows readers to go beyond what has been explicitly stated in the text, they are considered unnecessary for comprehension. They are assumed to be generated online only when a strong constraining context precedes the focal statement. Instantiation of noun category and causal consequence belong to this category.

Among these elaborative inferences, causal consequence is especially focused on in the current thesis. Causal consequence is also called *predictive inferences*. These inferences anticipate the subsequent information that is likely to occur in the text, characterized as "what-happens-next" inferences (Allbritton, 2004; Iseki, 2003). They allow readers to go beyond the text and presumably construct a richer understanding of a narrative (Allbritton, 2004). For example, if we read the following story, "The large ship was cruising around Greenland. The captain was in his cabin. It was a pitch-black night. Suddenly the ship ran into a gigantic iceberg and caused a terrible noise." Reading up to this point, we are likely to generate the inference that "the ship is going to sink" as an upcoming event. That is because we unconsciously use the general knowledge that the damage caused by a collision with a gigantic iceberg causes a ship to sink. Such inferences about what will happen next are called *predictive inferences*. These inferences are important in reading comprehension for the following three reasons. First, to generate predictive inferences, readers need to connect text events with general world knowledge and it allows readers to predict what is likely to occur next. This connection is essential to building situation models of texts (Fincher-Kiefer, 1996; Linderholm, 2002). Second, the generation of predictive inferences helps to activate the processing of future text events and enables future text to be processed more smoothly. Third, there are some classes of

predictive inferences that must be generated to maintain text coherence during reading (Klin, Murray, Leviene, & Guzman, 1999; Murray, Klin, Myers, 1993). One of the controversial issues regarding predictive inferences is whether they are generated during reading or not, and many studies have found that the online generation of predictive inferences depends on the textual characteristics, readers' background knowledge about the text, and their reading goals or strategies (Campion, 2004; Fincher-Kiefer, 1996; Klin, Murray, Levine, & Guzman, 1999; Lassonde & O'Brien, 2009; Linderholm, 2002; Murray & Burke, 2003).

On the other hand, in the field of L2 reading research, although predictive inferences have been dealt with as one of the reading strategies (Minaminosono, 1997), or one of the various inferences that the learner generates during reading (Horiba, 1996, 2000, Muramoto, 2000; Yoshida, 2003), there is little research that focus solely on the generation of predictive inference online. Therefore, this research has an interest in the generation process of predictive inferences in EFL learners' reading comprehension.

In sum, research on L1 reading comprehension has shown that necessary inferences which are always generated online during reading as it concerns the understanding of the text consistency. On the other hand, elaborative inferences are not always required to understand the consistency; thus, they are not always generated during reading (Graesser et al., 1994). So far, the classification of various types of inferences has been described. The following section introduces some reading comprehension models, and reviews how these inferences are organized in each reading model.

2.2.2 Theories and Reading Models of Reading Comprehension

In this section, I will provide an overview of the five reading comprehension models concerning the L1 and L2 studies on inference generation processes: the constructionist theory

(Graesser & Kreuz, 1993; Graesser et al., 1994), the minimalist hypothesis (McKoon & Ratcliff, 1992), the framework of van den Broek, Flecher, and Risden (1993), memory-based processing view (also referred to as the resonance model; Myers & O'Brien, 1998; O'Brien & Myers, 1999), and the event indexing model (Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995; Magliano, Zwaan, & Graesser, 1999). The main purpose of this section is to introduce the framework for interpreting Japanese EFL learners' comprehension processes.

2.2.2.1 Constructionist Theory

The constructionist theory proposed by Graesser et al. (1994) accounts for many empirical facts about the construction of situation models during reading comprehension. This theory embraces the search-after-meaning principle that has three critical assumptions: (a) *the reader goal assumption* represents that the reader constructs a meaning representation according to the reader's goals. These goals and meaning representations are normally set at deep levels of processing such as meaning identification and situation models rather than at shallow levels such as word recognition and syntax; (b) *the coherence assumption* represents that the reader attempts to construct a meaning representation which is coherent at both local and global levels. Local coherence refers to the causal relationship within small portions of texts, usually not longer than a paragraph. Meanwhile, global coherence is established when local chunks of information are interrelated into higher order chunks to achieve coherence between paragraphs; (c) *the explanation assumption* represents that the reader attempts to explain why actions, events, and states are mentioned in the text. Comprehension is achieved by "why questions" to a much greater extent than other types of questions (when, where, how, what-happens-next questions). Furthermore, these explanations of "why questions" involve knowledge structure about psychological and physical causality to construct coherent mental

representation.

In addition, following the three assumptions stated above, Graesser et al. (1994) categorize what kind of inferences will be generated in reading comprehension (see Table 2.1).

Table 2.1
Classification of Different Types of Inferences

Inference Type	Description	Classification	Source
(1) Anaphora	Forming a link between two terms referring to the same thing: e.g., "The car came racing round the corner. Everybody scattered as the *vehicle* crashed into the wall "	Cohesive	Text-based
(2) Pronoun resolution	Linking a pronoun to its previous referent: e.g., "*John* picked up Mary's book. *He* had wanted to read it for ages.	Cohesive	Grammatical knowledge
(3) Case-structure role assignment	Assigning the role of agent, object, recipient, time, or location to a noun phrase: e.g., "The elephant (agent) gave his bananas (object) to the monk (recipient)."	Cohesive	Grammatical knowledge
(4) Causal antecedent	Providing an explanation for the actions and events in a text: e.g., "The campfire started to burn uncontrollably. Tom grabbed a bucket of water."- inference: Tom grabbed the water to put out the fire.	Local coherence	Knowledge-based
(5) Superordinate goal	The overall goal that motivates the characters in the text. If the following sentence is encountered at the beginning of the story "It was Sam's mum's birthday and Sam wanted to by her a present," then the superordinate goal, *Sam wanted to buy his mum a present*, would be inferred if the following sentence was encountered at a later point in the story "Sam woke early and went to the shops to find something special."	Global coherence	Knowledge-based

Inference Type	Description	Classification	Source
(6) Thematic inference	The overall goal of moral of the passage: e.g., never play with fire.	Global coherence	Knowledge-based
(7) Character emotional reaction	The reactions of a character to the actions and events in the text: e.g., Sum gave his mum a lovely a present" would lead to the inference that his mum was really pleased.	Global coherence	Knowledge-based
(8) Causal consequence	The predicted consequences of the actions and events in the text: e.g., "The dragon turned towards the knight and let out a fiery roar" might lead to the inference that the knight was wounded by dragon.	Elaborative	Knowledge-based
(9) Instantiation of noun category	Elaboration of a specific exemplar from a generic noun, i.e., "fish" becomes "shark" after reading "the fish attacked the swimmer."	Elaborative	Text-based
(10) Instrument inference	Inferring a particular object used by an agent to complete an action, i.e., inferring the knight used a sword from the sentence "the knight lunged at the dragon and pierced his shiny scales."	Elaborative	Knowledge-based
(11) State inference	Static properties of objects, characters, etc., not related to the causal structure of the text: e.g., a dog has a tail.	Elaborative	Knowledge-based
(12) Subordinate goal action	How an agent achieved a goal not relevant to the superordinate goal of the text: e.g., you might infer that Sam took the bus to the shops to buy his mum a present.	Elaborative	Knowledge-based

Note. The classification is adapted from Graesser, Singer, & Trabasso (1994, p. 375) with minor revision.

According this Table 2.1, *anaphora* (class 1), *pronoun resolution* (class 2), *case structure assignment* (class 3), and causal antecedent (class 4) are made to establish the cohesive and

local coherence of the text. By contrast, *superordinate goal* (class 5), and *thematic inferences* (class 6) *character emotional reaction* (class 7) are needed to establish global coherence. In addition, classes 8 through 12 are elaborative inferences that are not needed for establishing coherent explanatory meaning representations. Therefore, they supposed that inferences relevant to global coherence are generated automatically during reading, whereas elaborative inferences are not generated during reading without reader's strategic processing.

From the viewpoint of the source of the inference, eight of 12 inference types (class 4-8 and 10-12) are generated on the basis of readers' knowledge as a source, whereas two types of inferences (class 1 and 9) are generated on based on the text and the other two types of inferences are generated on grammatical knowledge (class 2 and 3). This classification is beneficial for understanding what kind of inferences can be generated by readers in reading comprehension. Some researchers have used this framework to analyze readers' inference activity in the L2 field (e.g., Horiba, 1996, 2000; Muramoto, 2000; Yoshida, 2003).

In sum, the point of this reading model is that readers strategically generate inferences through a controlled intentional search of background knowledge to establish text coherence at both local (adjacent ideas) and global (distant ideas) level and to explain why various ideas (e.g., events, actions and states) are mentioned in the text. Many findings from prior research showed that readers more frequently and more strongly generate bridging inferences, which contribute to the understanding of a text, than elaborative inferences, which are less necessary for text comprehension (e.g., Long & Golding, 1993; Long, Golding, & Graesser, 1992; Magliano, Bagget, Johnson, & Graesser, 1993). It has been take as an indication that readers generate inferences in a strategic manner to facilitate comprehension. Likewise, the constructionist model is supported by the empirical findings that different reading goals (e.g., reading for entertainment or reading for study) evoke different patterns of inference generation (e.g.,

Linderholm, 2002; Lorch, & O'Brien, 1995) and that reading interventions, which target active strategic generation of bridging inferences, improve readers' comprehension (e.g., McNamara, 2004; McNamara, O'Relly, Best, & Ozuru, 2006).

2.2.2.2 Minimalist Hypothesis

The minimalist hypothesis was proposed by McKoon and Ratcliff (1992). They state that readers automatically generate only those inferences necessary for establishing local coherence or based on readily available information. The greatest difference between the constructionist theory and the minimalist hypothesis is that the constructionist theory claims that inferences are activated regarding a global coherence in on-line reading. Long and Lea (2005) defined local coherence and global coherence as follows: local coherence refers to the knowledge and processes involved in forming conceptual connections among short sequences of clause. Local coherence is often achieved by means of referential cohesion, the understanding that two or more text elements refer to the same entity. In contrast, global coherence refers to the knowledge and processes involved in connecting incoming information with one or more overarching themes. According to the constructionist theory, global coherence often involves finding explanations for the actions, events, and states that are explicitly mentioned in a text.

Table 2.2 shows a summary of prediction of on-line inferences from the constructionist theory and the minimalist hypothesis. Both the constructionist theory (Graesser et al. 1994) and the minimalist hypothesis (Mckoon & Ratcliff, 1992) agree that predictive inference are unlikely occur on-line. The reason is that they are not necessary to establish coherence in texts. However, both models admit that predictive inferences are generated on-line if they are highly constrained by contexts, that is, when they have sufficient supporting contexts. (Fincher-Kiefer, 1996; Klin et al., 1999). However, the above two models are not sufficient enough to explain

why predictive inferences are generated when there are strong contextual constraints.

Table 2.2 *Predictions of On-Line Inference Processing* (Graesser et al., 1994, p. 384)

	Types of Inference	Constructionist Theory	Minimalist Hypothesis
(1)	Anaphora	X	X
(2)	Pronoun resolution	X	X
(3)	Case-structure role assignment	X	X
(4)	Causal antecedent	X	X
(5)	Superordinate goal	X	
(6)	Thematic inference	X	
(7)	Character emotional reaction		
(8)	Causal consequence		
(9)	Instantiation of noun category		
(10)	Instrument inference		
(11)	State inference		
(12)	Subordinate goal action		

Note. X = On-line prediction.

2.2.2.3 Framework of van den Broek, Flecher, and Risden (1993)

Although many researchers classify inferences by focusing on their functions, van den Broek et al. (1993) approached the problem in terms of the source of information from which an inference is drawn, and the direction of inferences related to the focal statement. Figure 2.2 illustrates types of inferences in reading on the basis of these two perspectives. van den Broek et al. assumed three resources of information (i.e., text in short-term memory, text in long-term memory, and background knowledge) and three direction of inferences (i.e., backward, forward, and orthogonal).

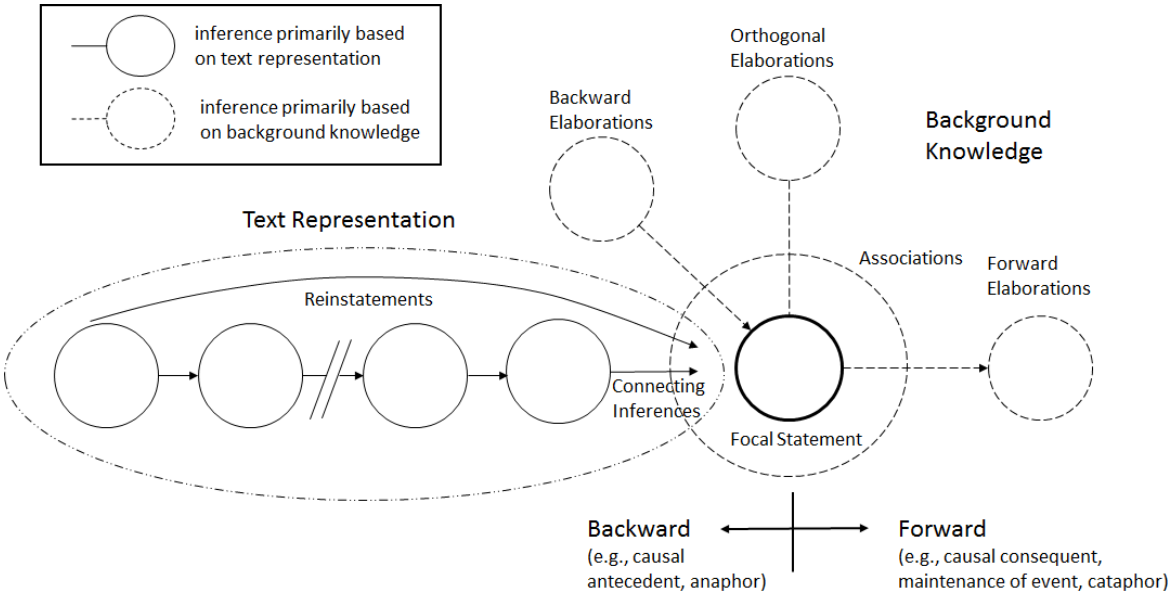


Figure 2.2. Types of inferences in reading (adopted from van den Broek et al., 1993, p. 171).

Then, they divided the inferences into the following six categories: (a) connecting inferences, (b) reinstatements, (c) backward elaborations, (d) forward elaborations, (e) orthogonal elaborations, and (f) associative inferences.

First, inferences (a), (b) and (c) are generated in backward direction, that is, they connect the statement that currently is being read (focal statement) to events or states that have occurred earlier, thereby providing coherence to the readers' representation of text. (a) *connecting inferences* connect the focal statement most directly to the text information that was processed recently processed in short-term memory, (b) *reinstatements* connect the focal statement to the text information that is reinstated in long-term memory. (c) *backward elaborations* draw heavily on the readers' general background knowledge in connecting the focal statement.

Second, (d) *forward elaborations* (i.e., predictive inferences) anticipate information that is yet to be described in the text (Graesser & Clark, 1985; McKoon & Ratcliff, 1986; van den Broek et al., 1993). Forward elaborations predict the occurrence of new events or future

relevance of text information. As mentioned before in the section 2.2.1, many researchers have found evidence for the frequent generation of backward inferences (Klin & Myers, 1993; Lear, & Andrusiak, 1992; McKoon & Ratcliff, 1992; Singer, Halldorson,), there is much more debate about the spontaneous occurrence of predictive or forward inferences (see Keefe & McDaniel, 1993; Magliano, Baggett, Johnson, & Graesser, 1993; Potts, Keenan, & Golding, 1988).

Third, (e) *orthogonal elaborations* involves the activation of information that is implied by and coexistent with the information in the focal statement. Orthogonal inferences may take several forms, but a particularly interesting and well-investigated type is the inference of spatial and visual information. Also, (f) *associative inferences* activate information that is associated with a focal statement. These inferences are assumed to be made automatically through a mechanism of spreading activation (Kintsch, 1988).

Table 2.3
Summary of van den Broek et al.'s (1993) Inference Types

Type	Source	Direction
Reinstatements	text representation	backward
Connecting Inferences	text representation	backward
Backward Elaborations	background knowledge	backward
Forward Elaborations	background knowledge	forward
Orthogonal Elaborations	background knowledge	orthogonal
Associative	background knowledge	orthogonal

Note. This table is summarized by the author based on van den Broek et al. (1993); Source = the source of information from which an inference is drawn; Directions = the direction of inferences related to the focal statement.

This framework is profitable in the point of showing the on-line processes generated two or more inferences through reading comprehension. This framework is partially used by Horiba (1996, 2000) in the L2 field. The present study also use it to figure out the EFL learners' reading

process how various inferences are generated by linking text information with their background knowledge.

2.2.2.4 Memory-Based Processing View (Resonance Model)

In contrast to the constructionist theory, *memory-based processing view* (resonance model) holds that inferences are activated automatically due to the spread of activation from textual ideas that are currently active in working memory to associated ideas in background knowledge via long-term memory networks (e.g., Myers & O'Brien, 1998; O'Brien & Myers, 1999; van den Broek, Rapp, & Kendou, 2005). In favor of this view, a large body of research has shown that elaborative inferences are more reliably activated when they constitute strong semantic associations with specific textual ideas (e.g., Calvo, Castillo, & Estebez, 1999; keefe & McDaniel, 1993). According to this model, when new text ideas are encoded into working memory, a signal is sent from these text ideas to all of long-term memory. Relevant information in long-term memory that share features in common with the contents of working memory will "resonate" in response both associatively and strategically. After that those information that resonate the most are also most likely to be incorporated into working memory.

For example, when reading about a delicate porcelain vase that is being thrown against a hard surface, readers routinely infer the strongly associated outcome of the vase breaking, although this predictive inferences is not necessary for understanding the inference-evoking information. More evidence supporting the memory-based view comes from studies that have demonstrated the contributions of *textual elaboration* (e.g., Rizzella & O'Brien, 1996), *textual cues* (e.g., Albrecht & Myers, 1995) and *feature overlaps* of textual ideas (e.g., Duffy, Henderson, & Morris, 2003) to the reactivation of prior textual ideas in working memory. Especially, the intensity of this resonance is decided by feature overlap. To put it in other terms,

this refers to how much information with similar characteristics exists in long-term memory. Thus, memory-based processing view focus primarily on the bottom-up components involved in reading and constructionist theory view take into account top-down components. From recent studies, it was clarified the framework of two views can explains both the associative and strategic reactivation of episodic and semantic information. That is, the framework allows for the possibility that early stages of reading are controlled by memory-based processes, whereas later stages of reading are controlled by explanation-based processes.

2.2.2.5 Event Indexing Model

The event indexing model proposed by Zwaan, Langston, and Graesser (1995) describes how readers construct a coherent representation of the story structure. Narrative events are understood as being causally linked with the manner in which we understand events in the real world. This model assume that the main purpose of constructing a situation model is to monitor what character are involved in a story, what is happening to them, what their goals are, and how they are achieving those goals, all within a narrative time and space. As each story event and action is comprehended, readers monitor changes in continuity in multiple levels of dimensions: characters, objects, temporality, spatiality, causality, and intentionality. Changes along these dimensions cue the mental representation for a story must be updated.

It is a crucial point that, before this model is advocated, most of research on situation model construction has only focused on one dimension of a situation model. However, Zwaan et al. (1995) have investigated the extent to which readers construct multidimensional situation models. In this model, when a focal sentence describes an event that is continuous with respect to these situational five dimensions, there is high correspondence between the discourse structure and the story structure. However, discontinuities along these dimensions often occur

when there are discrepancies between the discourse and story structures. It is under these conditions that readers must engage in effortful processing to construct a coherent situation model of the story structure.

As mentioned above five reading comprehension models, for many researchers in cognitive psychology and discourse processing, the central concern regarding the generation of inferences over past decades has been which types of inferences are normally generated online (e.g., Graesser et al., 1994; Kintsch, 1994; McKoon & Ratcliff, 1992). *An online inference* refers to an inference that is generated during reading comprehension, whereas an *off-line inference* is constructed in a later retrieval task (Zhang & Hoosain, 2005). The identification of online status distinguishes whether a cognitive process occurs in the comprehension process or memory representation. Therefore, the identification of online inferences is very important for understanding mental representation and its processes. The above five reading models mainly explain what inference is to be generated online, and what kind of textual information and readers' knowledge are activated for constructing the mental representation of the text.

2.2.2.6 The Recent Integrated Model

In recent years, the constructionist theory (see 2.2.2.1) and the memory-based processing view (see 2.2.2.4) hold a prominent position within the research literature on text processing. These two views are often presented as conflicting with one another. Both views acknowledge the role that general world knowledge plays in the comprehension process, although constructionist model have more appealed to the influence that general world knowledge has on comprehension. According to constructionist theory, both automatic and controlled strategic processes interact to influence on-line processing. When a new information encoded, readers search long-term memory for any meaningful information that is relevant to the concepts in

working memory to develop a rich and complete representation of the situation described by the text (Yeari & van den Broek, 2015).

The evidence for the involvement of both associative processes and strategic processes in the generation of knowledge-based inferences has led to a growing consensus that both types of processes are involved in inference generation. Even the constructionist and the memory-based views considered the possibility that under specific circumstances inferences may be generated via associative (e.g., Graesser et al., 1994, p. 377) or strategic processes (Myers & O'Brien, 1998, p. 152), respectively. As a result, researchers have proposed integrative models that examine the coexistence and interaction of associative and strategic processes (e.g., Gerrig, & O'Brien, 2005; Long & Lea, 2005; van den Broek et al., 2005). Yeari and van den Broek (2015) explained about this two-phase (i.e., associative and strategic) process as follows: "It has been suggested that, in the first phase, associative processes are the default primary processes that take place continuously during reading without control or effort on the part of the reader, whereas, in the second phase, strategic processes are initiated by the reader when the products of the associative processes are not sufficient to attain the desired coherence and comprehension (e.g., Bohn-Gettler, Rapp, van den Broek, Kendeou, & White, 2011; van den Broek et al., 2001; van den Broek et al., 2005) (p.1194)."

2.3 Knowledge Structure and Inference Generation

The purpose of this section is to explain about several knowledge-based inferences. Even among the inferences, what types of knowledge-based inferences are generated during reading will be overviewed based on the inference categorization from some famous reading models (e.g., the constructionist theory, the minimalist hypothesis). Furthermore, the findings in L2 research related to the inferential process will be reviewed.

2.3.1 Knowledge-Based Inferences

Inferences may also be classified in terms of the source of the information used to generate the inference. Sources for inferences include the text and readers' prior knowledge. Trabasso & Magliano (1996) and van den Broek, Young, Tzeng, and Linderholm (1999) further differentiate text-based sources into two categories: text ideas from the immediately preceding comprehension cycle that are carried over and text ideas from previous cycles that are retrieved from the reader's mental representation. Prior knowledge sources include readers' use of world, domain, and topic knowledge and include carryover of knowledge activated in the preceding cycle (Trabasso & Magliano, 1996).

Although these dimensions are typically used to describe inferences made during reading, similar qualities are used to describe inferences constructed after reading as well. Specifically, inferences made after reading are usually classified in terms of type and source (McNamara & Kintsch, 1996). This includes distinctions between elaborative and coherence inferences (Cain, Oakhill, Barnes & Bryant, 2001), knowledge-based and text-based (McNamara & Kintsch, 1996), and global and local (Ozuru, Best, Bell, Witherspoon, & McNamara, 2007). The majority of research on readers' inferences after reading has focused on the construction of bridging inferences. These inferences are often referred to as bridges because they require readers to establish a meaningful relationship between two or more text propositions, using ideas from the text and world knowledge (Magliano et al., 1999; McNamara & Kintsch, 1996). They are similar to explanatory inferences, but vary in terms of their necessity for comprehension.

Knowledge-based inferences are essential to the reading comprehension process, since reading is much involved with knowledge structure, and the reader's mental representation is constructed owing to their knowledge. While reading a text, the reader makes knowledge-based

inferences on the basis of their contextual knowledge (information about things that have been mentioned earlier in a text) and general knowledge (information concerning objects, actions, events, and abstract concepts attained from daily experiences, or relevant information about the text they are reading) (Graesser & Kreuz 1993). Since they are based on both contextual and general knowledge, the quality and the number of knowledge-based inferences completely depend on the quality of the reader's text representation made while reading and the structure and the amount of the reader's general knowledge. Of course, the latter is crucial because the reader will not be able to understand the context and fail to construct a coherent mental representation without any general knowledge.

Researchers have different viewpoints about how this general knowledge is represented in the reader's mind. Some researchers claim that knowledge-based inferences are inherited (i.e., activated and copied) from readers' general knowledge structures during text comprehension (e.g., Rumelhart & Ortony 1977; Schank & Abelson 1977). On the other hand, many other researchers adopt the standard assumptions of the so-called *the constructionist theory* of discourse comprehension (e.g. Graesser & Clark 1985; Graesser, et al., 1994; Hannon & Daneman, 1998; Lorch & O'Brien 1995; Shears & Ciarello, 2004). The constructionist theory assumes that when the readers read a text, they create its mental representation. Simultaneously, they compare the mental representation to knowledge structure of everyday experiences in their world. Such knowledge structure is called *schemata* (Bartlett, 1932; Rumelhart & Ortony, 1977; Rumelhart, 1980). In order to achieve this comparison between mental representation and knowledge structure, readers need to make a number of knowledge-based inferences that connect these pieces of knowledge to the situation model that the readers construct while reading a text (Graesser & Kreuz 1993). In other words, in order to make knowledge-based inferences readers must associate what they have read t with what they already know and

remember from their daily experiences or from other knowledge that they have obtained from the text.

According to the explanation assumption in the constructionist theory (see Section 2.2.2.1 in detail), knowledge-based inferences are most generated when the readers respond to "why question" which they have at subconscious level. Graesser and Wiemer-Hastings (1999) have given the explanation as shown in the following sample scenario (Table 2.4) that the readers have the potentiality to generate knowledge-based inferences when reading the scenarios. This Table is very informative to capture what sort of thinking processes are going through when readers generate several inferences.

Table 2.4
Sample Scenario and Potential Knowledge-Based Inferences

Sample Scenario
Two brothers were always in competition over money and success. George unfortunately went bankrupt the same year that Bill made his first million. Bill threw a New Year's Eve party that ended up being quite a festive occasion. But everyone sobered up very quickly when George was caught kissing Bill's wife.
A Set of Potential Inferences
(1) "George wanted to hurt Bill." This is a superordinate goal of the act of George kissing Bill's wife. Superordinate goals explain "why" acts occur, so this inference would have a high likelihood of being construction of being constructed in the situation model.
(2) "George touched the cheek of Bill's wife." This elaborates the subplan (i.e., subgoal, procedure, method, style) of the act of George kissing Bill's wife. This information specifies "how" the kissing occurred, but does not explain "why" it occurred, so it would have a low likelihood of being constructed in the situation model.

(3)	"George was jealous of Bill." This is an important causal antecedent that leads to the kiss. George's bankruptcy and Bill's financial success caused George to be jealous of Bill, which in turn triggered the goal of George wanting to hurt Bill. This chain of causal antecedents to the kiss explains "why" the kiss occurred, so it would be constructed in the situation model.
(4)	"Bill slugged George." This is a plausible causal antecedent (expectaion) of George's kiss. It answers a what-happen-next question, but not a why question, so it is probably not generated on-line in the situation model.
(5)	"George was in the bedroom." This is a plausible spatial setting for a kiss. But it fails to explain the kiss, so it is not likely to be constructed in the situation model.
(6)	"George's kiss was a romantic kiss." The romantic kiss inference is constructed on-line because it explains why the author used the words "but" and "was caught", and "why" the author bothered mentioning that the guests quickly sobered up.

Note. From Graesser and Wiemer-Hastings (1999) with minor revision.

2.3.2 Goal-Based / Situation-Based Inferences

In contrast with the former classifications of inferences, some researchers believe that inference generation during text comprehension is mainly driven by not knowledge but goal. Their opinions are based on the fact that much information in short-term memory which the readers need to assess as relevant information in a limited time appears to be *goal-based information* (e.g. O'Brien, Rizzella, Albrecht, & Halleran, 1998; Lorch & O'Brien 1995). This corresponds to the protagonist's goals (or more generally, any characters' goals). Consequently, the reader searches for causal relationships in a text with the protagonist's or other characters' goals. Then, they make inferences in order to establish the connections between events that take place and the protagonist's goals.

To illustrate, in the example of "John in the library", a reader may have inferred that John's goal in going to the library is to borrow a book. When a reader comes across the sentence "John put it into his bag and went home", he/she may infer that 'putting it into his bag' refers to the

fact that he borrowed the book from the library. Thus, with respect to this specific goal ("borrowing a book"), the reader will assess the sentence that John's borrowing the book is more relevant than the other sentences. In this way, most inferences can be reduced to the characters' acts that take place and with that, the goals that these characters pursue in a text.

So far, in terms of how inferences are driven, two types of inference classifications have been introduced: knowledge-based inferences, which are mostly text-based (i.e. the reader generate an inference from his/her general knowledge on the basis of a piece of text), and goal-based inferences, which are based on characters' goals and motives. In addition to the above inference classifications, some researchers claim the *situations-based inference* (e.g., Zwaan & Van Oostendorp, 1993) as a point of departure for making inferences during text comprehension. They argue that readers primarily draw inferences from situations which they recognize in a text and they fill up these situations with the reader's ideas and knowledge about the situations (places, objects and actions).

Whereas knowledge-based inferences often seem to be "backward directed", that is, the reader draws inferences at a later point to connect subsequent sentences into a coherent text representation (McKoon & Ratcliff, 1986), situation-based inferences are rather "forward-directed", as they predict situations. For example, when the reader reads a sentence like 'Peter is at the swimming pool', he/she will link this to their general knowledge about this spatial situation (i.e. being at the swimming pool) and its causal relationships (e.g., starting to swim or splashing water). According to the event-indexing model proposed by Zwaan and his colleagues (Zwaan, Langston, & Graesser, 1995; Zwaan & Ratvansky, 1998), during reading comprehension, the reader connects the text events into five different situational dimensions: time, space, causation, motivation, and protagonist. Among these situational dimensions, the components which are not written explicitly are supplemented with the situational-based

inferences.

To sum up, it is crucial to understand the causal relationship between text's events and objects which is drawn from the character's goal information in the narrative comprehension (van den Broek & Gustafson, 1999). Therefore, the idea of goal-based inference gives a good prediction for the inference generation process. Also, situation-based inferences explain the reader's predictive reading processes. This will enable the readers to easily foresee what is going to happen in the story because he/she looks for familiarities with the type of situations and knows what to expect. However, it appears to be that the reader does not necessarily generate only such efficient types of inferences. For example, the reader often infers to the appropriate instrument for the action mentioned in the text (i.e., instrument inferences; Lassonde, & O'Brien, 2009) that instrument is a supporter in the preceding text. These elaborative inferences could not be explained sufficiently by the framework of goal / situation-based inferences. Meanwhile, the knowledge-based inferences are more-exhaustive because they consider the generation of inferences which are drawn from various text information and knowledge sources. Therefore, in this thesis, I shall mainly use the classification of knowledge-based inferences.

2.4 The Effects of Text Factors

As mentioned before, reading process and inference generation are affected by several text and learner factors. The identification of text and learner factors that determine which processes occur at what moments during reading is primarily focused in the reading research (e.g., Sanford & Graesser, 2006; van den Broek, et al., 1993). C*haracteristics of the text* such as text genre, contextual constraint, and reading units of the text are a major influence on the types of processes that occur. Another major influence consists of general *individual*

characteristics of the learner, such as learners' L2 reading proficiency, working memory, and the interaction of reading purposes and standards of coherence. In this section 2.4 and the next section 2.5, the author introduce text and learner factors affecting EFL learners' reading process and inference generation.

2.4.1 Text Genre

In this thesis, my primary focus is on narrative text rather than on other text genres such as expository texts, descriptive texts, and essays. The reason is because, narrative text has a close correspondence to everyday experiences in contextually specific situations (Schank, 1986). Both narrative texts and everyday experiences involve people performing action in pursuit of goals, the occurrence of obstacles to goals, and emotional reactions to events. Knowledge about these actions, goals events and emotions are deeply embedded in our perceptual and social experiences (Graesser, Singer, & Trabasso, 1994). The inferencing process and knowledge structure that are used during the comprehension of everyday experiences are also likely to be used during the comprehension of narratives.

In contrast to narrative text, expository text is normally written to inform the reader about new information, concepts, and facts. The typical reader does not have extensive general knowledge about the topics in expository texts, so the readers generate fewer inferences than they generate during the comprehension of narrative text. Therefore, narrative text is an important genre to investigate in terms of inference generation and mental representation construction based on knowledge structures.

Although many studies have suggested that an inferential skill for text integration plays an important role in L2 reading (Carrell, 1984a; Horiba, 1993, 1996, 2000; Yoshida, 2003), it has also been pointed out that the cause of problems in inference generation in L2 reading

processes cannot simply be attributed to L2 proficiency alone and some other unknown factors are involved.

2.4.2 Contextual Constraint

In the field of L1 inference research, many reading models argue that various types of inferences are generated automatically during reading (e.g., Graesser et al., 1994; McKoon & Ratcliff, 1986). These inferences are called necessary inferences, and they are not consciously recognized by the readers because they are generated through the automatic and effortless processes. However, there are also elaborative inferences, which do not always occur during reading, but are generated after reading through the activation of knowledge given by clues such as text comprehension questions and recall tasks. elaborative inferences require a certain conscious level of work on the part of the reader regarding the contents that have been read; that is, elaborative inferences are generated through the strategic and time-consuming processes (Peracchi & O'Brien, 2004). The activation of necessary or elaborative inferences depends on whether the inference itself is necessary to establish coherence in the text. For example, if information about the protagonist is in the preceding context, readers are necessary to connect the information to the following pronoun refers to the same character (i.e., anaphoric inferences), so as to maintain coherence between the already processed information and the information currently being processed. Thus, the types of inference that require established coherence during reading are activated on-line. However, inferences about what type of instrument used to perform action by characters in the text (i.e., instrumental inferences) will not be activated on-line because they do not provide information that is directly connected to maintain coherence of the text (McKoon & Ratcliff, 1986).

As mentioned above, predictive inferences are categorized into elaborative inferences

that anticipate likely events or elaborations about "what will happen next." It is controversial whether these inferences are activated during reading. Predictive inferences are also assumed to make an important contribution to comprehension because they enable us to anticipate forthcoming information. Thus, they both facilitate the processing of new information in texts as well as when drawing an elaborative image about the text. On the other hand, activation of such prediction is considered "optional" for comprehension. Predictive inferences do not necessarily contribute to maintain coherence because of a quality that predicts future developments. As such, many earlier studies have claimed that predictive inferences are less likely to activate on-line (e.g., Duffy, 1986; Singer & Ferreira, 1983). However, in a landmark study of predictive inferences, McKoon and Ratcliff (1986) suggested that with greater contextual support, a larger proportion of features would have been activated with a corresponding increase in the level of activation and encoding of the inference. For example, they used the following scenario: "The director and the cameraman were ready to shoot close-ups when suddenly the actress fell from the 14th story building." Responses to recognition probes reflecting the inference concept (e.g., dead) immediately after the sentence ended were slower and less accurate than in a control condition. On the basis of the result, McKoon and Ratcliff argued that high support context activate the predictive inference concept. Such contextual support is also called *contextual constraint* (hereafter CC). The subsequent L1 studies have argued that under certain conditions, there is a possibility of on-line activation and also suggested the possibility that predictive inferences are more fully activated under conditions where CC is high (e.g., Calvo, Castillo & Estevez 1999; Keefe & McDaniel, 1993; Klin, Murray, Levine, & Guzman, 1999; Murray, Klin, & Myers, 199; Potts, Keenan, & Golding, 1988). In other words, predictive inferences could be generated on-line if the predictable events are readily available in the context, or if they are highly constrained by the context.

The degree of CC in the text also influences the specificity of the activated information. For instances, consider the following text from Sanford and Garrod (2005): "Cathy needs to make a dessert for a party. Cathy carefully follows a recipe and makes a delicious dessert that everyone at the party enjoys." In this case, the reader is left to infer the type of dessert. Cathy made for the party from several possible outcomes (e.g., pie, cake, pudding), each of which could be equally appropriate. However, if the story is further specified by adding the CC sentence that "Cathy makes a special dessert for a birthday party, which everyone at the party enjoys," the reader is likely to infer that she baked a cake, and other outcomes (e.g., pie and pudding) are no longer appropriate alternatives. Lassonde and O'Brien (2009) clarified when the context was constrained, activation of predictive inferences was also more constrained. Thereby, CC has a strong influence on the generation of specific predictive inferences.

The text elements of CC that have been described so far are terms that are frequently used in studies that focused mainly on inference generation. However, the fact that inference generation is affected by the quality of context has been supported by other types of inferences (e.g., causal bridging inferences, instrument inference) as well. For example, some studies demonstrated that the ease of generating bridging inferences was predicted by the degree of *causal relatedness* or the number of *causal connections* (Langston & Trabasso, 1999; Myers, Shinjo, & Duffy, 1987; Shimizu, 2011). In spite of slight variations in the names of factors and their definitions, the following universal statement can be made: activation and non-activation of these inferences is dependent on the amount and strength of the available information supporting the inference, and the conceptual overlap between the supporting information and the inferences.

In the past, the effect of CC on various types of inference generation has been examined. However, there are few studies have taken the amount of text into consideration. It is therefore

necessary to comprehensively test the effects of conceptual overlaps between individual words and the inferred context, and differences in causal relatedness between propositions on the sentence-level, text-level, and between-texts-level reading. Moreover, few studies have closely investigated the effect of CC and learners' L2 reading proficiency on the generation of knowledge-based inferences (e.g., predictive and thematic inferences) although some L2 or EFL research (e.g., Horiba, 1996, 2000; Yoshida, 2003) has examined inferential processes during reading. Consequently, this research focuses on sentence-level reading in Study 1, text-level reading in Study 2, and between-texts-level reading in Study 3. In Studies 1 and 2, the impact of the degree of CC is examined. In Study 3, the concept of analogy research is incorporated, and the impact of surface similarity that reflects the conceptual overlap between texts and that of structural similarity that reflects causal relatedness on inference formation is examined (see 2.4.3 Reading Units of the Text in the subsequent section for further details on the reading unit items of texts).

2.4.3 Reading Units of the Text

In L1 reading comprehension, native readers actively generate several types of inferences and change their use of reading strategies flexibly according to reading units. As the previous study examined the difference in sentence-level and text-level reading, Till, Mross, & Kintsch. (1988) and Long, Oppy, & Seely (1994, 1997) examined sense selection and knowledge-based inference generation by means of a lexical decision task.[2]

Till et al. (1988) found thematic inferences in sentence-level were not generated until later

[2] In the task, participants were presented four types of targets: (a) context-appropriate associates of homograph primes, (b) context-inappropriate associates of the primes, (c) words related to the topic of the sentence, (d) words unrelated to the topic of the sentence, and (e) nonwords. Then, they were required to judge these targets were words or nonwords.

in processing. Faster response time for appropriate theme words compared to inappropriate ones were observed only in the 1,000ms and 1,500ms SOA conditions. Meanwhile, Long et al. (1994, 1997) investigated individual difference in readers' sentence- and text-level representations. In sentence-level reading, they found no difference in proficient and less proficient readers' ability to generate knowledge-based inferences. However, in text-level reading, they found that only proficient readers generate knowledge-based thematic inferences spontaneously during reading whereas less proficient readers do not, even though both groups had similar knowledge about sentence topics and constructed adequate sentence-level representations.

This conclusion is based on the evidence that proficient readers were slower to reject targets in appropriate-theme targets than in inappropriate-theme targets in their priming recognition task, however, less proficient readers did not show any difference. In addition, proficient readers were slower to reject the appropriate topic words than less proficient readers and they made more errors. This is due to the fact in proficient readers, words that do not actually appear in the sentences are generated in their minds as inferences, thus resulting in the latency in their recognition time.

Why do less proficient readers fail to generate knowledge-based inferences? The straightforward explanation about their failure of inference generation is attributed to their deficits of basic linguistic abilities, such as word decoding or syntactic analysis. Perfetii (1997) has argued that less proficient readers generally have slower and less efficient word decoding skills than do proficient readers. Word recognition speed and accuracy play central roles in comprehension ability. Slow and inefficient word recognition processes inhibit the higher sentence- and text-level processes in text comprehension. However, there are some evidences that less proficient readers' inference problems are not the only cause of deficits in basic linguistic processes. Oakhill and her colleagues have examined the inferential abilities of good

and poor readers with similar word recognition skills (Oakhill, 1993; Oakhill, Yuill, & Parkin 1986). Specifically, they could identify readers who differed in their ability to answer questions about texts, but could not in their ability to read words aloud or to understand the meaning of isolated words. In addition, from the latter research, it was found that the reading comprehension was inhibited due to deficits in the reader's inference skills even if the reader had an enough cognitive ability. (Cain, Oakhill, Barnes, & Bryant, 2001).

Although such findings suggest that inference problems occur in spite of accurate word recognition skills, the possibility still remains that these problems are secondary to deficits in other component reading abilities. Less proficient readers may have accurate word recognition skills, but poor syntactic or semantic skills. Alternatively, they may have accurate but very slow word recognition processes. Thus, word recognition may consume resources needed to execute other component processes. If so, then their failure to make inferences may be due to more general comprehension problems such as limited working memory capacity, slow knowledge access and so on.

2.5 The Effects of Learner Factors

2.5.1 L2 Reading Proficiency

Compared to L1 inference studies, few studies on inference generation have been conducted in L2 reading comprehension. However, as well in L1 reading process, an inferential skill for text integration plays an important role in L2 reading comprehension in order to construct a coherent mental representation. Some significant L2 studies have pointed out the relationship between learners' inference-making skill and L2 reading proficiency. For example, some Japanese L2 researchers investigated the types of inferences are generated during L2 reading using a think-aloud methodology (Horiba, 1993, 1996; Yoshida, 2003). Horiba (1996)

examined the number of inferences constructed by native English speakers learning Japanese as L2 in reading Japanese text. The results demonstrated that L2 leaners with higher L2 proficiency generated inferences more frequently than learners with lower L2 proficiency. These findings were also supported by subsequent studies (Muramoto, 2000; Shimizu, 2006). This study attributed the limited production of inferences among lower L2 proficiency readers to the fact that these readers needed to allocate more cognitive resources to lower level reading processes such as word recognition and syntactic analysis. As a result, lower L2 proficiency readers did not have sufficient cognitive resources to engage in higher level processing such as inference generation.

Furthermore, Yoshida (2003) reported that Japanese university EFL students with high L2 proficiency more frequently produced elaborative inferences (including predictive inferences and thematic inferences) than coherence inferences compared to lower L2 proficiency learners. These observation suggested that less proficient L2 readers used insufficient resources for higher-level processing including the inference generation because they spent many of their cognitive resources on lower-level processing. Consequently, unlike L1 reading comprehension, the lower-level processing strongly influence the higher-level processing in the L2 reading comprehension.

Therefore, this study will examine the interaction between L2 reading proficiency as one of learner factors and inference processes. In this study, the participants' L2 reading proficiency were measured with the same English test through experiments. The test aimed to examine how well readers comprehend sentence and text-level information, with 34 items from reading section of the pre-first, second, pre-second grades of the STEP test (2007). Actual L2 reading proficiency of EFL learners who participated in this study will be described in detail in chapter 6 (section 6.1 p. 174).

2.5.2 Working Memory

Working memory is one of the important factors that affects readers' reading process and inference generation. Working memory refers to the pattern of cognitive neural network activations at any information given moment. As Koda described, "once extracted from print, lexical information must be consolidated into larger, meaningful chunks, such as phrases, sentences and paragraphs. Working memory plays a pivotal role in this critical process" (2005, p.198).

Several previous L1 and L2 studies have found readers' working memory capacity is highly related with the readers' inference generation skill (e.g., Calvo, 2001; Yohida, 2003). Furthermore, Osaka and Osaka (1992) examined the relationship between L1 and L2 working memory capacity among Japanese college students. With English and Japanese versions of reading span tests, they found clearly high correlations between L1 and L2 working memory scores. They interpreted the results that working memory resources in large part are shared across languages, and its capacity therefore, is somewhat independent of linguistic knowledge.

As a whole, it is clear that working memory as a shared cognitive resource is involved in almost every aspect of reading lower-level processing among both in L1 and L2 readers. Thus, there is a high possibility that working memory capacity is a key factor distinguishing proficient and less proficient readers.

2.5.3 Reading Purposes and Standards of Coherence

The fact that readers' specific purpose of reading affects their cognitive processing has been shown from some empirical research (van den Broek, Lorch, Linderholm, & Gustafson, 2001). Generally, readers change their reading process in terms of strategies employed, time spent reading, which affects the recall performance after reading. Also, it has been shown that

readers have various standards of coherence that affect how readers process the text to meet specific goals (van den Broek et al., 2001). *Standards of coherence* refer to "the types and strengths of coherence that the reader aims to maintain during reading" (van den Broek et al., 2001, p.124). These standards vary between individuals as well as within an individual from one reading situation to the next. Also, the strictness of these standards theoretically influences the kinds of cognitive processes readers engage in during reading, and these standards vary due to reading purpose, text difficulty, or the motivation level of the student.

For instance, students may have stricter standards for understanding textbook materials the day before an exam, but more relaxed standards three weeks before an exam. The strictness of these standards theoretically influences the kinds of cognitive processes readers engage in during reading, and these standards vary due to reading purpose, text difficulty, or the motivation level of the student.

In one suggestive study, van den Broek et al. (2001), it was investigated the effects of readers' purposes on inference generation and memory for expository text from a science magazine. L1 College students were assigned to *a read for entertainment purposes condition* (i.e., participants were asked to imagine that they were reading a magazine when they came across an interesting article that captured their attention) or to *a read for study purposes condition* (i.e., participants were asked to imagine that they were reading an article to prepare for an exam in a college course). The students were reading the same texts and they were asked to imagine themselves reading for a different purpose. The researchers used a think-aloud method to measure inference generation, and recall task for measure the information retention in long-term memory. As a result, readers with an entertainment purpose generated more free associations more evaluative comments. In contrast, readers with a study purpose generated more coherence-building inferences such as bridging and predictive inferences, and

paraphrased more often. The results show that reading purpose strongly influence inferential activity and readers' standards of coherence, which in turn influence the types of inferences that they draw and the final memory representations that they construct.

2.6 Activation of Knowledge Structure in Text Comprehension

As described before, reading comprehension involves memory for the meaning of the text rather than from its surface structure. A meaningful representation is created not only through construction and retrieval of representation of explicit text, but also activation and the application of applying the readers' text-appropriate general knowledge to generate knowledge-based inferences.

Many studies concerning the role of prior knowledge in reading comprehension have been done in the field of schema theory in 1970's and 1980's. (e.g., Rumelhart & Ortony, 1977; Seifert, Abelson, McKoon, & Ratcliff, 1986) It has been asserted that in text comprehension, although it is a strong factor for the readers' predictive reading when they have specific background knowledge of the text, it is not always the case that the readers have the sufficient knowledge about the text. Especially. in expository reading, the existence or non-existence of specific knowledge strongly affects the inference generation. On the other hand, in the case of narrative reading, the comprehension that why the character had the certain goal and acted towards his/her own goal, and ended up , and what the writer wants to convey will be achieved based on not *specific knowledge* but *general knowledge*. General knowledge is the accumulation of typical patterns that the readers have learned through various cases that the readers have learned through their experiences or their readings.

In this section, I shall explain the mechanism of knowledge structure which is utilized in reading comprehension.

2.6.1 Theoretical Background of Schema Theory

The clear distinction between memory for the surface structure and the meaning of a text (i.e., the idea that general knowledge is used in both remembering and understanding new information) was first introduced by Bartlett (1932). He carried out a series of studies on the recall of Native American folktales, and noticed that many of the recalls were not accurate, but involved the replacement of unfamiliar information with something more familiar. They also included many inferences that went beyond the information given in the original text. In order to account for these findings, Bartlett proposed that people have *schemata*, or unconscious mental structures, that represent an individual's general knowledge about the world. It is through schemata that old knowledge influences new information. Schemata were originally proposed to represent knowledge structures of familiar events or relationships among events stored in hierarchies in long-term memory (Bartlett, 1932; Rumelhart & Ortony, 1977). Some researchers use different definitions for the concept of general knowledge such as *frames* (Fillmore, 1976), *scripts* (Schank & Abelson, 1977). Also, Seifert et al. (1986) define the structure of schema as follows: "the information in a schema is assumed to be organized into structure, reflecting, for example, temporal contiguity, importance, or more abstract relations such as that between a goal and a plan for its attainment. The content of the schema and its structure determine the inferences that can be made when the schema is activated in memory" (p. 220). As another important aspect of schema-theory, Rumelhart and Ortony (1977) explained in his reading model, *interactive model* that top-down and bottom-up processing should be occurring simultaneously. At first, the textual data that needed to activate schemata were input through bottom-up processing. Then, top-down processing facilitates their assimilation if they are anticipated or consistent with readers' conceptual idea; bottom-up processing ensure that the reader will be sensitive to information that is novel or that does not fit readers' ongoing

hypothesis about the content or structure of the text. Moreover, top-down processing helps the reader to resolve ambiguities or to select between alternative possible interpretation of the incoming data. Thus, a fundamental assumption of the schema theory is that the process of comprehending a text is an interactive process between the readers' general knowledge of content and structure, and the text itself. The text does not by itself carry meaning. Rather, a text only provides guidance for readers how they should construct the intended meaning from their own general knowledge.

In order to illustrate the schematic interpretation and the simultaneous top-down and bottom-up processing, Rumelhart and Ortony (1977) showed following short text: "Mary heard the ice cream man coming down the street. She remembered her birthday money and rushed into the house…." (p.265). After reading these few lines, most people generate the following inference: 'Mary is probably a little girl who hears the bell ringing on the ice cream man's vehicle. Because she wants some ice cream, she runs into the house to get her birthday money so he can buy the ice cream.' What is important here is that the text does not actually say any of this, but inferences beyond explicit textual information were generated by readers' schema activation when they interpret the text. However, what will happen if the text were to continue: " … and locked the door" (Fillmore, 1976). The reader needs to go back and revise their interpretation, and activate another schema against which to make the text compatible. Perhaps, for example, *Mary is afraid that the ice cream man will steal her birthday money*. In this way, in the interactive process between the reader and texts, readers often predict the upcoming events by activating the textual input information and their schema at the same time. If readers notice their prediction was incorrect, it will be immediately modified. This predictive reading is one of the particular ability of proficient readers, and is important in successful comprehension (Cain, Oakhill, Barnes, & Bryant, 2001). Comprehending words, sentences, and

discourse, then, involves much more than just relying on reader's linguistic competence. In fact, reader's linguistic competence is just one part of reader's total general knowledge.

So far, the importance of utilizing schema in text understanding has been described. However, all of these studies were performed in L1 research. It was second language (L2) reading research in the 1980s which based on the schema theory as the foundation that research was done most briskly about a reader's knowledge and reading comprehension. In particular, Carrell, Devine and Eskey (1988) claimed that schema theory has provided numerous benefits to ESL teaching and, indeed, most current ESL textbooks attempt schema activation through pre-reading activities. Also, Carrel divided schemata into two main types: *content schemata* (general knowledge of the world) and *formal schemata* (general knowledge of rhetorical structure) on the basis of its nature. Now, the significance of schematic knowledge is widely acknowledged in foreign language teaching and many researches in the schema-oriented area of ESL/EFL teaching have been carried out.

2.6.2 Thematic Abstraction Unit (TAU)

In reading comprehension, especially narrative comprehension is largely determined by the goals and plans of the central characters. Readers use the character's goals to draw causal connections among events and to focus attention on actors, places, and objects in the situation model that are likely to be relevant to achieving or preventing those goals. In these processes, schemata about goal-planning may be activated during reading. One particular example of such schemata, *the thematic abstraction unit* (TAU) proposed by Dyner (1983) is the abstract knowledge pattern of goals and plans reflected in common adages. For example, the adage "Closing the barn door after the horse is gone" expresses the point of the following two stories shown in Table 2.5.

Table 2.5

Example of the Story Shared TAU "Closing the Barn Door After the Horse is Gone"

Story 1: Academia
Dr. Procoff knew that his graduate student Mike was unhappy with the research facilities available in his department. Mike had requested new equipment on several occasions, but Dr. Procoff always denied Mike's requests. One day, Dr. Procoff found out that Mike had been accepted to study at a rival university. Now wanting to lose a good student, Dr. Procoff hurriedly offered Mike lots of new research equipment. But by then, Mike had already decided to transfer.
Story 2: Wedding Bells
Phil was in love with his secretary and was well aware that she wanted to marry him. However, Phil was afraid of responsibility, so he kept dating others and made up excuses to postpone the wedding. Finally, his secretary got fed up, began dating, and fell in love with an accountant. When Phil found out, he went to her and proposed marriage, showing her the ring he had bought. But by that time, his secretary was already planning her honeymoon with the accountant.

TAUs are explained that it represents the thematic structure of an episode but serves many episodes in memory representation (Zhang & Hoosain, 2005). The fact that the patterns of goal-plan interactions represented in TAUs are easily recognized was shown by Seifert and Black (1983). In one experiment, they gave participants example stories that shared a common TAU (like the two stories above) and asked the participants to identify similar pair stories. As a result, of the stories written by the participants, 82% matched the TAUs of the examples. In the second experiment, participants were sufficiently able to sort stories according to their TAU patterns. However, as a limitation of their experiment, the possibility remains that this activation of abstract knowledge structure occurred strategically only in the experimental context. Their participants were explicitly aware of the attempt to retrieve the prior stories. To address this limitation, Seifert et al. (1986) examined whether the activation of abstract knowledge pattern will occur automatically when the participants were unaware of the relevance of prior stories.

They conducted the experiment on the basis of following assumption: if TAUs provide a way of connecting episodes of the same structure, then one episode of a particular schema may activate another. In their Experiments 1 and 2, the participants read pair stories which were based on same TAU or different TAU, and then made verification judgments on test sentences. Target test sentences that expressed the conclusion of a story appeared in four different priming conditions, namely, the priming sentence was the setup (i.e., the initiating circumstances of a story) from the same story, the setup from the other story, the conclusion from the other story that had the same TAUs. The results of Experiment 1 showed that verification recognition time for target test sentences in pair stories with the same theme were not faster than that for pair stories with different themes. In Experiment 2, experimental conditions were identical to those of Experiment 1 except that participants were encouraged to conduct strategic processing through the interaction to think about the theme of the story during reading, and rating similarities between stories after reading. It was found that verification recognition time for the same TAU was faster than that for different TAU. In Experiment 3 and 4, stories were presented one word at a time at the rate of 250ms, plus a 500ms pause at the end of each sentence, and a 2000ms pause at the end of a story. In either experiment there was no evidence of the activation of thematic information, even in Experiment 4 where specific instructions were given. To explain this, Seifert et al. added a pre-task in Experiment 5 and 6. In the pre-task phase, the participants read stories, answered questions about them, and wrote summaries. The results of both experiments showed that recognition time for test sentences indicating the same TAU was faster than those indicating different TAUs. Seifert et al. estimated the participants might use strategies in reading the same stories (trying to remember them), and the pre-task led to their better memories for the same stories so that it was easier for the activation of test sentences in a study-test phase. As a result, Seifert et al. concluded that the activation of thematic

information is not automatic in text reading but it depends on the reader's strategies.

Another study related to activation of thematic information was done by Till, Mross, and Kintsch (1988), although their results are not directly relevant to the issue of whether thematic inferences are normally made during reading. In their experiment, the participants read two-sentence texts presented by a rapid serial visual presentation (RSVP) procedure and performed a lexical decision task on target items that followed ambiguous prime words in the texts. When the target was a word, it was either an appropriate associate of the prime word, an theme-appropriate inference word generated by the text, an inappropriate associate of the prime word, or an theme-inappropriate inference word. Till et al. found that there was no difference between theme-appropriate inference words and theme-inappropriate inference words at short (200ms to 400ms) stimulus onset asynchrony (SOA); but at longer (1,000ms to 1,500ms) SOA, thematic inference words were strongly facilitated relatively to unrelated control words. However, Long et al. (1994, 1997) found that proficient readers showed contextual effects on thematic words with the SOA of 500ms, using a procedure similar to Till et al.'s experience.

In this way, it has been clarified by some researchers that abstract knowledge structures as typified by TAU facilitate readers to connect their knowledge to text information and lead to understanding of the text. To summarize these findings, the activation of thematic information based on TAU occurs when the readers were given instructions to turn attention to the similarities of the stories. That is to say, such activation is not automatic and depends on readers' strategies (Seifert et al., 1986). Also, although it takes 1,000 ms to 1,500 ms for the readers to activate the thematic information (Till et al., 1988), proficient readers could activate the thematic information in just 500ms after reading the context (Long et al.,1994).

However, what kind of contextual information facilitates the activation of thematic information during reading has not sufficiently been examined. That is to say, there is a

possibility of accessibility to readers' knowledge depending on the content of the reading context. I believe that this point should be researched in much more detail. In next section, the relationship between knowledge accessibility and reading comprehension will be described.

2.6.3 The Effect of Analogy on L1 and L2 Reading

In order to activate the readers' knowledge, some methods such as adding specific information, illustration, or graphic organizer to the text have been used. Another way to activate comprehension is an analogy. Analogy is a comparison between two different things in order to highlight some point of similarity (e.g., the mechanism of *camera* is illustrated to explain the structure of *eye*). By using the analogy, the readers can relate their knowledge to the text information (Gentner, Loewenstein, & Thompson, 2003; Gentner & Markman, 1997).

To understand new information when reading, we often utilize similar concepts or situations derived from our prior knowledge. The cognitive process in which a well-known domain (i.e., source analog) is applied to a new domain (i.e., target information) is called *analogical transfer.* Many studies have established that analogical transfer works as a powerful mechanism for promoting the learning of new information through reading, conceptual change and problem solving (Blanchette & Dunbar, 2002; Gentner & Markman, 1997; Holyoak & Koh, 1987). Thus, reading to learn new information requires one to understand the relevant concepts in a similar domain, as well as to interpret the relations between those concepts. Previous research has demonstrated the *knowledge effect* in learning from text: readers who have more knowledge of the text topic tend to show better comprehension and learning from texts (Braasch & Goldman, 2010). However, it often happens that learners who are reading to learn do not have sufficient prior knowledge, and have difficulty making the relevant connections between the target and analogous domains. To begin with, it has been pointed out that it is difficult for

EFL learners to activate their prior knowledge, since it is higher-level text processing that requires cognitive resources such as attention and working memory (Nassaji, 2003).

One solution to this difficulty is to provide an analogical text as an advance organizer. The effects of giving an analogy prior to reading the target text could be explained in terms of two major points: (a) the activation of relevant knowledge and (b) the analogy as a basis for constructing a text representation, which assists learners in higher-level text processing. As for point (a), learners are able to compensate for the lack of prior knowledge by using analogical text information. Since information from the analogy activates the relevant schema at the same time, it is easier for them to construct a text representation. Taniguchi (1988) suggested that showing an analogy activates the readers' schema related to sentence information, and understanding is then promoted by assimilating new information into the schema.

Point (b) is that analogical information serves as a basis for constructing a text representation and helps with reading comprehension. The target text and source analog share a similar propositional structure within the text, so if learners construct a basis for a text representation from a source analog in advance, it helps them construct the target text representation, because they have only to map the propositions and their relations to the target text onto the representation base of the source analog. According to the *structure-mapping theory* (Gentner et al., 2003; Gentner & Markman, 1997), the comparison processes act to achieve a structurally consistent alignment between two representations. Structural consistency requires that the correspondences between the surface features and structural relations in two representations must satisfy one-to-one correspondence. Once a structural alignment has been established, readers can generate inferences about what is connected to the base structure but is not explicitly presented in the target (Day & Gentner, 2007).

Let us look at specific examples of how analogies facilitate the understanding of text.

Consider first a story: A hunter who shoots an arrow at a hawk, but misses because the arrow does not have any feathers on it to help stabilize its flight. The hawk then gives the hunter some of its own feathers, and the hunter is so pleased that he promises not to hunt hawks anymore. Consider a second story: An aggressive country attacks its neighbor with missiles. The missiles fail to do any damage because they were poorly guided and missed their targets. The neighbor, which makes supercomputers, offers to sell some to the aggressive country. The aggressive country is appreciative and promises never to attack its neighbor again. These two stories differ in terms of surface features (e.g., *hawk* and *arrow* vs. *countries* and *missiles*) but share a similar structure. If learners had read the first story before reading the second story, information on the relations between propositions (e.g., *A hunter shoots at a hawk with an arrow* vs. *An aggressive country attacks its neighbor with missiles*) is activated and facilitates the structure mapping of the two stories. That is to say, readers can easily build connections between propositions by applying the propositional structure of the source analog from the first story, thus promoting the understanding of the target of the second story.

Concerning both points stated above, reading is a complicated process that is achieved through lower-level text processing such as word recognition, lexical access and syntactic parsing, and through high-level processing, which includes grasping sentence structure, constructing a propositional text representation, and the global understanding of discourse. Hence, for less proficient readers such as EFL learners, it is very difficult to reach a comprehensive understanding of the text. However, through the presentation of analogies, the activation of their schema and text structure construction should become easier, thus decreasing the cognitive burden on higher-level processing and making the process of reading easier for less proficient EFL readers. The author believes that it is very instructive for English teaching to clarify the facilitative effects of analogical transfer in EFL reading comprehension.

2.6.4 Mechanisms of Analogical Transfer

Many previous studies have claimed that analogical transfer between the representations of the source analog and target is caused by two types of similarity: *surface similarity*, which refers to the semantic resemblance between the elements in the source analcog and target, and *structural similarity*, which refers to the resemblance in the underlying system of relations between the source and the target elements (Blanchette & Dunbar, 2002).

These two types of similarity are defined based on three different levels of matches: *entity match*; *first order relation (FOR) match*; and *higher-order relation (HOR) match* (Gentner, Ratterman, & Forbus, 1993). The features of each match are explained as follows. An entity match refers to the minimal surface feature overlap of the entities involved in relations. Each entity shares the same categorical or semantic attributes in terms of the characters, physical objects, and locations (e.g., *hawk* vs. *eagle; arrow* vs. *crossbow*). A FOR match, the second level of match, involves events or other relations between entities. Its relations take two or more objects and a similar predicate (e.g., *A hunter shoots at an eagle with an arrow* vs. *An aggressive country fires on a neighboring country with missiles*). A HOR match, the third level, points to causal relations or other kinds of plot structure. It shares information such as characters' actions and outcomes relevant to the causality between two storylines (e.g., an attack is made but fails; this causes the one being attacked to offer to provide an item to the attacker to help the attacker; this offer causes the attacker to be grateful and to promise not to attack again).

The similarity features of the analogical text are determined according to how these matches are shared between source analog and target text. Surface and structure similarities are separated by the presence or absence of entity matches. Also, a FOR is basically part of either similarity type if these texts are analogous, whereas the causality of the two texts grows as the number of FOR and HOR matches increase. Table 2.6 briefly shows the combination of matches

comprising each similarity feature.

Table 2.6

Components of Common Matches in Each Similarity Feature

	Similarity types	
Causality	Surface similarity	Structural similarity
Strong	Entity, FOR, HOR match	FOR, HOR match
Weak	Entity, FOR match	FOR match

Over the past 20 years, the role of surface and structural similarities in analogical transfer has been investigated through empirical research. Some of these studies have shown that analogical transfer is driven mainly by surface similarity (Gentner et al., 1993; Keane, 1987). For example, Gentner et al. (1993), using the recall paradigm, found that a source analog that shares many entity matches with the target text (as in the case of surface similarity) is more likely to lead participants to recall target text information compared with conditions in which only FOR is shared. In addition, they showed that a source analog that shares a HOR but had no entity matches (as in the case of structural similarity) is more likely to lead to reminding than conditions in which only FOR is shared. However, the effect of structural similarity was much smaller than that of surface similarity. In short, the results suggested that surface similarity plays a large role in analogical transfer, while the effect of structural similarity is relatively small.

On the other hand, subsequent studies have argued that structural similarity is also crucial for analogical transfer (Blanchette & Dunbar, 2002; Catrambone, 2002; Gentner & Markman, 1997). Catrambone (2002), for example, systematically manipulated the number of surface and structural similarities. His results involving reminding in recall tasks and reading time data indicated that surface and structural similarities affected the transfer process about as equally

when a HOR match was shared in both conditions. When a HOR was not shared, though the effect of surface similarity continued, the effect of structural similarity decreased. This is a particularly insightful finding in terms of showing that the degree of causality (i.e., the presence of HOR) is the key factor in the use of structural similarity.

However, these past studies have only focused on whether text information could be retrieved, because they were conducted in psycholinguistic contexts. Thus, they have not clarified the issue of what kind of text information was left by analogical transfer in readers' text representations after reading. Successful reading involves constructing a coherent representation of text. This is accomplished by actively integrating the important main ideas from text rather than detailed ideas of lesser importance (Graesser et al., 1994). Therefore, it is useful in this study to examine the extent to which the important information remained with different similarity features. As stated above, the effect of enhancing information understanding and promoting text representation construction by analogical transfer has been addressed mainly in first language (L1) research. Few studies concerning this issue have been conducted in second language (L2) contexts. Furthermore, the L2 studies that do exist report that analogies did not aid in the L2 reading comprehension process because the addition of analogies to the target text appeared to make the reading process even more multifarious (Brantmeier, 2005; Hammadou 2000). However, one limitation in these L2 studies was that surface and structural similarities between the target and source analog were not manipulated with any degree of precision. Also, as L2 reading proficiency was not given much consideration, it was not made sufficiently clear whether the effect of analogical transfer on EFL learners' reading varies with proficiency level.

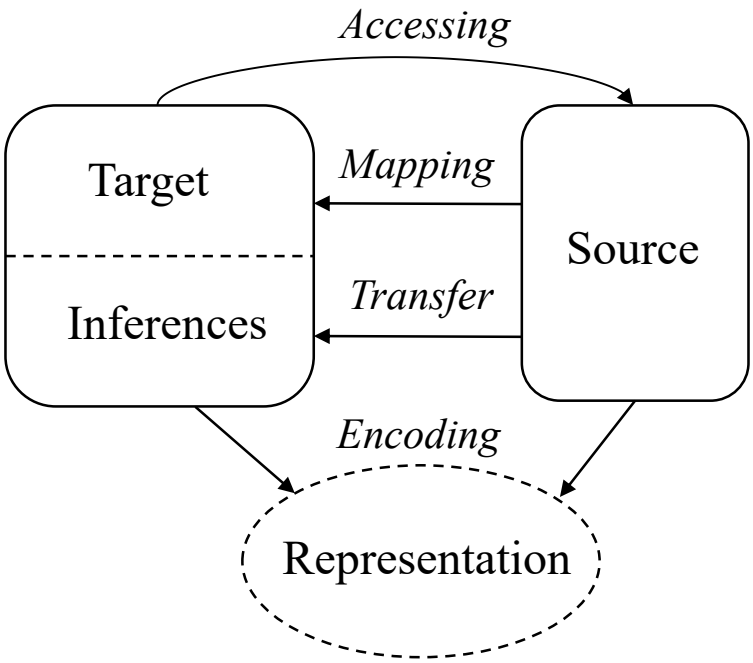


Figure 2.3. The serial process of analogical transfer.

Further, consideration should be given to the sequence processing by which the two types of similarity affect analogical transfer. Analogical transfer progresses through three basic steps: (a) *accessing* the source analog information relevant to the target; (b) *mapping* the components of the source and target; and (c) *encoding* the mapped information to the text representation (Holyoak & Koh, 1987). Figure 2.3 depicts the serial processes of analogical transfer. Blanchette and Dunbar (2002) have raised the possibility that the two types of similarity involve analogical transfer in different ways. They suggested that surface similarity mainly promotes the accessing and mapping processes while structural similarity influences mostly the encoding process. This issue has yet to be sufficiently studied, though it is crucial to understand the processes by which readers activate relevant information.

2.7 Methodologies for Inference Generation

Whether an inference is generated and when it is generated can be investigated by measuring specific aspects of the reading behavior. The methodologies for inference generation are generally divided into two types: on-line method (i.e., during the reading of the words or sentences that trigger the inference) and off-line method (i.e., after reading those words or sentences (Iseki & Kaiho, 2002).

2.7.1 On-line Methods

Generally, on-line methods are used to detect the inference activation process immediately during reading. The methods based on the assumption that inference generation is a time-spending process. Therefore, inferences are detected by longer reading time at the moment the inference is generated, relative to a control condition in which no inference is generated (Noordman & Vonk, 2015).

Reading Times

The measurement of reading times for target sentence is often applied in order to detect the inference activation during reading (i.e., on-line inferences). In the case of self-paced reading procedure, readers are asked to press a button exposes successive units of text (words or sentences) on a computer screen. The interval between button presses is collected as the reading time for the unit of text. Calvo, Meseguer, and Carreiras (2001) employed this procedure and found that facilitation did not occur in reading the target word itself but rather in the post target and final regions of the continuation sentence.

Furthermore, some studies have used the reading time paradigm based on the readers' inconsistency detection process as an indicator of inference activation. The method is called is

called *the contradiction paradigm* and was developed by O'Brien and Albrecht (1992). In this paradigm, a target sentence is presented that match well within the local context but conflict with the context information located earlier in the global discourse. A slowdown in reading time for the target sentence is regard as evidence that readers activate relevant information from a previous text when they read a target sentence, and thus interfered with comprehension. Under these condition, it is reasonable to interpret a slowdown in reading time as an indication that activation occurred (Peracchi & O'Brien, 2004). Reading times are also used in eye-movement assessment that reveal what the reader looks at and how long. This makes it possible to measure exactly when the inferences are generated.

Probe Recognition, Lexical Decision and Naming Tasks

Other inference detection methods such as the probe recognition, lexical decision, naming tasks also be defined as online methods. In these tasks, a word or sentence is presented as a probe immediately after the context that evokes the inference.

In *a probe recognition task*, participants are asked to decide whether or not the probe word had been presented in the previous sentence. For example, McKoon and Ratcliff (1986) presented the probe word 'dead' after the sentences such as 'The director and cameraman were ready to shoot close-ups when suddenly the actress fell from the 14th story.' Recognition time for the probe word were slower than those in a control condition that the inference probe 'dead' had been activated and was difficult to discriminate from explicitly written information. Concerning the result, they interpreted that the predictive context activate the inference concept, but the recognition time pattern showed that the level of inference activation was weak.

In *a lexical decision task*, participants are assigned to decide whether or not the probe's letter string is a real word. For example, Allbritton (2004) measured the lexical decision time

to for inference related probe words to detect the evidence of on-line generation of predictive inferences. As the critical sentence that strongly suggest a predictive inference, 'He [Superman] grabbed the steel bars and slowly began to pull,' was presented, immediately after reading the sentence, participants were asked to judge whether the lexical decision target "bent" was a word or nonword. Their results showed the lexical decision time were shorter for target related to the inference but this only occurred when there was a strategy to encourage monitoring what happens next. They suggested the possibility that predictive inferences are generated relatively strategically.

In *a naming task*, participants are required to pronounce the probe word. Murray et al. (1993) employed the following context that supporting the predictive inference: 'Stenven had been married for years, and his resentment had been building up. One day, no longer able to control his anger, he threw a delicate porcelain vase against the wall.' Naming time for the predicted target word 'break' were significantly faster than control condition.

These evidences of inference activation in the above methods are based on the assumption that if the inference is encoded into the text representation, the lexical decision time for the probe and its naming time are shorter. Also it takes more time to judge that 'break' did not occur in the text is longer than in a control condition in which the inference is not generated.

2.7.2 Off-line Methods

Contrary to on-line methods, off-line methods are used to detect the inference generated after reading. The methods based on the assumption that inference generation are encoded into the mental text representation, as well as information that is written explicitly in the text. Recall and verification tasks are classified as off-line methods and are often used in inference studies.

Recall Task

In *a recall task*, the reader is asked to describe everything they remember after reading the text. Information that was not stated explicitly in the text but that is recalled by the reader is supposed to be inferred by the reader. However, recall task has a negative aspect in that it can also function as a memory test. The task performance is easily influenced by task conditions. For instance, there is the evidence that recall protocols produced in L1 and L2 differ both qualitatively and quantitatively (Koda, 2005). Also, due to our limited cognitive resources, as the text used in the recall task becomes longer, it is more difficult for readers to remember and write all the information in the text. The fact imply that the task does not measure their understanding but rather their memory of a text. Therefore, researchers when researchers use a recall task to assess learners' understanding, they should take into consideration its advantages and disadvantages.

As one of the ways to reduce the burden on memory, retrieval cues are sometimes presented while a learner recalls a text. This type of recall task is called *a cued recall task*. The presentation of retrieval cue enables learners to recall the information more easily. Moreover, especially in the inference research field, a cued recall task is used for examining how generated inference information is integrated in learner's mental representation. That is because, the information that learners recall while making use of cues should highly related to cue concepts in their representations. For this reason, the present study mainly used recall task with inferential cues to examine whether EFL readers generated inferences and encoded them into their mental representation.

Verification Task

In *a verification task*, readers are required to judge whether the content of a sentence is true or false with respect to the content of the text. It is important not only the correct answer rates but also the recognition time for the verification. If inference information is recognized as quickly as explicit information, this indicates that inference is generated during reading, and not at the moment of off-line task (Noordman & Vonk, 2015). Off-line methods are used frequently in combination with on-line methods to answer the question of when inferences are generated.

2.8 Connection to Previous Research and Significance of This Thesis

The purpose of the present study is to clarify the process of the knowledge-based inference generation which are drawn by connecting textual ideas and readers' prior knowledge in Japanese EFL learners' reading. Empirical evidence from L1 and L2 reading research indicates that when readers comprehend a text, they attempt to build a mental representation by making inferences to coherence with the explicit contents in the text (e.g., Graesser et al., 1994; Horiba, 1996, 2000). However, since very little research has been done on inference generation in L2 context compared with that of L1 (e.g., Collins & Tajika, 1996; Horiba, 1996, 2000; Yoshida, 2003), little is known about the mechanism of L2/EFL readers' inference generation. The purpose of the current dissertation is to investigate the generation of EFL learners' knowledge-based inferences by considering the texts and learner factors.

Chapter 3, Study 1 examines the generation of thematic inferences in Japanese EFL learners' sentence-level reading. In sentence-level reading, processes are necessary to form

structures that represent the syntactic and conceptual relationships among words in phrases or clauses. These processes help in encoding propositions, abstract units that represent the meaning of a sentence.

Chapter 4, Study 2 investigates the generation of predictive inferences in EFL learners' text-level reading and the effect of the text factor and learner factor on the generation of predictive inferences in EFL reading. At text-level reading, processes are required to form connections among successive propositions in a text. The processes mentioned above involve accessing not only knowledge about the language (e.g., word meaning, syntax), but also knowledge about the world. Study 2 includes two experiments which examine the generation of predictive inferences. Experiment 2 uses a probe recognition task for the activation and encoding of predictive inferences. Corrected data are analyzed in terms of the strength of contextual constraint and L2 reading proficiency. Experiment 3 is conducted to clarify the influence of contextual constraint over the content of readers' predictive inferences. In this experiment, the strategy instruction "*What will happen next?,"* which promotes the generation of predictive inferences was given to readers, and a qualitative analysis was conducted on the open-ended answers according to their categories.

Chapter 5, Study 3 looks into how the text information works as access cues to the readers' knowledge structure. Experiment 4 examined how the similarity features between texts (i.e., similarity types and causality) and L2 reading proficiency affect EFL learners' reading comprehension through analogical transfer. Experiment 5 checked whether EFL learners were activating the between-text-level theme of the text.

In sum, the present study formulated three research questions (RQs):

RQ1: How do the text and learner factors affect EFL readers' knowledge-based inference generation?

RQ2: How does the presentation of analogy with either surface or structural similarities affect EFL readers' knowledge-based inference generation?

RQ3: How do the text and learner factors affect EFL readers' explicit text comprehension after reading?

In order to clarify these points, three studies were conducted. RQ1 was investigated through Study 1 and 2 (Experiment 1, 2 and 3). The Study 1 examined sentence-level reading and Study 2 dealt with text-level reading. RQ 2 was investigated through Study 3 (Experiment 4 and 5). Study 3 examined between texts-level reading using analogical texts. Moreover, RQ 3 was investigated through all studies 1 to 3 (Experiment 1 to 5). Table 2.7 summarize the variables and measurements used in the present study.

Table 2.7
Summary of Variables and Measurements Used in the Present Study

Study	EXP	Learner Factor	Text Factors	Measurements	
				On-line	Off-line
1	1	L2 reading proficiency (2: Upper, Lower)	Contextual constraint (2: High, Low)	Lexical decision (2: appropriate, inappropriate)	Cued recall
2	2	L2 reading proficiency (2: Upper, Lower)	Contextual constraint (2: High, Low)	Probe recognition (3: Explicit, Inference, Control)	Cued recall
	3		Contextual constraint (2: High, Low)		(a) Inference generation task, (b) Cued recall
3	4	L2 reading proficiency (2: Upper, Lower)	Similarity types (2: Surface, Structural) Causality (2: High, Low)	Probe recognition (3: Explicit, Inference, Control)	Cued recall
	5	L2 reading proficiency (2: Upper, Lower)	Structural similarity text (2: Same-theme, Difference-theme)	Reading time Cue text types (3: Complete,, Partial, Predictive)	Cued recall

Chapter 3

Study 1: Generation of Knowledge-Based Inferences in Sentence-Level Reading

3.1 The Purpose of Study 1

Good reading comprehension involves reading the words in a text, accessing their meanings, and grasping the sense of each sentence. To understand text in a meaningful way, readers need to integrate the meaning of successive sentences and to establish local coherence. Readers also need to establish how the information fits together as a whole, that is, global coherence. For both local and global coherence, readers need to incorporate general knowledge and ideas (retrieved from long-term memory) to make sense of details that are only implicit. On the other hand, poor reading comprehension is often associated with a failure to make appropriate knowledge-based inferences during reading. Less proficient readers have difficulty making inferences to integrate ideas in a text, to answer questions about it, and to identify main ideas and themes (Long et al., 1994, 1997).

In sentence-level reading, processes are necessary to form structures that represent the syntactic and conceptual relationships among words in phrase or clause. These processes help in encoding propositions, abstract units that represents the meaning of a sentence. At text-level reading, processes are required to form connections among successive propositions in a text. The processes mentioned above involve accessing not only knowledge about the language (e.g., word meaning, syntax), but also knowledge about the world.

The purpose of study 1 is to clarify the generation of knowledge-based inferences in Japanese EFL learners' sentence-level representations. Specifically, I examine proficient and less proficient readers' ability to generate inferences related to the topic of discourse. One issue in inferential processing research is to what extent do readers actually generate knowledge-

based inferences during reading. According to the minimalist hypothesis approach (e.g., McKoon & Ratcliff, 1992), readers generate only a limited amount of inferences during reading. The theory claims that, without specific goal-directed strategic processes, only inferences that are based on easily available information are encoded automatically during reading. On the other hand, the constructionist theory approach argues that knowledge-based inferences, even though they are not based on easily available text statements or general knowledge, are still encoded during reading. That is because they are important for establishing overall discourse coherence (e.g., Graesser et al., 1994). In Study 1, thematic inferences will be investigated as one of the knowledge-based inferences which are generated based on not easily available information. When readers try to understand the main points and morals in narrative readings through knowledge-based inferences, readers' general knowledge gained from their own experiences (e.g., goal-planning information) will be necessary.

It has been made clear from previous studies that knowledge-based inferences while reading require a certain amount of time for readers, (Till et al., 1988), and also they are generated when the text incorporated a question inviting the inference (Hannon & Daneman, 1998). It has also been found that the generation of knowledge-based inferences is an ability only possessed by proficient readers while it is difficult for less proficient readers to generate inferences while reading, although they had built an appropriate propositional textbase and had necessary information (Long et al., 1994, 1997). The above findings indicate that the automatic generation of knowledge-based inferences is very difficult for readers.

However, it has not been made clear if the activation of knowledge-based inferences occurs during EFL reading context and what kind of textbase representation is being built in the process. It is also assumed that when we compare knowledge-based understanding at the sentence-level and knowledge-based understanding at the text-level, the text-level appears to

be more difficult for readers because it needs to integrate information from different parts of a text. Long et al. (1997) have examined this point by using a pair of stories based on TAU. They have concluded the proficient readers' recognition time for the conclusion sentence was faster in the same TAU shared condition than in a different TAU shared condition so this fact indicates that proficient readers can generate knowledge-based inferences while reading.

Nevertheless, there is little research that compares sentence-level and text-level reading. In addition, a full investigation has not been conducted as to the effect of other text conditions (e.g., contextual constraint). The previous materials used in L1 research are quite varied in that both contexts and inference target words are strongly connected or weakly connected. So far, there has not been a study which investigates the influence of the constraints of contexts and inference words. I shall examine knowledge-based inferences in EFL reading as the text factors by manipulating the contextual constraints. The intensity of contextual constraints refers to the accessibility to inference information in readers' minds.

3.2 Experiment 1: Effects of Contextual Constraint and L2 Reading Proficiency on Thematic Inference Generation

Experiment 1 examines the generation of thematic inferences in sentence-level reading. It refers to the ability of readers' to make elaborate connections among concepts that appear in the same proposition and their ability to elaborate their sentence-level representation. The research questions of Experiment 1 are set as follows:

RQ1-1: Do proficient and less proficient EFL readers activate thematic inferences automatically in sentence-level reading?

RQ1-2: What kind of effects do contextual constraints have on the activation of thematic

inferences?

RQ1-3: How do thematic inferences and contextual constraints influence the construction of the readers' mental representation?

RQ1-4: How sensitive are proficient and less proficient readers' perceptions of the relationship between thematic inferences and passage information?

3.2.1 Method

3.2.1.1 Participants

The participants in this Experiment 1 were 40 first- to fourth-year Japanese students at a national university with diverse majors: international studies, science and engineering, medical science, education and so on. The participants were recruited from general English classes and were paid a small amount of remuneration for their participation in this Experiment 1. All participants were divided into two proficiency groups based on their scores in an L2 reading proficiency test. One student was excluded from the data analysis because his score in the comprehension questions was under 80%, which was regarded as not comprehending the passages sufficiently. As a result, 39 students participated in the analysis (see 3.2.1.4 Scoring and Data Analysis for details).

3.2.1.2 Materials

Two types of materials were prepared for Experiment 1; one was to assess the participants' L2 (English in the present case) reading proficiency, and the other was Reading passages and thematic target words for a lexical decision task and a cued written recall task in which participants' thematic inference generation was studied. I shall explain the L2 proficiency test first, and then the reading passages for the lexical decision task.

(1) *L2 Reading Proficiency Test*

In respect to testing the participants' English reading proficiency, total of 34 test items were adopted from the Society for Testing English Proficiency (STEP) pre-Grade 1 (14 items), Grade 2 (17 items), and pre-Grade 2 (3 items) tests (2007). The test comprised of 14 fill in the blank questions and four reading text with comprehension questions from a reading section.

(2) *Reading Passages and Thematic Inference Target*

Prior to the main experiment, a pilot study was conducted in order to select the material texts and to divide the materials into the high or low contextual constraint conditions. As stated before, in previous studies, there was the problem that accessibility to thematic inferences derived from texts were not sufficiently controlled. See the following sample passages and inference targets in Table 3.1.

Table 3.1
Sample Passages and Inference Targets Used in Prior Studies

Passage	Thematic target Items	
	Appropriate	Inappropriate
(1) The townspeople were amazed to find that all the buildings had collapsed except the *mint*. Obviously, it had been built to be ready for natural disasters.	earthquake	breath
(2) The bird loved to compete with others. But no matter how hard he worked, he could not keep up with the *swallow*.	flying	birthday

Note. Passages and target words are used in Till et al. (1988), Long et al. (1994, 1997), and Honnon and Daneman (1998); words in italics are homographs to examine the sense selection according to context in prior study.

These thematic target words (e.g., earthquake, flying) were the modal responses made by a group of pilot participants in Till et al.'s (1988) study. From passage (1), we can easily infer

that its theme is *earthquake* when we read the expressions such as "all the buildings had collapsed except the bank" and "natural disasters" because the thematic inference is generated as a plausible explanation for why the buildings had collapsed.

On the other hand, in passage (2), "the bird loved to compete" or "He could not keep up with the *swallow*" may be associated with the word *flying*. However, this word was not drawn from the question why such situation had happened. As such, it could be assumed that contextual constraint is weaker in paragraph (2) compared with paragraph (1).

Therefore, the purpose of pilot study was to eliminate passages which are clearly difficult for EFL learners to generate inferences. In addition, in this pilot study, inference generation tasks and rating tasks were conducted so as to divide experiment materials according to their contextual constraints (high contextual constraint, low contextual constraint; hereafter these are called high CC and low CC in each case).

A total of 10 undergraduate students at the same national university in Japan participated in the Pilot study. Their majors were humanities and international studies and they were 18-19 years of age in the first year of their courses. The participants were recruited from general English classes and were paid a small amount of remuneration for their participation in this Pilot Study. As the pilot study materials, 56 experimental passages comprised of 28 pair sets were adopted. Each passage in the pair was two sentences in length and was accompanied by a topic-appropriate lexical[3] decision target and topic-inappropriate lexical decision target. These passages and accompanying thematic targets were identical to those used by Till et al. (1988), Long et al. (1994, 1997), and Hannon and Daneman (1998).

The manipulation of passages was conducted in two phases. Previous studies have used

[3] The topic-appropriate lexical decision targets were modal responses made by the pilot study in Till et al., (1988). In this study participants had been asked "to write down a word reflecting their understanding of what the paragraph was about" (p.296).

passages which include homographs, so as to examine the ability to reject context-inappropriate meanings of ambiguous word (see words in italics in Table 3.1). However, in this Pilot study, the semantic understanding of ambiguous words will not be a focal issue. Therefore, homographs for which the participants were not familiar with the meaning were substituted with synonymous words (e.g., *mint* → bank). Also in the Pilot study, passages for which appeared to be difficult for EFL learners to form images due to their cultural backgrounds (e.g., The rabbi looked for something to use in putting up his announcement. Finally, he used his shoe to tack it to the door of the temple; inference target is *Church*) were eliminated. In addition, difficult words were replaced with easier ones. As a result, the 56 materials were narrowed down to 28.

Next, as the second phase, the pilot study was conducted. First, as *an inference generation task*, 10 participants were given leaflets in which the 28 passages were written and were instructed to "write down a word in Japanese reflecting your understanding of what the paragraph was about" (reference from Till et al., 1988). They marked the sentences that were not comprehensible to them as well as the words that they did not understand.

Then, *an appropriate rating task* was conducted. The participants were shown 28 passages and paired thematic target words on different sheets of paper and were given the following instruction: "grade the levels of the words on a 7-point Likert scale as to how the target word matches the image of the reading passage." In addition, the participants' familiarity toward target words was evaluated on a 7-point Likert scale.

The data gained from the pilot study were analyzed in terms of three viewpoints. First, for the lexical decision task in Experiment 1, it was necessary for the thematic target words to have as similar properties as possible in terms of word familiarity, word frequency, and the number of syllables. Therefore, seven passages that had a low familiarity rating score (1.4-4.0),

low frequency score (over 4000 word level in the JACET 8000 word level checker), and words over three syllables were excluded. Secondly, in order to divide the reading passages according to the three levels of contextual constraints, the remaining 21 passages were distributed into three conditions. Namely, seven passages were allocated in high CC condition (M = 6.17, SD = .66), another seven in intermediate CC condition (M = 5.02, SD = .21), and the last seven in low CC condition (M = 3.94, SD = .57) based on the results of an appropriateness rating task on a 7-Likert scale. To compare the differences among the three conditions, a one-way analysis of variance (ANOVA) was employed. As a result, the significant differences were found among three conditions, F (2, 18) = 31.768, $p < .001$. Moreover, when the participants were instructed to write the associative inference words in Japanese, seven target words in high CC condition highly corresponded with the words. It reflected the fact that their associative understanding from the passage was around an 80% match. That is to say, the levels of contextual constraint were reflected in EFL learners' cognitive perception appropriately.

Thirdly, from the result of the above pilot study, 21 passages were selected for Experiment 1. Out of 21 passages, each seven passages were allocated into high CC and low CC condition and it is used as the condition in which present topic-appropriate or topic-inappropriate target words. It was also necessary for topic-inappropriate target words to have similar properties with appropriate target words. Therefore, irrelevant content words which were similar in terms of part of speech, word familiarity, frequency, and number of syllables were created. On the other hand, the remaining seven passages which belonged to intermediate CC were allocated to filler condition and were used as the condition for presenting nonwords in the lexical decision task. Nonwords were newly created using the ARC Nonword Database (http://www.cogsci.mq.edu.au/~nwdb/). In doing so, special attention was paid so the number of letters or syllables would be approximately the same as the average of the other thematic

target words. That is to say, in Experiment 1, two types of presentations were conducted: version A and B. In version A, a passage was presented paired with an appropriate target word, while in version B, the same passage was presented with an inappropriate target word.

Table 3.2 shows one of the materials for each condition that was used in Experiment 1. (Refer to Appendix 1 and 2 for all the materials used in Experiment 1). In addition, 21 true or false comprehension questions were created in order to confirm whether readers could understand the meaning of the passages correctly (e.g., Had the bank building collapsed? / False). In addition, true or false answer questions were adjusted to be the same in number to maintain balance.

Table 3.2

Example of Passages and Inference Targets Used in Experiment 1

Condition	Passage	Thematic Target Items	
		Appropriate	Inappropriate
High CC	The townspeople were amazed to find that all the buildings had collapsed except the bank. Obviously, it had been built to be ready for natural disasters.	earthquake	stranger
Low CC	The little girl was very happy with the new doll from her grandmother. She reached up to hug her and give her little kiss.	love	rise
Filler	When the maid turned away from the laundry, the baby grabbed the iron. Later that day, the maid started looking for a new job.	fribbs (nonword)	

Note. CC = Contextual Constraint.

3.2.1.3 Procedure

The experiment was conducted individually. The process of one session went through the three phases. In the first phase, the L2 reading proficiency test was administered with a time limit of 30 minutes. The participants were handed leaflets in which they wrote down the answers.

In the second phase, the online passage reading and lexical decision task were administered. The participants read one text for practice and also checked how to operate the equipment, SuperLab 4.0, an apparatus that measured the experimental data. The participants undertook self-paced reading on the computer screen. The 21 sets of experimental passages composed of two sentences, thematic inference targets, and comprehension questions were presented in random order with presentation condition A or B.

In advance, the participants were informed that there would be comprehension questions after reading each passage, and were instructed to read the passages carefully. Following the instructions, the sentence, "From now on, the experimental text will be displayed" appeared on the computer screen. The passage was shown on the screen sentence-by-sentence, and subsequent sentences were displayed when the participants pressed the red key, which served as the advance key for SuperLab. When the participants finished reading the last sentence, the mark "***" was shown for 1,000 ms, and then the thematic target words were displayed and the lexical decision task was conducted. This SOA of 1,000 was set on the basis of the findings from Till et al. (1988) that readers take at least 1,000 ms for thematic inference activation. In this task, all lexical decision targets were marked by asterisks (e.g., ***earthquake***). The participants were required to press a key labeled *Yes* if the lexical decision target was a real word, or a key labeled *No* if the lexical decision target was a unreal pseudoword (i.e., nonword). The precise instructions were as follows: "You will read statements after reading two sentences, and you will have to determine whether each target word was a real word or a unreal word without meaning by pressing either the *Yes* or *No* key as quickly as possible." Thus, the participants' responses and recognition time were recorded by SuperLab, an apparatus that measures the experimental data. Following the lexical decision task, the mark "???" was shown for 500 ms on the screen, and a Yes-No comprehension question was presented. If the

participants believed the comprehension question was correct, they would press the *Yes* key; in contrast, if they believed the comprehension question was incorrect, then they would press the *No* key. Their responses were recorded by SuperLab.

Finally, as the third phase, the cued written recall and the appropriateness rating task were conducted. Regarding the cued recall task, the participants were given a piece of paper on which only the thematic target words were printed as a cue. Then, the participants were instructed to write down in Japanese everything that they could remember about the 14 reading passages (except the seven reading passages paired with nonwords). No time limit was set for this task, but it took about 20 minutes for the participants to perform it. The appropriateness rating task was adopted in the same way as the pilot study. After the cued recall task, participants were given another sheet where both passages and paired thematic target words were printed and were instructed to grade the levels of the words on a 7-point Likert scale to show how the paired target word matched the image of each passage they read.

3.2.1.4 Scoring and Data Analysis

First, as mentioned earlier, one participant was excluded from the data since his mean score on the Yes-No comprehension questions was under 80%. As a result, the mean score by the remaining 39 participants was 93.21%.

Secondly, in respect to the data of L2 reading proficiency test, the examiner gave one point for the correct answer and the total score was 34. In order to confirm the difference of two groups, *t*-test was administered.

Thirdly, when the responses of the lexical decision tasks were confirmed, all the participants proved to have the correct answer rates exceeding 80% (M = 98.17%). However, the recognition time for this task, the 25 cells of data (3.05% of the total data) that exceeded

±2.5*SD* from the mean, were deleted from the data as outliers. The correct recognition time[4] for thematic target words was analyzed statistically using a three-way ANOVA, with the factors thematic target (appropriate vs. inappropriate), CC (high CC vs. low CC), and proficiency (higher vs. lower). Both thematic targets and CC were treated as within-subject factors, and the proficiency level factor was a between-subject factor.

Fourthly, for the scoring of the cued recall data, the 21 passages were divided into idea units (IUs) by two raters in accordance with Ikeno's (1996) criteria. The inter-rater agreement was 96.24% and all disagreements were resolved through discussion among raters. One point was given when each IU was correctly recalled. The two raters scored 30% of the recall protocols independently. Since the inter-rater reliability was sufficiently high (r = .97), the remaining 70% of the protocols were scored by the author. Production rate data was calculated according to how many IUs out of the total data were recalled. This data was analyzed by a 2 (thematic target: appropriate vs. inappropriate) × 2 (contextual constraint: high CC vs. low CC) × 2 (proficiency: higher vs. lower) three-way mixed ANOVA.

Finally, for analyzing the data of the appropriateness rating task, a 2 (thematic target: appropriate vs. inappropriate) × 2 (contextual constraint: high CC vs. low CC) × 2 (proficiency: higher vs. lower) three-way mixed ANOVA was performed in a similar way. The above analysis was necessary in order to investigate how the participants perceive the relationship between the thematic inference target and the context of the passages.

[4] Correct recognition time is the time which participants could judge whether the knowledge-based target words were a word or a nonword correctly.

3.2.2 Results

3.2.2.1 L2 Proficiency Test

In order to confirm the reliability of the L2 proficiency test, the Eiken Test, Cronbach's alpha reliability was calculated. The reliability of the test was sufficiently high (Cronback's coefficient alphas was .82).

Table 3.3 shows the descriptive statistics for the L2 reading proficiency test. On the basis of their English test scores, 39 participants were divided into two proficiency groups: an upper group and lower group, t (36) = -7.96, $p < .001$, $d = 2.82$.

Table 3.3
Descriptive Statistics for L2 Reading Proficiency Test

Proficiency group	*n*	*M*	*SD*
Upper	20	26.95	1.99
Lower	19	20.35	2.71
Total	39	23.65	4.08

Note. A full score for the test was 34.

3.2.2.2 Lexical Decision Task

A lexical decision task was performed in order to investigate whether the participants activated the thematic inferences while reading. If the recognition time for topic-appropriate targets was shorter than that of topic-inappropriate targets, it proved that EFL learners activated the thematic inferences while reading (Hannon & Daneman, 1998; Long et al. 1994, 1997; Till et al. 1988).

The results of the lexical decision task are summarized in Table 3.4. The table shows the means and standard deviations of correct recognition time by CC (high CC, low CC), thematic target (appropriate, inappropriate) and proficiency (upper, lower). In addition, Figure 3.1

depicts a bar graph of the mean recognition time (*ms*).

Table 3.4

Descriptive Statistics for Recognition Time for Thematic Target Words

	High				Low			
	Appropriate		Inappropriate		Appropriate		Inappropriate	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Upper (n = 20)	847.96	217.98	1017.68	316.67	847.63	232.21	907.15	310.37
Lower (n = 19)	1200.96	379.61	1226.44	426.18	1021.14	330.18	1077.01	282.13
Total (N = 39)	1019.94	327.37	1119.38	383.93	932.17	283.38	989.90	315.61

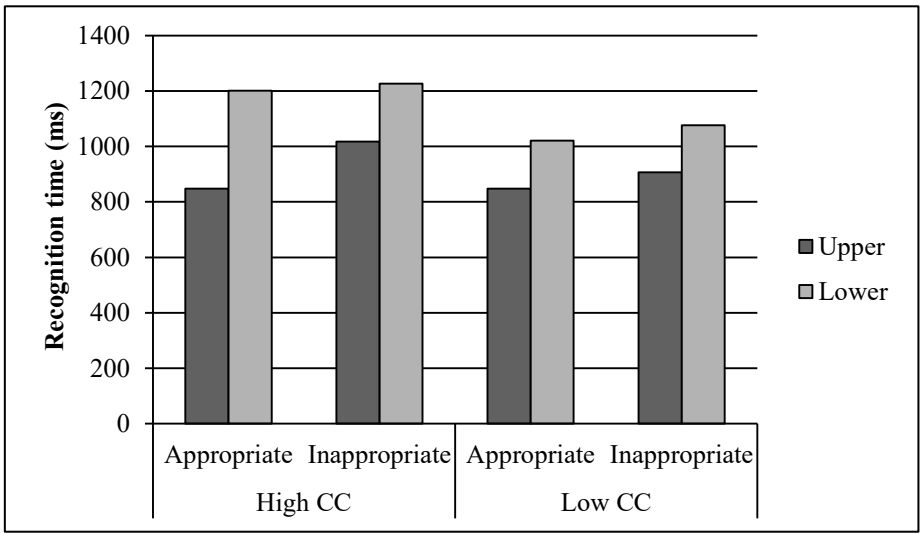


Figure 3.1. Recognition time in lexical decision task; CC= contextual constraint.

A three-way ANOVA was conducted to investigate the difference between appropriate and inappropriate targets in the correct recognition time by CC and proficiency, respectively. The following results were gained: the main effect of thematic target, F (1, 37) = 5.92, p = .042, η_p^2 = 107; CC, F (1, 37) = 24.07, p < .001, η_p^2 = 394; proficiency, F(1, 37) = 6.84, p = .013, η_p^2 = .156. Furthermore, a significant one-way interaction of CC × proficiency, F(1, 37) = 5.92, p

= .020, η_p^2 = .138 was shown. However, other interactions with thematic target and two-way interactions were not statistically significant (see Table 3.5).

Table 3.5

Results of Three-Way ANOVA for Recognition Time for Thematic Target Words

	SS	*df*	*MS*	*F*	*p*	η_p^2
Within-Subjects Effects						
Contextual Constraint (CC)	471832.33	1	471832.33	24.06	.000	.394
Thematic target	234946.33	1	234946.33	4.44	.042	.107
CC x Thematic target	15517.52	1	15517.52	.69	.410	.018
CC x Proficiency	116202.80	1	116202.80	5.92	.020	.138
Thematic target x Proficiency	53283.81	1	53283.81	1.01	.322	.027
CC x Thematic target x Proficiency	48132.36	1	48132.36	2.15	.150	.055
Between-Subjects Effects						
Proficiency	1995701.02	1	1995701.02	6.84	.013	.156

Note. η2 = effect size; 0.010 ≤ η2 (small) < 0.059, 0.059 ≤ η2 (medium) < 0.138, 0.138 ≤ η2 (large; Cohen, 1988, pp. 284-287); Effect size of 0.009 and below is regarded as having no effect.

From the results of the main effect of thematic target, it was found that the participants reacted significantly faster to the topic-appropriate target than they did to the topic-inappropriate target. This could be evidence that the EFL readers were activating thematic inferences while reading. On the other hand, as for CC, the recognition time for low CC was shorter than high CC. In terms of proficiency, the recognition time of the upper group was faster than that of the lower group. There was a significant interaction between the two factors. As the result of post hoc analyses, the recognition time of the lower group was shorter in low CC compared with high CC, but such difference was not found in the upper group. It should also be noted that there was no significant interaction between thematic target and the other two

factors (contextual constraint, proficiency).

The result indicates that the factor of CC did not have any effect on the upper group readers' access to thematic information, but rather high CC inhibited the lower group readers' access to the thematic target. That is to say, contrary to my expectation, the levels of CC did not affect the generation of EFL learners' thematic inferences in sentence-level reading. However, the obtained result shows that there is a possibility that the factor of CC did not function correctly in the reading comprehension of the lower group. Although the reading passages were allocated to each CC condition based on the pilot study, the difference of CC levels may not have been sufficiently perceived by lower group readers. Therefore, the appropriateness rating task conducted in Experiment 1 in order to confirm how learners in the lower group perceived the relationship between thematic target and passages (see 3.2.2.4 Results of Appropriateness Rating Task) is necessary. From the above findings, it is possible that the participants are generating thematic inferences regardless of their levels of proficiency or contextual constraints. This does not coincide with the assertions by Long et al., (1994) or Hannon and Daneman (1998) which state that only proficient readers generate thematic inference during reading. There is a possibility that the materials used by the above researchers were made easier for readers by excluding some passages which were difficult for them to generate inferences. However, in this study, it was found that the EFL readers were generating thematic inferences while reading.

3.2.2.3 Cued Recall Task

The amount of information that was produced as the result of cues of the thematic target shows how strongly thematic inference generated in the readers' minds is connected to the sentence information they have understood in the reading.

The results of cued recall tasks are summarized in Table 3.6. The table shows the means and standard deviations of recall production rates by CC (high CC, low CC), thematic target (appropriate, inappropriate) and proficiency (upper, lower). Furthermore, Figure 3.2 depicts a bar graph of the mean recall production rates (%).

As the results of a three-way ANOVA, the following results were gained: the main effect of thematic target, $F(1, 37) = 87.11$, $p < .001$, $\eta_p^2 = .702$; CC, $F(1, 37) = 23.46$, $p < .001$, $\eta_p^2 = .388$; proficiency, $F(1, 37) = 7.28$, $p = .009$, $\eta_p^2 = .164$. In addition, a significant one-way interaction of CC × thematic target showed the following result: $F(1, 37) = 4.76$, $p = .036$, $\eta_p^2 = .114$. However, other interactions with proficiency and two-way interactions were not statistically significant (see Table 3.7).

As for the main effect of proficiency, it was found that the upper group produced significantly more IUs than the lower group regardless of the CC levels or the types of thematic targets as cues. In addition, in respect to the interaction of CC and thematic target, a multiple comparison by Bonferroni correction clarified that the participants could more significantly recall the amount of IUs when they were given the topic-appropriate targets as cues than when they were given the topic-inappropriate targets.

Table 3.6
Descriptive Statistics for Cued Recall Rates

	High				Low			
	Appropriate		Inappropriate		Appropriate		Inappropriate	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Upper (n = 20)	.73	.14	.44	.20	.56	.18	.39	.23
Lower (n = 19)	.56	.21	.31	.20	.43	.25	.26	.22
Total (N = 39)	.65	.20	.38	.21	.50	.22	.32	.23

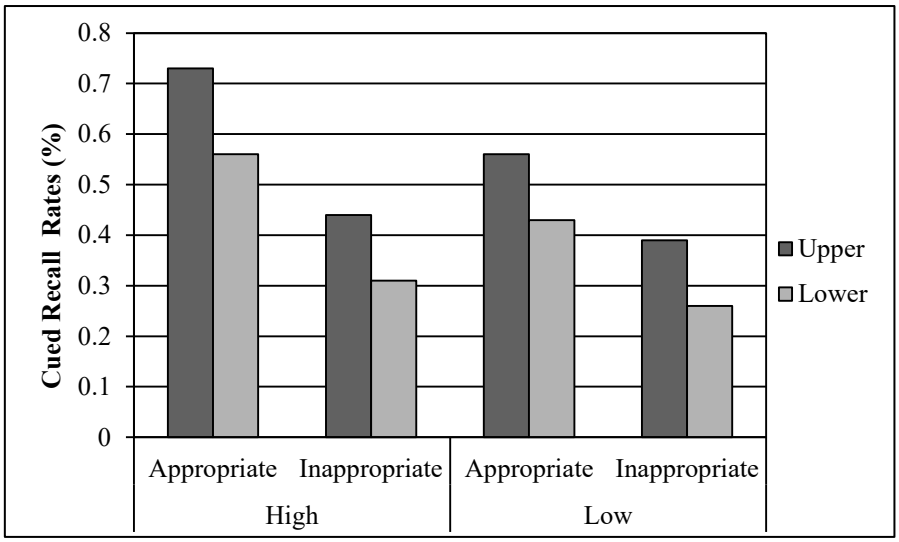


Figure 3.2. Cued recall rates for each condition.

Table 3.7
Results of Three-Way ANOVA for Cued Recall Rates

	SS	*df*	*MS*	*F*	*p*	η_p^2
Within-Subjects Effects						
Contextual Constraint (CC)	.41	1	.41	23.46	.000	.388
Thematic target	1.89	1	1.89	87.11	.000	.702
CC x Thematic target	.08	1	.08	4.76	.036	.114
CC x Proficiency	.01	1	.01	.46	.503	.012
Thematic target x Proficiency	.01	1	.01	.17	.687	.004
CC x Thematic target x Proficiency	.01	1	.01	.25	.620	.007
Between-Subjects Effects						
Proficiency	.79	1	.79	7.28	.010	.164

Note. η2 = effect size; 0.010 ≤ η2 (small) < 0.059, 0.059 ≤ η2 (medium) < 0.138, 0.138 ≤ η2 (large; Cohen, 1988, pp. 284-287); Effect size of 0.009 and below is regarded as having no effect.

The above result means the exact opposite of the result in the lexical decision task in which the recognition time for the thematic target was not affected by high CC. Namely, for EFL learners, the constraints of the passage are not especially important for activating the thematic inference while reading, whereas high CCs have a key role in recalling the passage information. This could be attributed to the fact that learners may draw much information from

their memory depending on the strong relationship between topic-appropriate targets and passages.

3.2.2.4 Appropriateness Rating Task

Appropriateness rating task was adopted to figure out how strategically the participants perceive the relationship between thematic information and passage in their mind. In particular, as stated above, the lower group readers' lexical decision time for thematic target in high CC condition was shorter than that compared with that in low CC condition. Therefore, it is necessary to examine the lower group readers' perception toward thematic targets and constraint of passages in their reading processes.

Table 3.8 shows the descriptive statistics for the appropriateness rating task and Figure 3.3 depicts a line graph of the mean rating score. The results of the three-way ANOVA toward rating data showed the main effect of thematic target, $F(1, 37) = 701.83$, $p < .001$, $\eta_p^2 = .95$; $F(1, 37) = 11.47$, $p = .002$, $\eta_p^2 = .237$. Furthermore, a significant one-way interaction of the thematic target × proficiency proved to be, $F(1, 37) = 4.76$, $p = .036$, $\eta_p^2 = .114$. However, other interactions with CC and two-way interactions were not statistically significant (see Table 3.9).

Table 3.8
Descriptive Statistics for Appropriateness Rating Task

	High				Low			
	Appropriate		Inappropriate		Appropriate		Inappropriate	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Upper (n = 20)	6.08	.83	1.93	.76	4.77	.87	1.75	.62
Lower (n = 19)	5.83	.93	2.55	.83	4.56	.99	2.11	.61
Total (N = 39)	5.96	.88	2.23	.85	4.66	.93	192	.64

Note. The appropriateness of the thematic target was rated on a 7-point Likert scale.

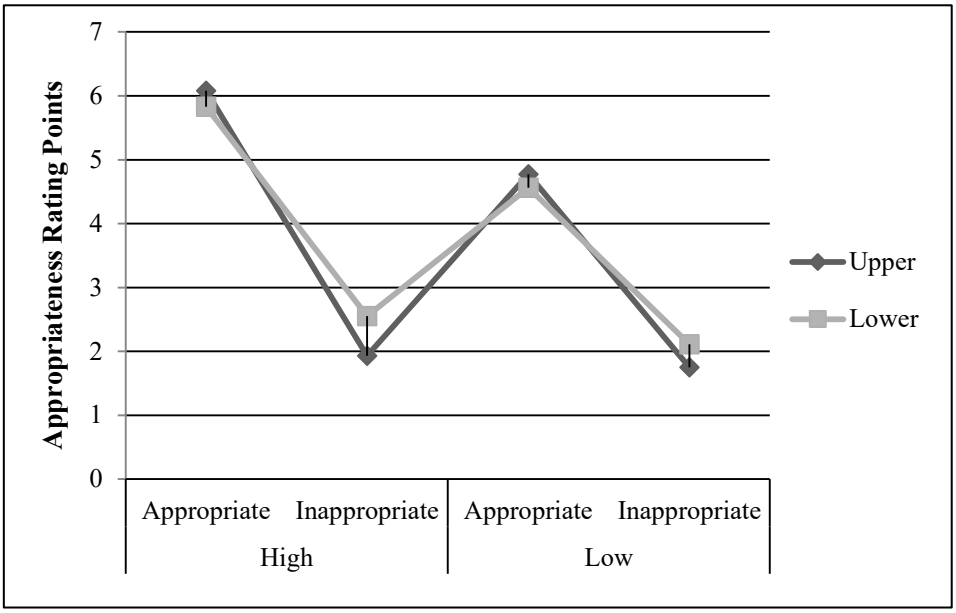


Figure 3.3. Mean points of appropriateness rating task.

Table 3.9
Results of Three-Way ANOVA for Appropriateness Rating Task

	SS	*df*	*MS*	*F*	*p*	η_p^2
Within-subject Effects						
Contextual Constraint (CC)	3.573	1	3.573	11.476	.002	.237
Thematic target	540.702	1	540.702	701.863	.000	.950
Proficiency	.620	1	.620	.479	.493	.013
CC × Thematic target	.003	1	.003	.010	.920	.000
CC × Proficiency	.135	1	.135	.433	.515	.012
Thematic target × Proficiency	4.976	1	4.976	6.459	.015	.149
CC × Thematic target × Proficiency	.234	1	.234	.767	.387	.020
Between-subject Effects						
Proficiency	.620	1	.620	.479	.493	.013

Note. η2 = effect size; 0.010 ≤ η2 (small) < 0.059, 0.059 ≤ η2 (medium) < 0.138, 0.138 ≤ η2 (large; Cohen, 1988, pp. 284-287); Effect size of 0.009 and below is regarded as having no effect.

The results of the two main effects in thematic targets and CCs showed that the participants rated the topic-appropriate target higher than the inappropriate target and also rated

high CC conditions higher than low CC conditions. This fact indicates that these two factors functioned well in the reading of Experiment 1. Of course, since the perception by readers was drawn strategically by rating task, this participants' perception cannot exactly be identified with their actual perception during reading. As for the significant interaction of the thematic target × proficiency, a multiple comparison by Bonferroni comparison revealed that the lower group rated topic-inappropriate targets higher than the upper group did. This could be interpreted that for readers who had high L2 reading proficiency, their sensitivity toward thematic information led from passages were more accurate.

3.2.3 Discussion for Experiment 1

In Experiment 1, I examined the generation of thematic inferences in sentence-level reading. I also investigated the recognition time in the lexical decision task to see inference activation while reading. Furthermore, I looked at cued recall tasks to find out how much text information is drawn based on thematic information after reading. Lastly, in the appropriateness rating task, I examined how readers recognize the relationship between thematic targets and passages.

Firstly, finding (a) is the answer to RQ 1-1 (Do proficient and less proficient EFL readers activate thematic inferences automatically in sentence-level reading?). The results of the above findings are as follows: (a) both proficient and less proficient EFL readers activate thematic inferences during sentence-level reading, (b) contextual constraint and L2 reading proficiency do not affect the activation of thematic inferences during reading, (c) EFL learners are building mental representations based on thematic information and the information that is strongly related with the passages is much recalled after reading, and (d) although EFL learners correctly recognized the topic of the passages, it was the proficient readers of the upper group who

responded more sensitively to the connection between passages and thematic information.

As for the result of recognition time for the lexical decision task, the upper group readers' 847 ms mean latency for topic-appropriate targets was 115 ms faster than their mean latency for topic-inappropriate targets. Furthermore, the lower group readers' 1110 ms mean latency for topic-appropriate targets was 41 ms faster than their mean latency for topic-inappropriate targets. In other words, both upper and lower group readers could activate thematic inferences during reading, although lower group readers were less proficient at this than upper group readers. This result is supportive of previous studies in that there is the activation of thematic inference while readers are conducting sentence reading. On the other hand, it did not support the finding of Long et al. (1994) and Hannon and Daneman (1998) that proficient readers generate thematic inferences, whereas the less proficient readers do not.

Secondly, finding (b) is the answer to RQ 1-2 (What kind of effects do contextual constraints have on the activation of thematic inferences?). When comparing the recognition time for topic-appropriate target, upper group readers' 847 ms mean latency for high CC condition was the same latency for low CC condition (847 ms). In addition, although lower group readers' recognition time had no significant difference between high CC (1200 ms) and low CC conditions (1021 ms), this reverse tendency in high CC and low CC contradicted my prediction about the CC effect. Since there was a possibility that the lower group readers' recognition of CC was not accurate, I conducted the appropriateness rating task. However, from the result of the task, it was found that both upper group readers and lower group readers correctly recognized that it was thematic target which had the strong connection with passages in high CC condition rather than in low CC condition. Therefore, the manipulation of contextual constraint was intended to improve the activation of inferences by narrowing down the potential inferences generated from context. However, is could be concluded that EFL learners are able

to activate thematic inferences regardless of the quality of context, i.e., the intensity of contextual constraint.

Thirdly, finding (c) is the answer to RQ 1-3 (How do thematic inferences and contextual constraints influence the construction of the readers' mental representation?). As a result of the cued recall task, readers could more significantly recall the amount of sentence information when they were given the topic-appropriate targets as cues than when they were given the topic-inappropriate targets. This is the opposite of the result in the lexical decision task in which the recognition time for the thematic target was not affected by contextual constraints. That is, the strong relationships between topic-appropriate targets and passages have an important role not for the activation of the thematic inference while reading, but for drawing much information from the readers' memory after reading. In other words, the factor of contextual constraints had an effect not on the encoding process, but on the retrieving process of thematic information.

This result is similar to that of Hannon and Daneman (1988). They showed that thematic inferences were promoted by using integrated questions which emphasized the relationship between the important words in the passage and encouraged the reader to draw a conclusion about the relationship between them. For example in order to generate thematic inference (e.g., burglar) for the short passage "The old woman awoke to a sound from downstairs. She reached into her purse and found only a file," it was necessary to fill in the gaps with integrated questions such as "The old woman awoke and said, '*why is there a sound downstairs?*' She reached into her purse and found only a file." In this example, integrated question encourages the readers to draw inferences from prior knowledge a 'why' conclusion about the relationship between the important words in the passage (e.g., because *burglar* stolen her money). In their experiment, integrated questions helped less proficient readers make knowledge-based inferences.

In cued recall tasks in Experiment 1, I have given such manipulated questions, but EFL

learners had to recall why such inferences were drawn when they accessed passage information taking the cues from topic-appropriate targets. In order to do so, it is assumed that the strong relationship between inference targets and passages played important roles when retrieving passage information in readers' mental representation.

Lastly, finding (d) is the answer to RQ 1-4 (How sensitive are proficient and less proficient readers' perceptions of the relationship between thematic inferences and passage information). From the result of the appropriateness rating task, both upper group readers and lower group readers recognized the relationship between thematic inferences and passage information correctly. However, the interaction of the thematic target × proficiency revealed that the lower group readers rated topic-inappropriate targets higher than the upper group readers. This could be interpreted that for readers who had high L2 reading proficiency, their sensitivity toward thematic information retrieved from passages was more accurate than that of readers with lower L2 proficiency.

3.3 Summary of Study 1

In Study 1 (Chapter 3), it was examined the effect of contextual constraints and L2 reading proficiency on thematic inference generation and explicit text comprehension in sentence-level reading. Consequently, two main findings were indicated as follows:

1. When constructing representations with short passages, it is not necessary to integrate information that has been separated, and the cognitive load involved in low-level processing is low. For that reason, appropriate thematic inferences were generated from the information in the passages without the effects of constraints or levels of proficiency.
2. In information reproduction prompted by inference words obtained after reading, the most information was reproduced when the constraints were strong, but it was found that the

constraint factors affected not the generation of inferences but rather the process of searching text information in long-term memory.

As mentioned above, I have clarified the generation of thematic inference in EFL learners' sentence-level reading, and the how CC as well as L2 reading proficiency affect the generation process and the cognition of readers. In Experiment 1, I have used short passages which are comprised of two sentences. However, when reading longer passages, the process is to integrate more information such as the background of the story, the protagonist's goal, actions or their emotions. As such, there is a possibility that inference generation will become more difficult for EFL learners compared with sentence-level reading. Therefore, in the next Study 2, I shall investigate the generation of thematic inferences in text-level reading.

Chapter 4

Study 2: Generation of Knowledge-Based Inferences in Text-Level Reading

4.1 The Purpose of Study 2

The results of Experiment 1 suggested that both proficient and less proficient EFL readers have reasonably accurate propositional representation and generate thematic inferences. However, do they also construct accurate representations of longer passages in text-level reading? It refers to the ability to integrate propositions that are relatively distant in the surface structure of a story and the readers' ability to make inferences about story topics. Cain et al. (2001) examined proficient and less proficient readers' representations of connected discourse and found significant differences. In particular, they found that less proficient readers failed to integrate information from different parts of a text. That is because such inferences in text-level reading are often based on information about the text as a whole. If readers fail to build integrated text representations, then they are unlikely to generate knowledge-based inferences. The purpose of Study 2 was to examine EFL readers' ability to integrate story ideas and to make knowledge-based inferences.

Moreover, Study 2 particularly focus on predictive inferences as one of knowledge-based inferences. These inferences anticipate the subsequent information that is likely to occur in the text, characterized as "what-happens-next" inferences (Allbritton, 2004; Iseki, 2003). They allow readers to go beyond the text and presumably construct a richer understanding of a narrative (Allbritton, 2004). These inferences are important in reading comprehension for the following three reasons. First, to generate predictive inferences, readers need to connect text events with general world knowledge and it allows readers to predict what is likely to occur next. This connection is essential to building situation models of texts (Fincher-Kiefer, 1996;

Linderholm, 2002). Second, the generation of predictive inferences helps to activate the processing of future text events and enables future text to be processed more smoothly. Third, there are some classes of predictive inferences that must be generated to maintain text coherence during reading (Klin, Murray, Leviene, & Guzman, 1999; Murray, Klin, & Myers, 1993).

As with Study 1, Study 2 also aimed to clarify the effect of text factor and learner factor on generation of predictive inferences. Study 2 was composed of Experiment 2 and 3. Experiment 2 was designed to test whether EFL learners could generate predictive inferences on the basis of contextual features which has been strongly biased (context constraints), and whether L2 reading proficiency influences their inference generation. To observe whether generated predictive inferences were encoded into the learners' long-term memory representation of the text, a recognition task was employed. Also, to observe what explicit text information they finally understood from a text, a cued recall task was used. Both of the tasks are categorized into the off-line method.

Experiment 3 was designed to investigate the effect of contextual constraint on the specificity of predictive inferences. So as to promote their predictive inferences intentionally, learners were provided a inference evoking question: "What will happen next" after reading a passage, and required to write down the story outcome freely according to their prediction. Then, their open-ended answers to the question were categorized from the standpoints of whether the contents were general or specific. In addition, it is examined readers' long-term memory representations after they read a passage when they were instructed to generate predictive inferences through the inference generation task. As a off-line approach, inference generation task and cued recall task were employed in Experiment 3. Figure 4.1 depicts relationships among factors in the reading processes assessed in Experiment 2 and 3.

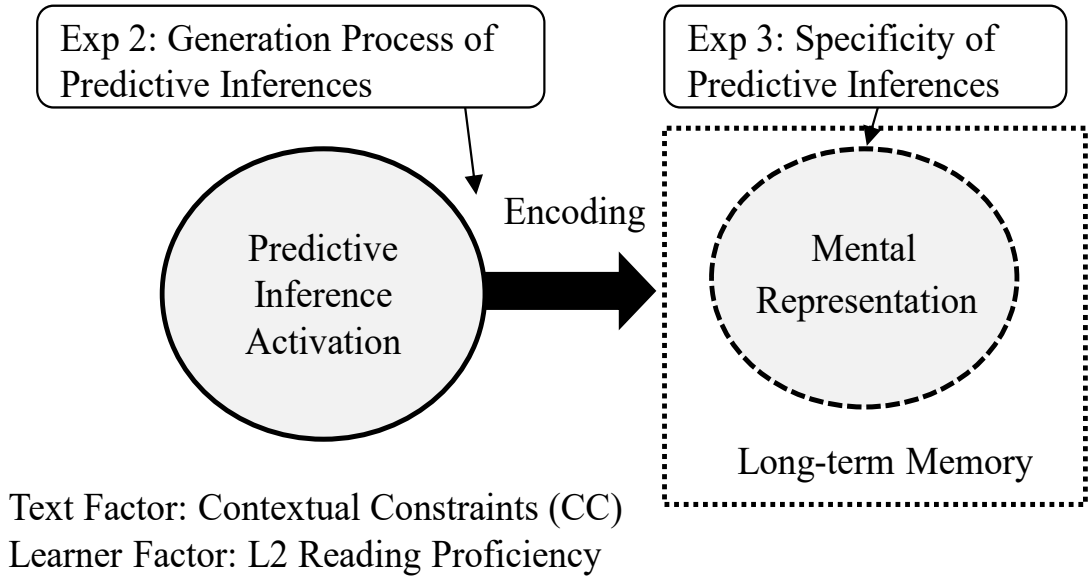


Figure4.1. Association chart for Experiments 2 and 3.

4.2 Experiment 2: Effects of Contextual Constraint and L2 Reading Proficiency on Predictive Inference Generation

The aim of Experiment 2 was to investigate the generation process of predictive inferences on the basis of contextual features in EFL learners' reading comprehension. I also aimed to clarify the effects of their L2 reading proficiency on the reading processes. That is because, most studies that have shown the generation of predictive inferences have concentrated on the L1 field (e.g., Calvo et al. 1999; Fincher-Kiefer, 1995, 1996; Keefe & McDaniel, 1993; Klin et al., 1999; Murray et al, 1993). As such, little attention has been paid to readers' proficiency levels (e.g., Murray & Burke, 2003; Horiba, 1996, 2000). In an EFL reading context, it is possible that unproficient readers may be unable to perform higher-level processing such as inference because their lower-level reading processes (e.g., word decoding or syntactic analysis) require most of their' cognitive resources. Research questions were set as follows:

RQ2-1: How do contextual constraints and EFL learners' L2 reading proficiency affect the encoding of predictive inferences into mental representation?

RQ2-2: How do contextual constraints and EFL learners' L2 reading proficiency affect the comprehension of explicit text information in their long-term memory?

4.2.1 Method

4.2.1.1 Participants

A total of 45 EFL Japanese college students participated in the experiment. They were first-year students at a national university with diverse majors. Most had received formal education in English for six and a half years in Japan. Three students were excluded from the data analysis because their reading times were outliers, resulting in 42 students in the analysis.

4.2.1.2 Materials

(1) *L2 Reading Proficiency Test*

The L2 reading proficiency test was the same as Experiment 1. All questions in the L2 reading proficiency test were adopted from the previously administered the Society for Testing English Proficiency (STEP) pre-Grade 1 (14 items), Grade 2 (17 items), and pre-Grade 2 (3 items) tests (2007). The number of items was 34, and only the reading comprehension section was used.

(2) *Reading Materials*

The materials were two narrative passages revised from Klin et al. (1999) and Cook, Limber, & O'Brien, (2001), which had different levels of CC, as shown in previous studies. Each passage began with an introduction of the protagonist and a background of seven to eight

sentences. This was followed by an elaboration section containing contextual constraints embedded in about seven to eight sentences. The contextual constraints sentences are the revised part that differentiates the two versions. It was manipulated to minimize the difference in the number of words between the texts and versions.

The contexts varied in the extent to which, in combination with later information, they would lead to the generation of specific predictive inferences. For example, in the text of Jimmy's story (see Table 4.1), Jimmy and his friends are playing a game of throwing objects at the target for points. The objects are either 'rocks' in high CC condition or 'sponge ball' in low CC condition. Also, the text in high CC condition emphasized in the elaboration section that the car is made of vulnerable materials. In the low contextual constraints version of the text, an episode about a neighbor is inserted instead. The last inference-evoking sentence states that Jimmy missed the target and accidentally hit the door of a new car. Neither the context nor the inference-evoking sentence contained the phrase "damaged it." However, if contextual features contained within the situation model can activate the predictive inference, the specific outcome would be predict in high CC condition.

In the same way, the second text, Brad's story was varied in the extent to which, in combination with later information, they would lead to the generation of specific predictive inferences. The outline of the story is as follows: the protagonist, Brad goes to buy a birthday present for his wife, but he is unable to buy the ring he really wants because he is poor. The text in high CC condition emphasizes that he has no money because he was laid off three months ago. By contrast, the low CC condition has an episode inserted instead that he has only a little money as he has just started his new job recently, and that another accessory is also on display at more reasonable prices. That is, although neither the context nor the inference-evoking sentence contained the phrase "steal the ring", the texts in high CC condition were

intended to draw the specific predictive inferences than that of the texts in low CC condition.

Table 4.2 shows the number of words and readability of the passages used in Experiment 2. Some difficult and unfamiliar words for EFL learners were revised on the basis of the *JACET List of 8000 Basic Words* (JACET Basic Words Revision Committee, 2003).

Table 4.1

An Example of Material Used in Study 2 (Part of Each Passages)

High CC condition
Jimmy was delighted and ran across the street to play with them. They taught him a fun game that involved throwing *rocks* at a target to get points.
Low CC condition
Jimmy was delighted and ran across the street to play with them. They taught him a fun game that involved throwing *sponge balls* at a target to get points.
Inference-Evoking Sentence
Jimmy missed the target and accidentally hit the door of a new car.

Table 4.2

Number of Words and Readability of the Passages in Experiment 2

Text	CC	Words	Sentences	Reading Ease	FKGL
Jimmy's story	High	201	16	85.9	4.3
	Low	202	15	85.1	4.6
Brad's story	High	212	17	77.2	5.5
	Low	208	17	78.4	5.3

Note. The Readability was calculated using Microsoft word 2007; CC = contextual constraints; FKGL = Flesch-Kincaid grade level.

Pilot Study

First, as a pilot study, a rating task was conducted to ensure that the contexts created a high or low likelihood for the predicted outcome. 12 university students were participated in a the pilot study. They were given the contexts and subsequent outcomes which were composed

of a single phrase for both high or low CC conditions. The participants were asked to rate the likelihood of an outcome on a 7-point scale from (1) *very unlikely to occur* to (5) *very likely to occur*. The mean ratings for low- and high-CC conditions were 2.38 and 6.72, respectively. As a result of two-way analysis of variance (ANOVA), there was a significant main effect of context, $F(1, 20) = 51.32$, $p < .001$. On the basis of the result, it was confirmed the appropriate materials were allocated to high and low CC conditions.

Next, reading times for inference-evoking sentences were measured to ensure if the materials function as the texts lead predictive inference activation in EFL learners' reading. The materials were modified according to the methodology by Cook et al., (2001). For half the passages, the inference-evoking sentence was written to explicitly state the concept representing the predicted event (i.e., explicit condition). For the text to be used in the explicit condition, a target inference in two to three words was added to the inference-evoking sentence (e.g., "He accidentally hit the door of a new car *and damaged it*"). The other half of the passages, in implicit condition, were not added any information about the predicted event (e.g., "He accidentally hit the door of a new car _____").

If a specific predictive inference is activated during reading in the implicit condition, then the result of the mental representation should be the same as the mental representation when the inference was explicitly stated in the explicit condition. Thereby, the data obtained from the above reading conditions could be interpreted as follows: the reading times for the inference-evoking sentence in the implicit condition should not differ from in the explicit condition. Meanwhile, if predictive inference is not activated during reading, the reading time in the implicit condition should be longer than in explicit condition. Following CC and explicitness conditions, four texts were assigned four patterns, namely 2 (CC: high vs. low) × 2 (explicitness: explicit vs. implicit) (see Appendix 3).

Participants read the two passages, which were presented in a counterbalanced order. A single sentence was displayed on the screen, and the next sentence was displayed immediately after the participants pressed a Response Pad 730 button, which was the apparatus for recording the participants' reading times. They read the texts at their own pace, and the reading times for the inference-evoking sentences were recorded with SuperLab 4.0 for Windows. In scoring and analysis, the length of inference-evoking sentence differed between the texts (Jimmy's story: 13 words, 20 syllables; Brad's story: 15 words, 24 syllables); hence, each reading time was divided by the number of syllable sounds for uniformity. These results were analyzed statistically using a repeated-measures two-way ANOVA, with the factors CC (high vs. low) and explicitness (explicit vs. implicit). Both CC and explicitness of the inference-evoking sentence were treated as within-subjects effects.

Table 4.3 and Figure 4.2 shows mean reading times for inference-evoking sentences. Overall, the reading time tended to be shorter in conditions where the CC was high or the inference content was explicit. The ANOVA showed no significant main effect of either CC, F (1, 11) = 1.06, p = .325, η^2 = .087) or explicitness, F (1, 11) = 3.27, p = .098, η^2 = .229). The interaction between CC and explicitness was also not significant, F (1, 11) = .146, p = .710, η^2 = .013). From the above results, which showed neither significant interaction nor main effects of explicitness and CC factors, it means that the participants read the implicit and explicit target sentences at a similar speed in both CC conditions, implying that they predicted the story outcome from the text context during reading. That is, predictive inferences were activated during reading regardless of the level of CC conditions, and it was ensured that the materials function as the texts lead predictive inference activation in EFL learners' reading.

Table 4.3

Mean Reading Times (ms) for Inference-Evoking Sentence (N = 12)

Explicitness	Contextual Constraints			
	High		Low	
	M	*SD*	*M*	*SD*
Explicit	434.21	205.48	496.59	243.07
Implicit	517.49	309.09	632.91	299.88

Note. Each reading time was divided by the number of syllables for uniformity.

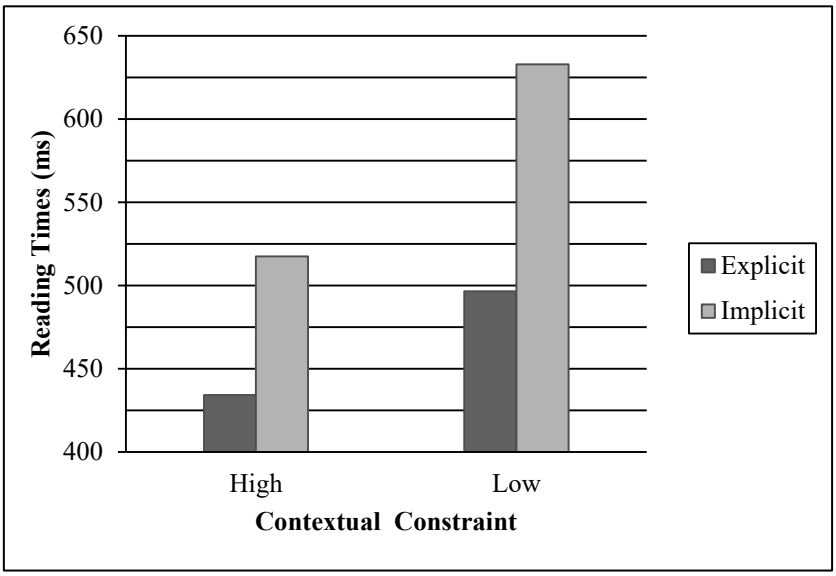


Figure 4.2. Reading times (*ms*) for inference-evoking sentences.

(3) *Probe Sentences for Probe Recognition Task*

As for a probe recognition task, following three types of probe sentences were created and participants' recognition (i.e., recognition time, yes response rates) were compared.

(a) Explicit probes: sentences that sentences as they appeared in the text

(b) Inference probes: sentences that are not explicitly mentioned in the text but can be inferred by readers from the surrounding context.

(c) Control probes: sentences that are not explicitly stated in the text nor can be inferred from the surrounding context.

In recognition time, a small difference between (1) explicit and (2) inference probes is regarded as an indicator of how strongly the implicit inference information is encoded into readers' representation, compared to explicit textbase information. Similarly, if there is the big difference between (1) explicit and (3) control probes, it is an index of the construction of accurate propositional textbase because readers can distinguish explicit from unstated information.

Out of these, most important is the difference between (2) inference and (3) control probes is regarded as an index of the construction of a situation model because readers discriminate the information that is true in the situation context from the information that does not exist in the context. As an evidence of encoding inference information into the readers' mental representation, the *delayed* recognition time for inferential stimulus was used. If inferential information is encoded into the readers' mental representation, then a correct response (i.e., "No" response) would be slow by relative inhibition effect. Fincher-Kiefer (1995, 1996) explained that relative inhibition effect (i.e., a delay in recognizing inference target) is caused by confusion about the source of the memory for the target word. That is, at the time of the recognition, readers experience difficulty in distinguishing inferential information from the reading and is part of the representation of the text's meaning. Therefore, such a delay in recognizing a target is regarded as evidence of inference encoding into mental representation.

Furthermore, previous research regarded Yes response to inference probes as an index that showed that a reader encoded inference into their understanding, although the correct reaction to the inference probe was actually No. For example, in L2 research, Muramoto (2000) investigated the effects of L2 proficiency on the recognition among explicit, inference, and control probes and showed that the Yes response of the explicit probes was higher than that of the inference and control probes, and that of the inference probes was higher than that of the

control probes. In addition, in inference probes, the Yes response of upper proficient readers was significantly higher than that of lower proficient readers. It means that upper groups generated more inferences than lower groups.

4.2.1.3 Procedure

The experiment was conducted through three phases. In the first phase, participants took the L2 reading proficiency test, with a time limit of 30 minutes. They were handed booklets in which they wrote down the answers.

In the second phase, the online passage reading and probe recognition task were administered. The participants read one text for practice and also checked how to operate the equipment. There were four experimental texts. The participants undertook self-paced reading on the computer screen, and their reading time on inference-evoking sentence was recorded by SuperLab 4.0, an apparatus that measures the reading time. Before starting, the instruction, "From now on, the experimental text will be displayed" was shown on the computer screen. The text appeared on the screen sentence-by-sentence, and subsequent sentences were displayed when the participant pressed the red key, which served as the advance key for SuperLab 4.0. As the participants finished reading the last inference-evoking sentence, the mark "***" was shown for 500 ms, and then the probe sentence was displayed and the probe recognition task was conducted. The participants were required to answer whether the probe sentence was in the text or not as quickly as possible by pushing the *yes* or *no* key. The precise instructions were as follows: "You will read statements after reading two passages in succession, and you will have to answer whether each statement was or was not in the text that you read by pressing either the YES or NO key." Thus, the recognition time and the percentage of questions answered correctly were measured. The two texts and the order in which the probes were

displayed were counterbalanced.

Finally, as the third phase, the cued recall task was conducted. The participants were given two pieces of paper on which only inference probe statements were printed as a cue. Then, the participants were instructed to write down in Japanese everything that they could remember about the two reading texts. No time limit was set for this task, but it took about 20 minutes for the participants to perform it.

4.2.1.4 Scoring and Data Analysis

First, the data that exceeded ±2.5 *SD* from the mean were deleted from the results for recognition times for the probe sentence. Thereafter, the remaining recognition time for probe sentences were divided by the number of syllable sounds for uniformity. Then yes reaction rates and recognition time were analyzed statistically with a 2 (CC: high vs. low) × 2 (proficiency: upper vs. lower) × 3 (probe types: explicit vs. inference vs. control) three-way ANOVA. Both CC and probe types were treated as within-subjects variable, and the proficiency level was a between-subjects variable. Since the each recognition probe had a different number of words between texts and probe types, the recognition time per syllable were calculated in order to analysis.

Second, the cued recall task data were scored with respect to clauses in the texts because the participants read the text by clauses in the experiment. A point was awarded when the participant provided two-thirds of the information in each clause. The scoring of recall was evaluated by two raters. The inter-rater reliability was high, $r = 90.55$ when 30% of items were scored by two raters. Thus, the remaining 70% of the recall protocol was scored by only one rater. Recall rates were analyzed statistically with a 2 (CC: high vs. low) × 2 (proficiency: upper vs. lower) two-way ANOVA.

4.2.2 Results

4.2.2.1 L2 Reading Proficiency Test

In order to confirm the reliability of the L2 proficiency test, the Eiken Test, Cronbach's alpha reliability was calculated. The reliability of the test was sufficiently high when one item was deleted (Cronback's coefficient alphas was .81).

Table 4.4 shows the descriptive statistics for the L2 reading proficiency test. The participants were divided into two groups on the basis of their scores on the L2 reading proficiency test, t (40) = 8.57, $p < .001$. As a result, the upper group consisted of 20 participants (M = 25.20, SD = 1.82), whereas the lower group consisted of 22 participants (M = 19.36, SD = 2.49).

Table 4.4
Descriptive Statistics for L2 Reading Proficiency Test

Proficiency group	*n*	*M*	*SD*
Upper	20	25.20	1.82
Lower	22	19.36	2.49
Total	42	22.28	3.61

Note. A full score for the test was 34.

4.2.2.2 Probe Recognition Task (Recognition Time)

The descriptive statistics for the recognition time for each type of probe are shown in Table 4.5. Also, Figure 4.3 shows the line-graph of the mean recognition time (*ms*) for the probe recognition task.

Obviously, mean recognition time for the inference probe in high CC condition seemed to be longer than the other probe types. As for the results of the ANOVA, the main effects of the CC, F (1, 40) = 18.12, $p < .001$, $\eta^2 = .312$, and the probe types, F (2, 80) = 43.16, $p < .001$, $\eta^2 = .519$, were both significant. Furthermore, both of the interactions of CC × proficiency, F

(1, 40) = 5.69, p = .022, η^2 = .125, and CC × probe type, F (2, 80) = 1206, p < .001, η^2 = .232 were significant. On the other hand, the main effect of the proficiency level and the interaction of the probe type × proficiency, the CC × probe type × proficiency were not significant at the .05 level.

Table 4.5

Descriptive Statistics for Recognition Time in the Probe Recognition Task

	Contextual constraints					
	High			Low		
	EXP	INF	CON	EXP	INF	CON
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Upper	382.19	734.33	456.79	393.20	473.18	304.45
(n = 20)	(193.81)	(261.49)	(195.14)	(133.20)	(103.81)	(118.70)
Lower	395.23	636.84	423.51	489.48	541.82	310.93
(n = 22)	(126.03)	(195.83)	(150.42)	(175.27)	(196.99)	(116.43)
Total	389.02	683.26	439.35	443.63	509.13	307.85
(N = 42)	(159.95)	(231.86)	(171.81)	(162.25)	(161.48)	(116.12)

Note. EXP = Explicit probe; INF = Inference probe; CON = Control probe.

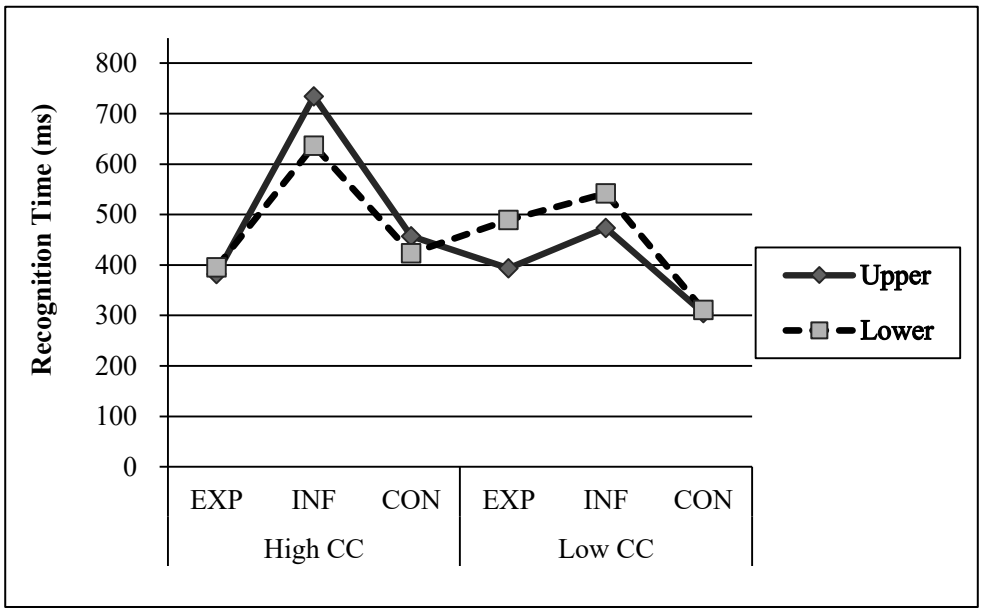


Figure 4.3. Recognition time (*ms*) for each probe in the recognition task;

EXP = Explicit probe, INF = Inference probe, CON = Control probe.

Table 4.6

Results of Three-Way ANOVA for Probe Recognition Time

Variable	*df*	*MS*	*F*	*p*	η2
Contextual Constraint (CC)	1	164010.42	18.12	.000	.312
Probe Type (PT)	1	871083.95	43.16	.000	.519
Proficiency (P)	2	31333.02	.55	.461	.014
CC x PT	2	610297.56	1206	.000	.232
CC x P	1	44633.23	6.69	.002	.125
PT x P	2	29244.47	.72	.490	.018
CC x PT x P	2	33510.70	.90	.409	.022
Within-cells error	80	37040.64			

Note. η2 = effect size; 0.010 ≤ η2 (small) < 0.059, 0.059 ≤ η2 (medium) < 0.138, 0.138 ≤ η2 (large; Cohen, 1988, pp. 284-287); Effect size of 0.009 and below is regarded as having no effect.

In order to examine the differences in both of the interactions, CC× proficiency and CC × probe type, multiple comparison analysis were conducted. First, as for the CC × proficiency interaction, it was found that compared to the low CC condition, the upper groups' recognition time was significantly delayed in the high CC condition than that of low CC condition, but the similar tendency was not attained from the lower groups' recognition time regardless of their CC level. Second, as for the CC × probe type, the results of the multiple comparisons showed that, in the high CC condition, the recognition time for the inference probe was significantly slower than that of control and explicit probe (inference > control ≒ explicit). On the other hand, in the low CC condition, the recognition time for the control probe was significantly delayed than that of inference and explicit probe, and there was no significant difference between inference and explicit probes (inference ≒ explicit > control). Taken together, these findings indicate that it is necessary that the specific outcome should be sufficiently constrained by CC for encoding process of predictive inference of the context. It was also found that the

sensitivity to the contextual constraint was different depending on the L2 reading proficiency levels of the readers. When CC was high, the upper groups' recognition time to the inference information was delayed due to the source confusion, since they had already encoded the inference information into their mental representation. Also, the upper groups' recognition time to the explicit information was faster than that of the lower group. It indicates high constrained text promoted the proficient readers to access the textbase information more easily than less proficient readers.

4.2.2.3 Probe Recognition Task ("Yes" Response Rates)

The descriptive statistics for "Yes" response rates in each type of probe sentence are shown in Table 4.7. It seemed that the rates gradually decreased in order from explicit, prediction, and control probes. Since the recognition rates of the control probe were below 20%, it was considered that the participants engaged in reading the text carefully.

Table 4.7

Descriptive Statistics for "Yes" Response Rates in Each Type of Probe

	Contextual constraints					
	High			Low		
	EXP	INF	CON	EXP	INF	CON
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Upper	.75	.35	.05	.75	.10	.00
(n = 20)	(.44)	(.48)	(.22)	(.44)	(.31)	(.00)
Lower	.77	.05	.00	.59	.05	.05
(n = 22)	(.43)	(.21)	(.00)	(.50)	(.21)	(.21)
Total	.76	.24	.02	.67	.07	.02
(N = 42)	(.43)	(.43)	(.15)	(.48)	(.26)	(.15)

Note. INF = Inference probe; CON = Control probe; EXP = Explicit probe.

The results of the factors 2 (CC: high vs. low) × 3 (probe types: inference vs. control vs. explicit), × 2 (proficiency: upper vs. lower) three-way ANOVA are shown in Figure 4.4 and Table 4.8. It was found that a significant main effect of CC, F (1, 40) = 4.68, p = .036, η^2 = 105, and probe types F (2, 80) = 95.63, p < .001, η^2 = 705. In addition, a marginally significant two-way interaction of CC × probe types × proficiency , F (2, 80) = 3.01, p = .055, η^2 = .071 was found. However, none of other main effects and interactions reached significance level.

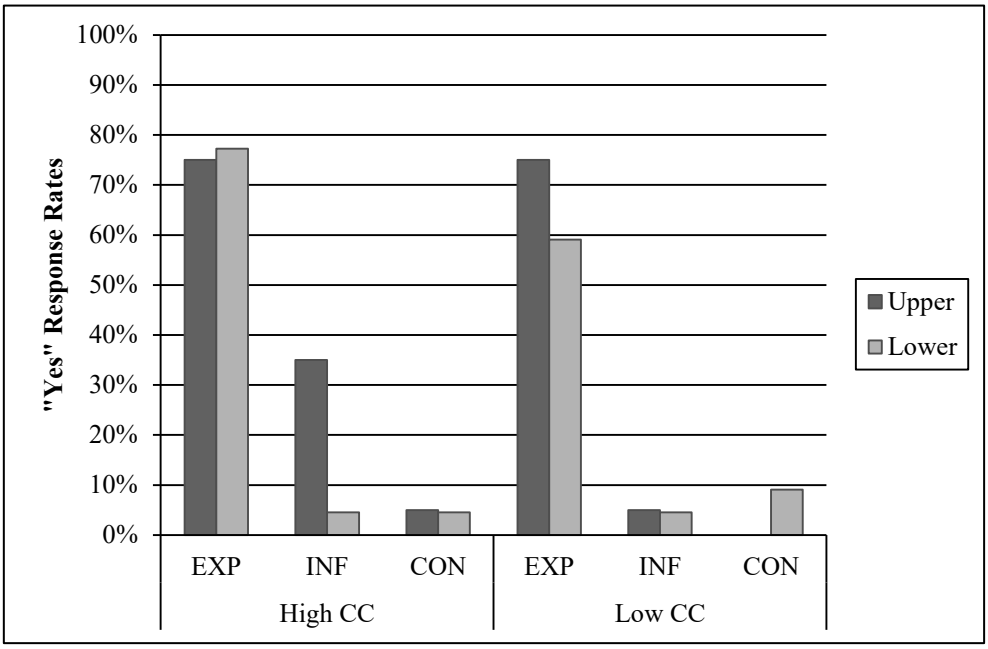


Figure 4.4. "Yes" response rates for each probes in the recognition task;
EXP = Explicit probe, INF = Inference probe, CON = Control probe.

Table 4.8

Results of Three-Way ANOVA for "Yes" Response Rates

Variable	*df*	*MS*	*F*	*p*	η2
Contextual Constraints (CC)	1	.02	4.68	.036	.105
Probe Type (PT)	1	197	95.63	.000	.705
Proficiency (P)	2	.00	.00	.974	.000
CC x PT	2	.08	.77	.467	.019
CC x P	1	.02	.14	.710	.004
PT x P	2	.41	2.25	.132	.053
CC x PT x P	2	.08	3.01	.055	.071
Within-cells error	80	.10			

Note. η2 = effect size; 0.010 ≤ η2 (small) < 0.059, 0.059 ≤ η2 (medium) < 0.138, 0.138 ≤ η2 (large; Cohen, 1988, pp. 284-287); effect size of 0.009 and below is regarded as having no effect.

In order to examine the differences in three factor interactions, multiple comparison were conducted. As a result, there was a significant difference between the upper group and lower group in the inference probe when contexts were highly constrained. Also, there was a significant difference between the high CC and low CC condition in the reaction rates for inference probe by upper group. Furthermore, the comparison of each probe types showed that the upper group responded to the inference probe (35%) more than the control probe (5%) when CC was high (control < inference < explicit). However, these two probes reaction rates (10%, 0%, respectively) were at the same level when CC was low (control ≒ inference < explicit). On the other hand, the lower group responded to the inference and control probes in a similar manner regardless of the CC level (control ≒ inference < explicit). Therefore, the evidence of predictive inference encoding was seen in the upper group for inference information only when the text was constrained, although explicit and control information were correctly recognized regardless of the L2 reading skills or the CC levels.

In sum, different from the inference activation as seen in Experiment 2, it was shown that the factor of readers' L2 proficiency levels and CC were closely related in the encoding process of predictive inferences.

4.2.2.4 Cued Recall Task

The descriptive statistics for the idea units (IUs) produced through cued recall protocols are shown in Table 4.9, and Figure 4.5 shows the mean percentage of IUs produced.

Table 4.9
Descriptive Statistics for Idea Units Produced by Cued Recall Protocols

	Contextual Constraints			
	High		Low	
	M	*SD*	*M*	*SD*
Upper (*n* = 20)	.51	.11	.40	.14
Lower (*n* = 22)	.47	.13	.24	.10
Total (*N* = 42)	.49	.12	.31	.14

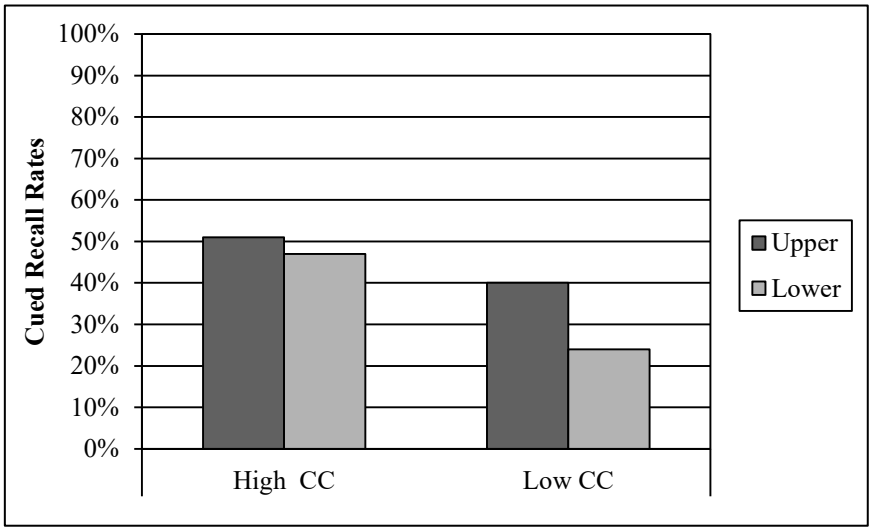


Figure 4.5. Cued recall rates for each condition.

The result of a 2 (CC: high vs. low) × 2 (proficiency: upper vs. lower) two-way ANOVA is represented in Table 4.10. The ANOVA found significant main effects of CC, F (1, 40) =

62.86, p = .001, η^2 = .611, and proficiency, F (1, 40) = 9.43, p = .004, η^2 = .191. There was also a significant one-way interaction of CC× proficiency, F (1, 40) = 4.67, p = .037, η^2 = .104.

Table 4.10
Results of Three-Way ANOVA for Cued Recall Performance

Variable	*df*	*MS*	*F*	*p*	η2
Contextual Constraints (CC)	1	.709	62.861	.000	.611
Proficiency (P)	1	.409	9.434	.004	.191
CC x P	1	.036	4.666	.037	.104
Within-cells error	40	.008			

Note. η2 = effect size; 0.010 ≤ η2 (small) < 0.059, 0.059 ≤ η2 (medium) < 0.138, 0.138 ≤ η2 (large; Cohen, 1988, pp. 284-287); effect size of 0.009 and below is regarded as having no effect.

The overall rate of production was higher in the upper proficiency group and in the high CC condition. Multiple comparisons revealed no significant differences between proficiency groups when CC was high. However, in the low CC condition, only the upper proficiency group could significantly (p < .05) recall much of the contents. These results indicate that by decreasing the contextual information which leads to specific outcome in the text, the lower proficiency readers were inhibited from creating elaborate representations.

4.2.3 Discussion for Experiment 2

The focus of Experiment 2 was to explore the extent to which L2 reading ability and CC is related to the encoding of predictive inferences in EFL learners' reading comprehension. From results of the probe recognition task, it was shown that the encoding of readers' predictive inferences was promoted due to both the EFL learners' L2 proficiency and text constraint by high CC. The results of recognition time indicated that it is necessary the specific outcome

should be sufficiently constrained by CC for encoding process of predictive inference of the context. It was also found that the sensitivity to the contextual constraint was different depending on the L2 reading proficiency levels of the readers. When CC was high, the upper groups' recognition time to the inference information was delayed due to the source confusion, since they had already encoded the inference information into their mental representation. Also, the upper groups' recognition time to the explicit information was faster than that of the lower group. It indicates high constrained text promoted the proficient readers to access the textbase information more easily than less proficient readers.

Moreover, yes response rates in recognition task gave the more clear evidence of the relation of the CC and L2 reading proficiency in inference encoding. The upper group responded to the inference probe more than the control probe when CC was high, however, these two probes' reaction rates were at the same level when CC was low. On the other hand, the lower group responded to the inference and control probes in a similar degree regardless of the CC level. Therefore, the evidence of predictive inference encoding was seen in the upper group as for inference information only when the text was constrained, although explicit and control information were correctly recognized regardless of the L2 reading skills or the CC levels. Also, explicit text information encoded in readers' long-term memory was investigated with the cued recall task. The results indicated that regardless of their proficiency levels, the high CC condition was recalled more than the low CC condition. This result clearly shows a strong influence of CC in constructing concrete text-representation, whereas the influence of the readers' skills is rather limited.

In sum, high contextual constraints promoted the inference encoding into participants' mental representation, and EFL learners' L2 reading proficiency was related to their sensitivity to the contextual constraint levels and encoding information to a mental representation. That is,

for EFL learners, the encoding of the predictive inferences into mental representation was clearly influenced by both the text and learner factors. Several previous studies have examined the influences on the encoding of predictive inferences, including reader-based factors such as reading comprehension skill, age, and working memory capacity (e.g., Linderholm, 2002). For example, Murray and Burke (2003) suggested that only highly proficient readers are able to automatically generate predictive inferences. They claimed that less proficient readers utilize poor-quality lexical representations, which lead to difficulties in word identification that, in turn, can affect higher-level inferential processes. Among "high-level" processing differences, one that appears especially important is an individual's ability to activate and accurately encode text information into mental representations. However, it contradicted the results of probe recognition task in Experiment 2, that is, participants had constructed the text representation properly, but the less proficient learners had not reached the stage to encode inference information into mental representations. On the contrary, the effect of CC as text-based factors showed the same tendency as many previous studies did. The activation and encoding process were promoted when EFL learners read the sufficient predictable text (i.e., high CC condition). The memory-based text-processing reading model (O'Brien, Rizzella, Albrecht, & Halleran, 1998) claims that more highly related information acts as a clue and therefore engenders stronger inference generation. In this study, it was presumed that the framework of the memory-based view would apply; that is, the higher the support, the more the inferential information would be able to resonate online. The results supported this hypothesis; CC affected the automatic online reading process. Moreover, this result also made it clear that CC has more influence on explicit text-based processing. This is evident from the fact that the high CC condition did shorten the reading time for explicit target sentences as well as the recognition time to access the explicit probe.

4.3 Experiment 3: Effects of Contextual Constraint on the Specificity of Predictive Inferences

The goal of Experiment 2 were to investigate the effects of contextual constraints on readers' predictive inferences in on-line and off-line processing. Experiment 3 examined readers' mental representations after they read a passage when they were instructed to generate predictive inferences through the inference generation task. Therefore, a methodology using predictive inferences question and cued recall tasks was employed as the off-line approach. In predictive inference question tasks, the inference qualities predicted by the readers as future developments were investigated. From the L1 studies, it has been found that in on-line reading with the higher CC inferences with specific inferences were generated; whereas with the lower CC, inferences with general inferences were inclined to be generated (Lassonde & O'Brien, 2009). Therefore, in this task, we investigated whether the inferences generated by the readers were general or specific from the standpoint of off-line processing. Also, in the cued recall task, the research was conducted so as to find out what kind of information was most likely to be retained in the readers' mental representation. The aim of the above research was to examine whether the information with higher CC would be retained in long term memory, and thus increasing the production output in the cued recall task (Casteel, 2007; Klin et al.,1999). In order to investigate these concerns, the following research questions were set.

RQ 3-1: Do the inference generation task and contextual constraints affect the specificity of the generated predictive inferences?

RQ 3-2: Does the difference in contextual constraints affect the information encoded in the reader's mental representation of the text?

4.3.1 Method

4.3.1.1 Participants

A total of 60 Japanese university students with diverse majors participated in Experiment 3. Those who failed to follow instructions were eliminated from the analysis. As a result, 54 participants were analyzed. In contrast to Experiment 2, participants of this experiment were not grouped based on their L2 reading proficiency. The reason was because since their average score of the reading part in TOEIC IP test was 196 points, which was comparable to A2 level in CEFR, all participants were categorized as less proficient L2 readers.

4.3.1.2 Materials

The materials were two narrative passages from Klin et al. (1999) and Lassonde and O'Brien (2009) same as Experiment 2. They were two versions for each passage: (a) a high contextual constraints version and (b) a low contextual constraints version. The first version strongly suggested a particular outcome in the story line. In the second version, another context was inserted so that inferable content was not suggested in order to avoid generating a specific inference.

4.3.1.3 Procedure

All the participants took the test at the same time. Each was handed a booklet consisting of passages, questions, and recall answer sheets. The participants were instructed to not turn the page before reading the comprehension text. Then, after reading the passage, they answered the predictive inferences questions (which asked readers what they think will happen in the story development) and questions to verify their understanding of the passage (e.g., Why did Jimmy delight when he was asked if he wanted to play with the neighborhood kids?).

This was followed by a cued recall task in which participants were given two pieces of paper on which only inference probe statements were printed as a cue. Then, the participants were asked to write in Japanese as much as possible about what they remembered of the text they had just read. The same procedure was repeated for the two passages. The distribution of the passages among the participants was adjusted so that every participant would read a high contextual constraints version for one passage and a low contextual constraints version for the other. The session lasted 45 minutes.

4.3.1.4 Scoring and Data Analysis

As for the predictive inferences question task, the participants' answers were categorized by content. The categorization was conducted mainly following three criteria:

(1) The participants generated the "specific" contents of the predictive inferences.

(2) The participants generated the "general" contents of the predictive inferences.

(3) The participants did not generate predictive inferences at all.

In Text A, the participants' answers were categorized into four, whereas the categorization was five in Text B. Category (a) indicates a specific type of inference, whereas category (b) indicates a general type of inference. On the other hand, since the answer patterns in Text B were more diverse than those of Text A, one more categorization has been added. In Text B, category (a) shows a specific type of inference, while (b) and (c) indicate general types of inference. Firstly, two raters scored 30% of answer data and inter-rater reliability was sufficiently high, $r = .97$. Therefore, remaining 70 % was scored by one of the raters.

As for the recall task, each passage was divided into idea units (IUs) based on Ikeno's (1996) criteria by two raters, and the inter-rater reliability was $r = .90$. When the participants wrote two-thirds of the information in each IU, one point was given. Firstly two raters scored

30% of recalled protocols and inter-rater reliability was sufficiently high, r = .94. Therefore, remaining 70 % was scored by one of the raters.

4.3.2 Results

4.3.2.1 Inference Question Task

Table 4.11 shows the number of responses divided into categories in each text. As for the results of chi-square test, the answering patterns of both texts varied according to the influence of the CC (Text A , χ^2 = 12.58, p = .006, and Text B, χ^2 = 13.55, p = .009).

Table 4.11
Number of Responses Divided Into Categories in Each Text.

Text A - Jimmy's story

Conditions	(a)	(b)	(c)	(d)
High contextual constraints	23	1	0	3
Low contextual constraints	14	9	3	1

Note. Responses were divided into following four categories: (a) Jimmy damages the car of friend's father, (b) Jimmy damages something, (c) no response, and (d) others.

Text B - Brad's story

Conditions	(a)	(b)	(c)	(d)	(e)
High contextual constraints	21	4	1	0	2
Low contextual constraints	8	10	7	1	1

Note. Responses were divided into following five categories: (a) Brad steals the ruby ring, (b) Brad steals something, (c) Brad buy another thing, (d) no response, and (e) others.

This categorization showed that the contextual constraints to text information made a difference in the tendencies of the answers provided by the readers. The texts with high contextual constraints showed that the readers strongly imaged specific inferential information. The texts with lower contextual constraints showed less specific answers, while more

ambiguous and general answers increased.

The target predictive inference in Text A was that Jimmy, the protagonist, damaged the door of the car belonging to a friend's father. A total of 23 participants out of 27 predicted the same ending in the storyline. For the text with low contextual constraints, several patterns were observed in the answers. Nine of the participants gave ambiguous answers, such as "he will damage something" or "the protagonist will be told off for doing something bad". In Text B, the target inference was that Brad, the protagonist, steals a ring from a department store. A total of 21 participants of 27 who read the high contextual constraints version gave the target inference for their answers. In the low contextual constraints condition, on the other hand, 10 people's answers contained the target inference in general terms, while seven participants gave answers that did not reflect the target inference.

4.3.2.2 Cued Recall Task

Table 4.12 and Figure 4.6 shows the descriptive statistics for cued recall rates of each IUs category. Since each category has a different number of IUs, recall is discussed in terms of the proportion of IUs.

Table 4.12
Descriptive Statistics for *Cued Recall Rates of Each IUs Category (N = 54)*

Recall protocols	Conditions	*M*	*SD*
(a) Entire passage IUs	High	.49	.23
	Low	.48	.25
(b) Contextual constraints IUs	High	.44	.30
	Low	.29	.23
(c) Inference-evoking sentence	High	.43	.41
	Low	.29	.33

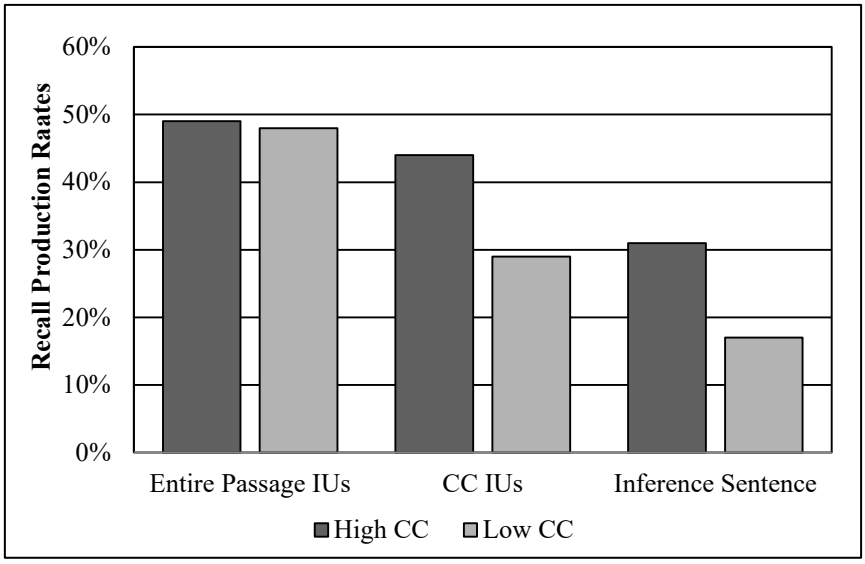


Figure 4.6. Cued recall rates for each IUs category.

These IUs were grouped into three categories. The first category included (a) the entire passage IUs except for the contextual constraints contexts and inference-evoking sentence. The second category consisted of (b) the contextual constraints IUs from the differently revised sections and between the two CC conditions. The third category consisted of (c) the inference-evoking sentence, which was the last line of each passage and lead readers to the inference.

In order to examine the effects of contextual constraints level on recall production, the paired *t*-test was conducted in three categories. The results showed that (a) there were hardly any differences in the entire passage IUs between the conditions, t (53) = .09, p = .930, and (b) more contextual constraints IUs were generated with high CC conditions, t (53) = 2.92, p =. 005. In addition, the more inference-evoking sentence information were generated significantly in the high CC condition than low CC condition, t (53) = 2.12, p = .038. The results showed that context information or inference-evoking sentence information that leads to suggest a particular ending and easier to be instantiated in the reader's mental representation.

4.3.3 Discussion for Experiment 3

This study examined the role of contextual constraints in the activation and instantiation process of predictive inferences in reading a text by EFL learners.

As for RQ3-1: Do the inference generation task and contextual constraints affect the specificity of the generated predictive inferences?, based on the qualitative analysis of the answers to questions regarding predictive inference, it was revealed readers tended to generate predictive inference with a more definitive and specific image when the contextual constraints was high, while they generated a more general and ambiguous image about the story development when the contextual constraints was low. With regard to RQ3-2: Does the difference in contextual constraints affect the information encoded in the reader's mental representation of the text?, the results of recall tasks after reading showed that the texts information concerning predictive inference with high contextual constraints tended to be kept in EFL readers' long-term memory and integrated into their mental representation.

Considering that off-line measures used in Experiment 3 reflect the reconstruction of text information (i.e., readers' understanding of a text after reading), on-line measures used in Experiment 2 reflect the incremental reading process (i.e., readers' generating predictive inferences during reading). It was found that EFL readers generated predictive inference after reading, influenced by the context, but predictive inference was not activated during reading. This finding is different from the "memory-based view processing" (McKoon & Ratcliff, 1992) in L1 research, which claimed that most inference information is obtained by activating passive concept on-line. This shows that the reading process of EFL learners is that new information is actively added by the reader to the text information off-line. As a limitation, in Experiment 3, inference question tasks were used to observe specific content of predictive inferences. Free response data for the task were classified by two evaluators, including the author. Although a

high inter-rater reliability was obtained (r = .97), responses should probably have been classified by evaluators other than the author so as to give the outcome better objectivity.

If an educational suggestion from this study is to be stated, the generation of predictive inference by learners is thought to be enhanced by going through an output process, by providing learners with a predictive inference question, or giving them a task of summarizing the content after reading. In addition, readers' ability to use predictive inference strategically is thought to be enhanced if the learners are instructed to predict the story development before and during reading, as appropriately.

4.3.4 Discussion and Analysis from Another Perspective

In this section, I will report the consideration about the recall data in Experiment 3 analyzed and examined from a difference viewpoints. So far, it was clarified the possibility that contextual constraints of the text greatly influences the reader's predictive inferences. However, there are two types of inferences among predictive inference; consequence predictive inferences and motivational predictive inferences (Murray et al., 1993; Klin, et al., 1999). Therefore, it is arguable points that whether or not the difference of predictive inference types affect readers' recall performance. The following points are not clarified yet. Do the different types of predictive inferences affect the readers' understanding of the text? Does the influence of contextual constraints differ according to the types of contextual constraints?

Klin et al. (1999) described the two types of predictive inferences as follows; consequence predictive inferences are simply the consequence of the events describing the story. When reading first example in Table 4.13, readers may generate the inference that the porcelain vase is going to break automatically. This is due to build up the activation of the inferred concept; breaking is related to delicate vases, to porcelain, to delicate things being thrown, and

to angry husbands. So breaking the porcelain vase is the highly probable consequence of the vase being thrown, and it serves to elaborate the learners' mental representations.

Table 4.13 *Two Subtypes of Predictive Inferences (Klin et al., 1999)*

(1) Consequence predictive inferences Steven had been married for years, and his resentment had been building up. One day, no longer able to control his anger, he threw a delicate porcelain vase against the wall. [inference stimulus: break]
(2) Motivational predictive inferences Brad had no money, but he just had to have the beautiful ruby ring for his wife. Seeing no salespeople around, he quietly made his way closer to the counter. [inference stimulus: steal]

On the other hand, motivational predictive inferences are generated from an understanding of the causal relationship between the character's motivation and the consequence. As shown in the second example in Table 4.13, it is not only the predicted action of stealing the ring. The inferred consequence of Brad walking quietly over to the counter is also the motivation or the cause. Brad is quietly making his way to the counter because he intends to steal the ring. Without the inference, the second sentence creates a break in the causal coherence of the passage; there is no cause for Brad's actions. Because these predictive inferences are needed to maintain the coherence of text, it has high possibility of being generated by the readers. Next, we would like to state the results of the examination of the above two types of predictive inferences in EFL learners' reading comprehension.

All of them read two types of texts (consequence predictive inference text vs. motivational predictive inference text). Then, immediately after reading the critical sentence which promotes inferences (e.g., the second line of Table 4.13, (2): Seeing no salespeople around, he quietly made his way closer to the counter), they were given a predictive inference

question immediately after they read the sentence. This task was aimed at observing what they predicted about an upcoming event in the text. Afterwards, a cued recall task was conducted requiring the participants to write down all the contents of the text which they could bring back to memory. This cued recall task reflected the readers' after reading textbase, and made it possible for us to learn how much the reader understood the text.

The materials used were two narrative passages (201-212 words) from Murray et al., (1999) and Lassonde and O'Brien (2009). They were partially modified for this study and two versions were prepared for each passage: (a) high contextual constraints version and (b) low contextual constraints version. The first version strongly suggested a particular outcome in the story line. In the second version, another context was inserted so that inferable content was not suggested in order to avoid generating a specific inference.

In the data analysis, the participants' recall performances were scored by idea units (IUs) text analysis. Table 4.14 shows that Descriptive statistics for the recall performances.

Table 4.14

Descriptive Statistics for Cued Recall Rates

Inference type	contextual constraints	*n*	*SD*	*M*
Consequence predictive inferences	High	27	.13	.33
	Low	27	.12	.34
	Total	54	.12	.34
Motivational predictive inferences	High	27	.15	.62
	Low	27	.15	.62
	Total	54	.15	.62

Note. This result shows the percentages of recalled IUs in the total text IUs.

Afterwards, statistical analysis was conducted using two-way ANOVA; 2 (inference types: consequence predictive inferences, motivational predictive inferences) x 2 (contextual

constraints: high contextual constraints, low contextual constraints) to compare the recall performance in the two types of predictive inferences. Figure 4.7 shows the proportion of recalled IUs in two types of inference conditions.

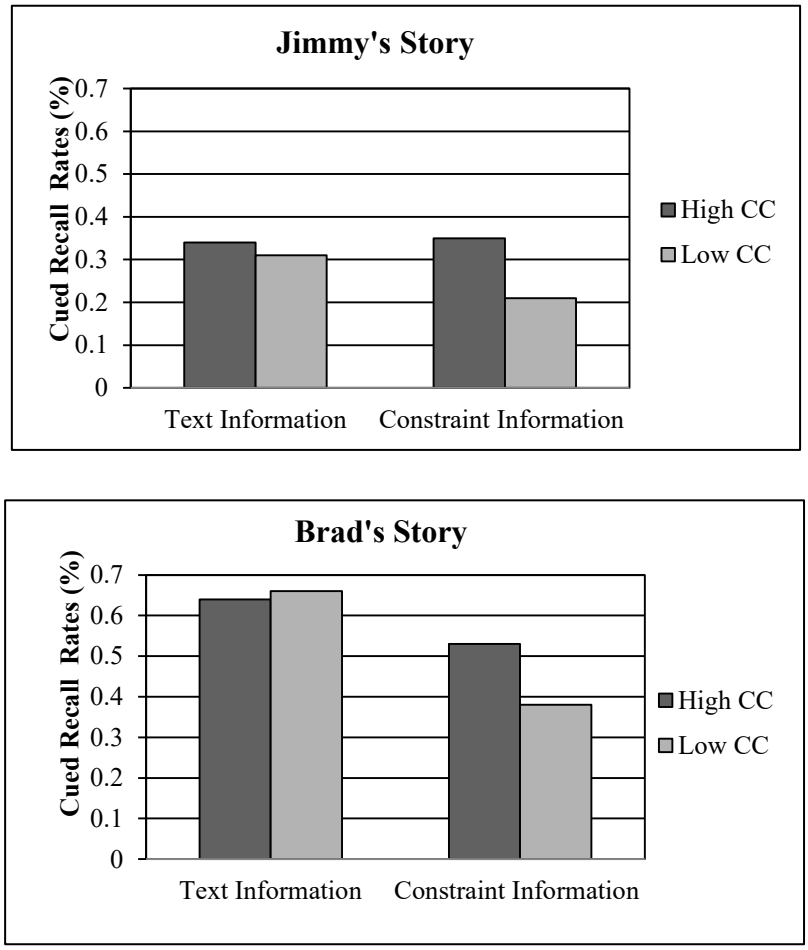


Figure 4.7. Proportion of recalled IUs in two types of inference conditions.

The results of ANOVA revealed that there was no interaction between inference type and contextual constraints, F (1, 52) = .03, p = .856, whereas the main effect of inference type was significant, F (1, 52) = 127.23, p = .001. The main effect of contextual constraints was insignificant F (1, 52) = .03, p = .866. As such, whether contextual constraints is high or low does not affect EFL learners' recall performances. However, in the texts with different

generative inferences, there was a big difference in productive rates (34% vs. 62%), which implies that the readers produce more information from motivational predictive texts.

The fact that there was a significant difference between recall performances under the condition of different inference types implies that there was the possibility of strong influence on readers' mental representations. According to the event-indexing model (E-I model) proposed by Zwaan, Magliano, and Graesser (1995), the readers construct a situation model from five independent dimensions such as time, space, protagonist, causality and intentionality during the process of reading. For example, if there is a big change in the protagonists' move or the elapse of time in the text, the situation model of "time" and "space" will be updated. Also, if the continuous information in each dimension is integrated, it will be easier to encode in situation model, thus making it able to be maintained in the readers' memory.

In the reading research prior to E-I model, only the viewpoint of each dimension of the situation model was considered. However, E-I model regards reading as the process of forming the network while relating propositions in each dimension as well as linking each dimension mutually. The texts used in this experiment were not so different in the four dimensions except for intentionality; "time, space, protagonist, causality." However, it is assumed that in "intentionality", especially in the case of motivational predictive inference text, the information as to why the protagonist performed the action which leads to the consequences is mutually related with other dimensions such as causality and protagonist consistently, thus contributing to encoding much information to mental representation. In addition, constructionist theory states that the inference is useful for maintaining coherence, and also claims the aspects concerning causality and intentionality are important (Graesser et al., 1994).

The results of the present study indicate that although the levels of contextual constraints do not influence mental representation after reading, the difference of inference types in the

texts affects the EFL learners' text comprehension. However, this is only the comparison of holistic recall performances and new findings could be discovered by examining the inference information in recall protocols. To sum up, it is important to clarify the relationship between EFL learners' inference process and characteristic features of the text.

4.4 Summary of Study 2

In Study 2 (Chapter 4), it was examined the effect of contextual constraints and L2 reading proficiency on predictive inference generation and explicit text comprehension in text-level reading. Consequently, three main findings were indicated as follows:

1. When strong predictions of outcome were made as a result of contextual constraints, predictive inferences were generated only among highly proficient readers. However, the inferential information that was generated was not as clearly encoded into representations as explicit information. In addition, it became clear that the stronger the contextual constraints, the more that explicit textual comprehension was facilitated, regardless of learner proficiency.
2. The higher the condition of the textual constraints, the more specific the conclusions were reflected in the content of the predictive inferences of the learner. Furthermore, a particularly high rate of recall of the text information was seen where the text had been manipulated to raise constraints. Because raising constraints meant reinforcing the causal coherence of the characters' goals and actions with the outcome (inferential information), it is thought that information regarding constraints was particularly strongly encoded into learners' text representations.
3. The fact that the types of predictive inferences that were generated differed depending on the text can be thought of as one factor for why there were variations in recall performance

across texts. It was found that motivational inferences, which are inferences drawn as a result of the causal links between the motivations of the main character and the consequences of his/her actions, were more readily generated than consequence inferences, which are simply drawn from the consequences of the events of the story.

Chapter 5

Study 3: Generation of Knowledge-Based Inference in Between-Texts-Level Reading

5.1 The Purpose of Study 3

Study 3 investigates how the text features work as access cues to the readers' knowledge structure using analogical texts. Analogical texts as advance reading organizer are used so as to strategically activate readers' prior knowledge. Showing an analogy activates the readers' schema related to sentence information, and understanding is then promoted by assimilating new information into the schema (Taniguchi, 1988). However, it was still unclear what kind of text information triggered activation of knowledge-based inferences in EFL reading.

5.2 Experiment 4: Effects of the Different Types of Similarities Between Analogical Texts on Analogical Transfer

The purpose of the Experiment 4 was twofold. The primary purpose was to investigate how much information (i.e., quantity) and what kind of information (i.e., quality) is facilitated through analogical transfer in EFL learners' reading comprehension. The secondary purpose was to reveal how similarity features such as the similarity types and causality of texts influence the process of analogical transfer (i.e., accessing, mapping, and encoding). The present study makes use of material in which the number of surface and structural similarities has been systematically manipulated, and incorporates participants' L2 reading proficiency into the experiment design as a between-subject factor. Furthermore, following the methodology of past research (Blanchette & Dunbar, 2002; Gentner et al., 1993), the processes of accessing and mapping in analogical transfer are explored using an immediate recognition task, and the encoding to text representation process is examined using a cued recall task. The following two

research questions (RQs) were posed:

RQ4-1: Do the quantity and quality of information facilitated by analogical transfer differ according to EFL learners' L2 reading proficiency and similarity features between texts?

RQ4-2: Do the serial processes of analogical transfer differ according to EFL learners' L2 reading proficiency and similarity features between texts?

5.2.1 Method

5.2.1.1 Participants

A total of 46 Japanese university EFL students participated in this study. They were students from a national university with diverse majors: international studies, science and engineering, and education. All participants were divided into two proficiency groups based on their scores on an L2 reading proficiency test. Two students were excluded from the data analysis because their experimental scores were outliers, resulting in 44 students in the analysis (see 5.2.1.4 Scoring and Data Analysis for details).

5.2.1.2 Materials

(1) *L2 Reading Proficiency Test*

The L2 reading proficiency test was the same as Experiment 1 and 2. All of the test items were adopted from the Society for Testing English Proficiency (STEP) pre-Grade 1 (14 items), Grade 2 (17 items), and pre-Grade 2 (3 items) tests (2007). The item numbers were 34, and only the reading section was used.

(2) *Reading materials*

Four sets of paired analogical texts were used. Each set consisted of a target text and four

paired source analogs. The four source analogs were assigned four conditions, namely similarity type (surface vs. structural) x causality (strong vs. weak) based on the components of common matches in each similarity feature (the number of entity matches, FOR matches, and HOR matches; see Appendix 4). In other words, a total of 20 narrative texts were used. Ten of the narrative texts used were modified versions of those found in Gentner et al. (1993), while six were modified from Catrambone (2002). The remaining four were by the author, written to conform to the same standards as the other texts.

Table 5.1

An Example of Similarities in a Surface-Strong Causality Condition

Target	Source analog
Karla, an old hawk, lived at the top of a tall oak tree. One afternoon, she saw a hunter on the ground with a bow and some crude arrows that had no feathers…	Once there was an eagle named Salam who nested on a rocky cliff. One day she saw a sportsman coming with a crossbow and some arrows that had no feathers…

Entity	FOR	HOR
hawk vs. eagle; bow vs. crossbow; feathers vs. tail-feathers...	Karla saw a hunter coming with a bow vs. Salam saw a sportsman coming with a crossbow …	An attack is made but fails; this causes the one being attacked to offer to provide an item to the attacker…

Note. The above represents one of the Text A pairs.

To manipulate the level of similarities in the texts, two raters identified the matches between the target and source analog. Following the methodology used by Catrambone (2002), the entity matches were identified at the individual word level, the FOR matches at the sentence level, and the HOR matches at the text level, as shown in Table 5.1. The two raters scored the texts independently, and the resulting inter-rater agreement rate was 84.82%. All the

disagreements were resolved through discussion. As for the results, surface similarity conditions (*M* = 4.38) shared more entity matches than structural similarity conditions (*M* = 1.25). The number of FOR matches by condition was, in descending order, surface-strong (*M* = 5.75), structural-strong (*M* = 4.75), surface-weak (*M* = 4.5), and structural-weak (*M* = 2.25). The HOR matches were shared only in strong causality conditions. Accordingly, it is reasonable to conclude that the materials used in the experiment accurately reflect the similarity features.

Table 5.2

Number of Words, Sentences, and the Readability in Target Text and Source Analog

	Target text				Source analog			
Text pair	(A)	(B)	(C)	(D)	(A')	(B')	(C')	(D')
Words	79	79	84	95	76.82	73.5	95.26	98.75
Sentences	6	6	5	6	6.51	5.75	5	6
FKG level	5.1	5.5	7.6	8.3	5.83	5.32	8.88	8.79

Note. The FKG Level (Flesch-Kincaid Grade) Level was calculated with Microsoft Word 2007; the source analog values are the averages of the four texts that were paired with the target texts.

Also, some difficult words in the texts which were higher than Level 6 on *JACET 8000* (JACET, 2003) were replaced with plainer words. If there was no appropriate substitute word, an explanatory note was added. The number of words and the levels of readability were controlled in each target and source analog. After modification, all passages were between 79 and 95 words in length, contained five to six sentences, and ranged in readability from Level 5.1 to 8.3 on the Flesch-Kincaid Grade (see Table 5.2). The text conditions and the order of given texts were counterbalanced across the participants.

(3) *Recognition Test Items*

A recognition test was designed to examine the activation of analogical information by

accessing and mapping in the participants' memory. In the methodology used, the target and source analog were presented in series, but the participants were not directly informed of the link between the two stories. In the experiment, the participants read a source analog which included a statement which was not written in the target text. Immediately after reading the target and source analog (in that order) to their respective ends, three types of probe statements were provided, and participants were asked to judge whether each was in the target text or not.

The three types of probe statements were as follows:

(a) Explicit statement: information that was actually in the target text

(b) Inference statement: information that was not in the target text but drew upon the source analog

(c) Control statement that was not in the text and that was not related to the analogy (see Appendix 4).

The most important type is (b) inference statement because the reactions to this statement represent the degree of activation by accessing and mapping between the two stories. For example, the inference statement for Text A, "The hunter promised never to shoot at the hawk again," was not written explicitly in the target text, although the source analog contained the sentence, "He (the sportsman) promised never to attack eagles again." If the source information was activated when they read the target text, "yes" response rates to the inference statement would increase compared with the reaction to the control statements. Also, (a) explicit statements were set for the purpose of examining the sensitivity to the target text information compared with the information that was not present in the text.

5.2.1.3 Procedure

All tasks were administered to participants individually over the course of about 80

minutes. First, the participants took an L2 reading proficiency test, which took 30 minutes. Second, the reading of analogical texts was conducted on computer screens. Before reading the materials, participants read a practice text to make sure they knew how to operate the device. Each participant read the four sets of analogical texts, which were paired with one source analog and one target text. Texts appeared in the screen sentence by sentence using Super Lab Pro 4.0 for Windows and the successive sentences were displayed whenever the participants pressed the "advance" key in the Response Box, RB730. Third, a recognition test was conducted. After reading each source analog and target text pair to the end, the reading stimulus "***" was presented for 500 milliseconds (ms), and then the probe statement was displayed. Participants were required to answer whether the probe sentence was in the text or not as quickly as possible by pushing the "yes" or "no" key. The precise instructions were as follows: "You will read statements after reading two passages in succession, and you will have to answer whether each statement was or was not in the text that you read by pressing either the YES or NO key." All analogical text pairs and probe statements were presented in random order. Afterwards, Yes-No comprehension questions were presented in order to verify that they understood these texts.

Finally, the cued recall task was administered. Participants were given four pieces of paper inscribed with the first story (source analog). Using it as a cue, they were instructed to write down the second story (target text), which they were reminded of in Japanese. There was no time limit set, but almost all participants took 20 minutes to complete this task.

5.2.1.4 Scoring and Data Analysis

First, two participant were excluded from the data since his mean score on the Yes-No comprehension questions was under 80%. The data that exceeded ±2.5 *SD* from the mean were deleted from the results for recognition times for the probe sentence. Thereafter, the remaining

recognition time for probe sentences were divided by the number of syllable sounds for uniformity.

Secondly, as for the "yes" response rates for the recognition test, a 3 (statement: explicit, inference, control) x 2 (similarity types: surface, structural) x 2 (causality: strong, weak) x 2 (proficiency: upper, lower) four-way mixed analysis of variance (ANOVA) was performed. The factor of proficiency was the between-subject factor, and the remaining three factors (statement, similarity types, causality) were within-subject factors.

Thirdly, for the scoring of the recall data, the target texts were divided into idea units (IUs) by two raters in accordance with Ikeno's (1996) criteria. The inter-rater agreement was 96% and all disagreements were resolved through discussion. One point was given when each IU was correctly recalled. The two raters scored 30% of the recall protocols independently. Since the inter-rater reliability was sufficiently high (r = .95), the remaining 70% of the protocols were scored by the author. Production rate data was calculated according to how many IUs out of the total were recalled. This data was analysed by a 2 (similarity types: surface, structural) x 2 (causality: strong, weak) x 2 (proficiency: upper, lower) three-way mixed ANOVA.

Finally, in order to examine what kind of information remained in participants' text representations relative to RQ4-1, all text information had to be divided into more important IUs (e.g., main ideas) and less important IUs (e.g., detailed information). A total of 10 graduate students who majored in English education helped with the task of importance rating. The degree of importance per IU for understanding text content was rated on a five-point scale. The mean of rated points was analyzed by cluster analysis using Ward's method. All IUs were classified as more important IUs [range: 3.44-4.89] or less important IUs [range: 1.67-2.89]. Afterwards, the proportions of more and less important IUs in each reading relative to the

participants' recall data were compared with a four-way ANOVA that added information types (more important vs. less important) as a within-subject factor.

5.2.2 Results

5.2.2.1 L2 Reading Proficiency Test

In order to confirm the reliability of the L2 proficiency test, the Eiken Test, Cronbach's alpha reliability was calculated. The reliability of the test was sufficiently high (Cronback's coefficient alphas was .82).

Table 5.3 shows the descriptive statistics for the L2 reading proficiency test. Based on their scores on L2 reading proficiency test, the 44 participants were divided into two groups by t (42) = 8.71, $p < .001$). As a result, the upper group consisted of 21 participants (M = 24.79, SD = 1.68), whereas the lower group consisted of 23 participants (M = 18.04, SD = 2.07).

Table 5.3
Descriptive Statistics for L2 Reading Proficiency Test

Proficiency group	*n*	*M*	*SD*
Upper	21	24.79	1.68
Lower	23	18.04	2.07
Total	44	21.46	3.05

Note. A full score for the test was 34.

5.2.2.2 Probe Recognition Task (Recognition Time)

To examine the effect of similarity types and levels on recognition time and "yes" response rates, a 3 (probe type: explicit, inference, control) x 2 (similarity types: surface, structural) x 2 (causality: strong, weak) x 2 (proficiency: upper, lower) four-way ANOVA was performed. The results for recognition time showed a significant main effect of probe type and proficiency, F (2, 42) = 10.289, p = .000, η^2 = .205 and F (1, 42) = 6.329, p = .016 η^2 = .137,

respectively. However, any interaction and the main effect of the similarity types and levels were not significant (p >.05). As for probe types, the recognition time (ms) of text items (M = 7170, SD = 407.09) and analogical inference items (M = 7153, SD = 340.71) were longer than that of control items (M = 6018, SD = 314.35) in every condition. This means that the participants were accessing analogical inference information which was non-existent in the text, with the same speed as the information explicitly written in the text. On the other hand, since the control item is not in the text and could not be generated by inference, its existence was immediately dismissed. This is an indication that EFL participants activation of analogical inferences and their processes was not influenced by the similarity types or levels.

5.2.2.3 Probe Recognition Task ("Yes" Response Rates)

Table 5.4 shows the descriptive statistics for the "yes" response rates. As a whole, the "yes" response rate for analogical information was relatively high. To clarify how similarity features and L2 reading proficiency affect the processes of accessing and mapping, the responses to each probe statement were analyzed. A 3 (statements: explicit, inference, control) x 2 (similarity types: surface, structural) x 2 (causality: strong, weak) x 2 (proficiency: upper, lower) four-way mixed ANOVA revealed the significant main effect of probe types, F (2, 84) = 136.61, p = .001, η^2 = .765. Moreover, it was shown that there was a significant three-way interaction among the statements, causality and proficiency, F (1, 42) = 5.27, p = .007 η^2 = .112. (see Table 5.5)

Table 5.4

Descriptive Statistics for the "Yes" Response Rates

Similarity type	Causality	Proficiency	Explicit		Inference		Control	
			M	*SD*	*M*	*SD*	*M*	*SD*
Surface	Strong	Upper	.90	.30	.57	.50	.00	.00
		Lower	.74	.49	.52	.51	.09	.28
	Weak	Upper	.86	.36	.48	.50	.10	.31
		Lower	.76	.21	.38	.51	.00	.00
Structural	Strong	Upper	.95	.21	.43	.43	.00	.00
		Lower	.78	.42	.48	.58	.09	.28
	Weak	Upper	.81	.30	.24	.49	.00	.39
		Lower	.66	.38	.35	.48	.17	.29

Note. The numbers in each proficiency group: upper group (n = 21); lower group (n = 23).

Table 5.5

Results of Four-Way ANOVA for the "Yes" Response Rates

	SS	*df*	*MS*	*F*	*p*	η_p^2
Within-Subjects Effects						
Similarity Types	.290	1	.290	2.180	.147	.049
Causality	.121	1	.121	1.211	.277	.028
Probe Types	56.009	2	28.004	136.616	.000	.765
Similarity Types×Causality	.121	1	.121	1.046	.312	.024
Similarity Types×Probe Types	.593	2	.296	2.437	.094	.055
Causality×Probe Types	.025	2	.013	.139	.870	.003
Similarity Types×Causality× Probe Types	.177	2	.089	.596	.554	.014
Between-Subjects Effects						
Proficiency	.053	1	.053	.243	.625	.006
Similarity Types×Proficiency	.139	1	.139	1.042	.313	.024
Causality×Proficiency	.000	1	.000	.003	.960	.000
Probe Types×Proficiency	.372	2	.186	.908	.407	.021
Similarity Types×Causality× Proficiency	.000	1	.000	.002	.963	.000
Causality×Probe Types × Proficiency	.965	2	.482	5.272	.007	.112
Similarity Types×Causality× Probe Types ×Proficiency	.299	2	.149	1.003	.371	.023

Note. η2 = effect size; 0.010 ≤ η2 (small) < 0.059, 0.059 ≤ η2 (medium) < 0.138, 0.138 ≤ η2 (large; Cohen, 1988, pp. 284-287); Effect size of 0.009 and below is regarded as having no effect.

The “yes” response rates of each statement by proficiency and the degree of causality are presented in Figure 5.1. Multiple comparisons with the Bonferroni post-hoc test showed that, in strong causality conditions, the upper group answered “yes” to the explicit statements ($p < .001$) and inference statements ($p = .005$) more frequently than to the control statements. Also, the explicit statements yielded “yes” responses more frequently than the inference statements ($p = .003$). In weak causality conditions, the same results were obtained: the answer rates to the explicit statements ($p < .001$) and inference statements ($p = .023$) by the upper group were higher than those to the control statement. There was also a significant difference between the explicit and inference statements ($p = .003$). These results indicate that proficient readers were significantly more likely to respond that the explicit and inference statements had been stated in the target texts when the relevant information could be accessed and mapped from the source analog regardless of causality between the two texts. The sensitivity of access to the information explicitly written in the target text was especially high.

Much like the upper group, the lower group, in strong causality conditions, answered “yes” to the explicit statements ($p < .001$) and the inference statements ($p = .008$) more frequently than to the control statements. However, there was no significant difference between the explicit and inference statements. Furthermore, in weak causality conditions, the “yes” response rates to the explicit statements were higher than those to the inference ($p = .002$) and control statements ($p < .001$). One particularly noteworthy point is that there was no significant difference between the inference and control statements ($p = .562$). These results indicate that the sensitivity of access to the explicit text information had decreased. Also, the accessing and mapping process did not occur in lower group reading when causality was weak, because the analogical information was not activated in their minds.

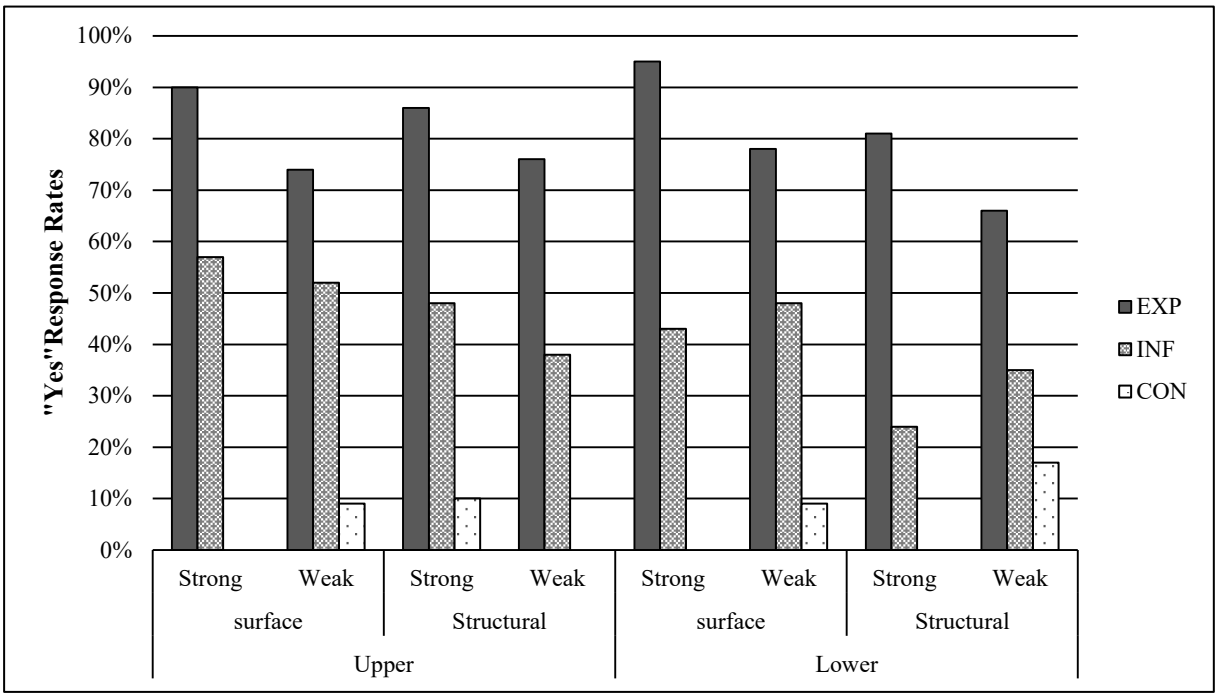


Figure 5.1. "Yes" response rates of each statement by proficiency and the degree of causality.

Table 5.6 *Sensitivity and Activation Value Scores for Different Conditions*

Proficiency	*n*	Sensitivity				Activation Value			
		SSH	SSL	RSH	RSL	SSH	SSL	RSH	RSL
Upper	21	.91	.76	.95	.81	.57	.33	.38	.24
Lower	23	.65	.96	.71	.65	.43	.48	.39	.18

Note. Calculation of response sensitivity: hits ("yes" answers to text items) minus false alarm ("yes" answer to control item"); activation value: false alarm ("yes" answers to analogical inference) minus false alarm ("yes answers to control item").

As additional analysis, response sensitivity and activation value were calculated. Means for the different conditions are reported in Table 5.6. Sensitivity indicates how accurately explicit information in the text has been responded to, while the activation value shows the rate of the activation of analogical inference information. The results show that when analogical inference information is high, response sensitivity to explicit information is promoted. Also, it

indicates that although the activation value does not show much difference among conditions, analogical level has relative influence.

In sum, the difference of similarity types did not influence the accessing and mapping process at all. More proficient readers were able to access analogical information and map the information from the source analog to the target regardless of the degree of causality. However, it was difficult for less proficient readers to perform accessing and mapping processes when the cue was not sufficient because the causality was weak.

5.2.2.4 Cued Recall Task

Table 5.7 shows the overall results of the recall rates produced in each condition. First, it was confirmed that there were no significant differences among the four texts A-D as the prerequisite for the following results did not lie in the differences in text difficulty. A one-way ANOVA with the text difference as a within-subject factor was conducted. It was confirmed that the four texts were equal in ease of reading comprehension, F (3, 129) = 1.15, p = .333, η^2 = .026.

Afterwards, to examine the quantity of remaining information in learners' representations, learners' recall performances were analyzed. A 2 (similarity types: surface, structural) x 2 (causality: strong, weak) x 2 (proficiency: upper, lower) three-way mixed ANOVA was performed (see Table 5.8). It revealed a significant main effect of similarity type, F(1, 42) = 12.6, p = .001, η^2 = .231, and proficiency, F (1, 42) = 4.58, p = .038, η^2 =.098. Also, there was a significant three-way interaction among similarity type, causality and proficiency, F (1, 42) = 6.47, p = .015, η^2 = .133.

Table 5.7 *Descriptive Statistics for the Recall Rates*

Similarity type	Causality	Upper (n = 21)		Lower (n = 23)	
		M	*SD*	*M*	*SD*
Surface	Strong	.55	.15	.52	.14
	Weak	.56	.16	.49	.15
Structural	Strong	.58	.19	.36	.24
	Weak	.41	.23	.39	.23

Table 5.8

Results of Four-Way ANOVA for the Recall Rates

	SS	*df*	*MS*	*F*	*p*	η_p^2
Within-Subjects Effects						
Similarity Types	.38	1	.38	12.61	.001	.231
Causality	.08	1	.08	3.46	.070	.076
Similarity Types×Causality	.04	1	.04	1.54	.222	.035
Similarity Types×Proficiency	.06	1	.06	1.82	.184	.042
Causality×Proficiency	.06	1	.06	2.90	.096	.065
Similarity Types×Causality×Proficiency	.15	1	.15	6.47	.015	.133
Between-Subjects Effects						
Proficiency	.31	1	.31	4.58	.038	.098

Note. η2 = effect size; 0.010 ≤ η2 (small) < 0.059, 0.059 ≤ η2 (medium) < 0.138, 0.138 ≤ η2 (large; Cohen, 1988, pp. 284-287); Effect size of 0.009 and below is regarded as having no effect.

The recall rates for each condition are shown in Figure 5.2. Multiple comparisons of interactions with the Bonferroni test revealed that both the upper (p = .012) and the lower (p = .048) groups produced more target information based on surface similarity than structural similarity when causality was weak. Also, there was neither a significant difference between the upper and lower group nor strong causality and weak causality in surface similarity conditions. This indicates that the upper and lower groups produced the same amount of information regardless of the causality level in surface similarity conditions.

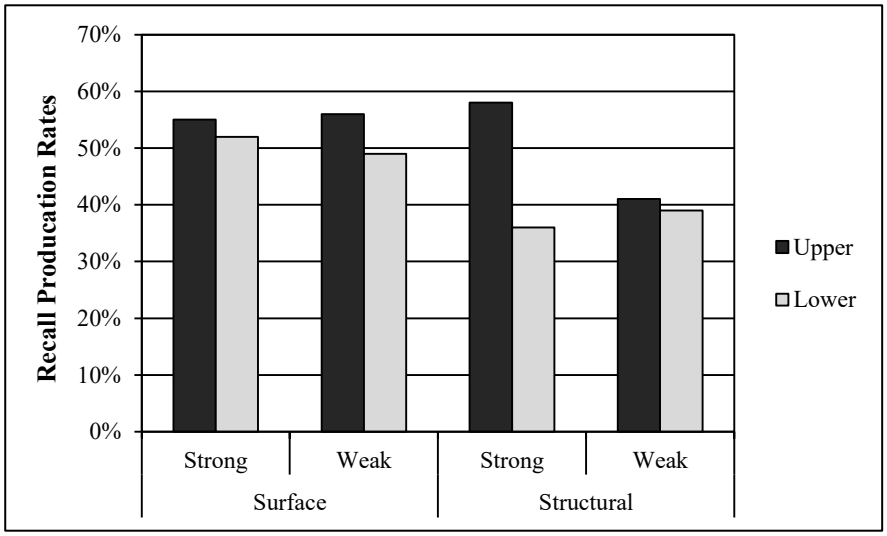


Figure 5.2. Cued recall rates for each condition.

Meanwhile, in structural similarity conditions, the uppers group's recall performance based on strong causality was significantly higher than that of weak causality (p = .006). Furthermore, there was a significant difference between the upper and lower groups (p = .002) in structural similarity conditions and strong causality conditions. However, there was no significant difference between the two groups in weak causality conditions. As a result, in surface similarity conditions, both groups were able to recall nearly half of the target text contents (49-56%) regardless of the degree of causality. In contrast, in structural similarity conditions, only the upper group was able to recall a lot of information (58%), compared with the lowers group's performance (36%), relative to strong causality. These results are quite suggestive because structural similarity refers to the resemblance in the underlying system of relations without semantic resemblance (i.e., shared the FOR and HOR matches without entity matches). Therefore, the results show that although proficient readers were able to encode much of the target information to their text representations with structural similarity as a basis for causal structure within two texts, it was difficult for less proficient readers. This suggests that information encoding based on structural similarity requires a certain level of L2 reading

proficiency.

Next, to examine the quality of information which remained in the learners' representations, their recalled contents were analyzed on the basis of importance-rating classifications (see 5.2.4 Scoring and Data Analysis for details). In regard to recall rates in each reading condition, a 2 (information types: more important, less important) x 2 (similarity types: surface, structural) x 2 (causality: strong, weak) x 2 (proficiency: upper, lower) three-way mixed ANOVA was performed. The results showed a significant main effect of similarity types, $F(1, 42) = 13.13$, $p = .001$ $\eta^2 = .238$, and information types, $F(1, 42) = 116.78$, $p = .001$ $\eta^2 = .735$. Moreover, there was a significant four-way interaction among information types, similarity types, causality and proficiency, $F(1, 42) = 6.27$, $p = .016$ $\eta^2 = .131$. Multiple comparisons of interactions with the Bonferroni test revealed that the upper group recalled both the more important ($p = .021$) and less important information ($p = .022$) in surface similarity conditions than in structural similarity conditions when causality was weak. However, when causality was strong, the effect was reversed for the upper group, yielding the production of more important information in structural similarity conditions ($p = .028$). That is to say, the upper group recalled more important information based on causal relations (as opposed to less important information) in structural similarity conditions. The task performance exhibited a significant difference between the upper and lower groups ($p = .011$). In contrast, the lower group produced both types of information in surface similarity conditions than they did in structural similarity conditions. Moreover, causality did not influence any type of information.

To summarize the results of the cued recall task, when surface similarities with semantic connections (i.e., many entity matches) were common to the target and source analog, the same amount of information was produced in both the upper and the lower groups, regardless of the strength of causality. Meanwhile, in structural similarity conditions, in which there were no

entity matches, only the upper group produced the more important information, which was linked to strong causality. According to Langston and Trabasso (1999), it has been shown that it is more important information that has many causal connections with other propositions. In sum, it can be assumed that the understanding of the upper group improved as a consequence of their ability to construct causal relations between propositions and thus transfer them analogically to the target.

In addition, based on the Macro-structure model (Kintsch, 1998), the comparison between conditions was conducted so as to judge the propositional level of recalled information in text structure. First, an important rating task was conducted with ten graduates who majored in English Education. The purpose was to have them rate the importance of each IU information in the target text. Against the obtained score, cluster analysis using Ward's method was carried out. The clusters were categorized as follows: high importance as Macro IUs, low importance as Micro IUs: Text A [Macro IUs = 9 (Importance rating task score range: 3.78-5.00), Micro IUs = 9 (1.67-3.56)], Text B [Macro IUs= 9 (1.78-3.44), Micro IUs = 10 (4.22-4.67)], Text C [Macro IUs = 11 (3.78-4.89), Micro IUs = 10 (1.78-3.44)] Text D [Macro IUs= 12 (3.78-4.89), Micro IUs = 9 (1.44-2.89)].

The cued recall production rates were calculated in each Macro IUs and Micro IUs. A 3 (propositional structure: Macro, Micro) x 2 (similarity types: surface, structural) x 2 (similarity levels: high, low) x 2 (proficiency: upper, lower) mixed-ANOVA was conducted. The results showed a significant main effect of propositional structure, F (1, 42) = 116.78, p = .000, η^2 = .735. However, the interaction with any other factors was not significant. This indicates that although participants retain more important Macro propositions in their memory, similarity features do not influence the result.

5.2.2.5 Questionnaire

The mean of responses to questionnaire items [Q1: How much was the story or situation similar in two passages? Q2: How easy was it to imagine the second passage (target) by using the first passage (source) as cue?] were described in Table 5.9. To examine participants' recognition of similarities, a 2 (similarity types: surface, structural) x 2 (similarity levels: high, low) x 2 (proficiency: upper, lower) mixed-ANOVA was carried out. The results indicated a significant main effect of similarity types and similarity levels in Q1 item, F (1, 42) = 16.88, p = .000, η^2 = .287, F (1, 42) = 116.78, p = .000, η^2 = .735, respectably. As well, same tendency was found in Q2 items, F (1, 42) = 81.15, p = .000, η^2 = .659, F (1, 42) = 16.63, p = .000, η^2 = .284, respectably.

Table 5.9

Mean Responses to Questionnaire Items (N =44)

Proficiency	Similarity Type	Similarity Level	Q1		Q2	
			M	*SD*	*M*	*SD*
Upper	Surface	High	4.62	.17	4.38	.16
		Low	3.67	.21	3.86	.19
	Structural	High	3.24	.22	3.95	.25
		Low	2.57	.24	3.24	.23
Lower	Surface	High	4.17	.16	4.13	.15
		Low	3.78	.20	3.87	.18
	Structural	High	3.26	.21	3.48	.24
		Low	2.48	.23	3.39	.22

Note. Both items were evaluated on five-point scale.

The above results indicate that with regards to surface similarity, it was easier for the participants to form an image when analogies were similar. Also, similarity levels had high scores in high conditions regardless of proficiency, thus showing the fact that participants

correctly recognized the similarities between analogies.

5.2.3 Discussion for Experiment 4

The goal of this research was to investigate whether analogical transfer occurs in the reading processes of EFL learners, and if so, whether it contributes to improvements in their text comprehension. The investigation was carried out with particular attention to the following two facets: the quantity and quality of information facilitated by analogical transfer (RQ4-1), and the serial process of analogical transfer (RQ4-2). As factors which influence these facets, the L2 reading proficiency of EFL learners and similarity features (similarity types and causality) between the target and source analog were examined.

First, with regard to RQ4-1, the results from the cued recall task showed the following findings: in surface similarity conditions, both upper and lower proficiency groups were able to recall about half of the text information (49-56%) regardless of causality level. Also, more information was yielded than in structural similarity conditions (39%, 41%) when causality was weak. However, in structural similarity conditions, the performance of the upper group (58%) was higher than that of the lower group (36%) when causality was strong. Meanwhile, the performance of the lower group in structural similarity conditions was consistently low. These results indicate that causal information and a reader's L2 reading proficiency level are strongly related in the use of structural similarity. Also, the quality of recalled information represented in the upper group tended to include more important information when the causality was stronger under structural similarity conditions. This effect, which did not influence the lower groups' production, occurred because proficient readers were more aware of the causal relations underlying the story they were reading. This promoted the analogical transfer of causally relevant information to similar text structures. To summarize these points, analogical transfer

on the basis of surface similarity occurred easily, irrespective of L2 reading proficiency. On the other hand, analogical transfer based on structural similarity occurred with a certain level of L2 reading proficiency and causality between the two texts as the key factors. This supports Catrambone's (2002) suggestion that the FOR matches are essential to the use of structural similarity.

Second, with regard to RQ4-2, the results from the recognition test showed that the upper group was able to activate analogical information regardless of causality level during reading. However, the lower group needed causal relations by way of shared the FOR matches for activation. Thus, the accessing and mapping processes in analogical transfer were affected according to only causality and L2 reading proficiency level. The presence of semantic features via the sharing of many entity matches between texts did not influence the processes. Furthermore, as the recall performance indicated, encoding in analogical transfer was a complex process that involved similarity types, causality, and the L2 reading proficiency of learners. The surface similarity of texts promoted the encoding of a lot of information to their text representations regardless of causality or L2 reading proficiency level, while the structural similarity of texts became available as a cue only if causal relations were recognized effectively. Taken together, we can conclude that analogical transfer did occur in EFL reading comprehension. However, each process in the sequence was influenced by different factors. During reading, it was possible for accessing and mapping to be driven by causal relations between information even in the absence of common surface features. And, after reading, the process of encoding the analogical information to the text representations of EFL learners was mainly driven by the surface features of texts. Moreover, when surface similarity was lost, structural similarity enhanced the understanding of more important information, but its effect was restricted to the proficient learners.

5.3 Experiment 5: Effects of Structural Similarity Between Analogical Texts on Thematic Inference Generation

Experiment 5 examined thematic inference generation in between-texts-level reading. The sentence-level examination in Experiment 1 showed that both proficient and non-proficient readers activate thematic inferences during reading. It also showed that contextual convergence of the text has a significant influence on searches that utilize activated appropriate inferred information as a cue. However, thematic inference generation at between-texts-level requires readers to integrate information from different parts of the text and also to combine it with their own knowledge. Especially in generating thematic inference while reading a story, it is important to understand the level of thematic situational similarity, which indicates the extent of the match between the abstract knowledge stored in the reader's knowledge structure about the character's goal, planning, planning failure, etc., and the information present in the text (e.g., the theme of the famous fable, 'sour grapes' is similar to the theme of the anecdote about the disgruntled job seeker) (Wharton, Holyoak, & Lange, 1996). Experiment 5 follows the method used by the previous studies (e.g., Johnson & Seifert, 1992; Seifert et al., 1986) and uses two texts that share the same theme to examine whether readers construct the theme of the text during reading.

Seifert et al. (1986) used TAU based texts (thematic abstraction unit; see 2.1.2 for details) to examine thematic inference generation. TAU is an abstract knowledge structure. It captures the relationship that is shared between concepts such as the interaction between the goal and planning expressed in the text. A text contains a problem which may occur, as well as planning information to avoid or solve the problem. This information is useful in organizing the memory to store individual episodes (Hano & Chihara, 1996). For example, the saying "Closing the barn door after the horse has gone" is characterized as insufficient planning. The character knows

the plan that is needed to achieve the goal, but does not implement the plan. Only when the achievement of the goal is hampered, the character attempts to implement the plan, but it is already too late and fails to achieve the goal.

The experimental study by Seifert et al. (1986) used the priming paradigm, where readers were asked to read a TAU-based story pair and then to determine if the target sentence was true or false. As a result, the study demonstrated that when a strategic instruction is given to pay attention to the thematic similarities, the time to respond to the target text becomes shorter in the story pair that shares the same TAU, compared to the story pair that has different TAU. Long et al. (1997) also used the same method and showed that proficient readers' response time was shorter in the same TAU, while non-proficient readers' response time did not change. It means proficient readers understood the common theme at the basis of the stories. The target text used in Seifert et al. (1986) and Long et al. (1997) was the conclusion sentence of the story. This is because the conclusion sentence of a story shows the character's goal, planning and results of actions. The shorter time to process the conclusion sentence in the story where the theme is constructed based on these elements was regarded as the evidence of inference generation in these studies. On the other hand, Catrambone (2002) focused attention on measuring the reading time as a method to measure the reading process more delicately than the priming paradigm. In general, when the reader generates inferences to maintain the consistency of the text, the reading time becomes longer (McNamara & Kintsch, 1996). However, many studies concerning predictive inferences have shown that when the reader predicts the text content, the time it takes to read the conclusion sentence that matches the prediction becomes shorter (e.g., Linderholm, 2002; Magliano et al., 1999). That is because, story representation match the readers' naive theories about a possible world, and two types of reader's knowledge structure: linguistic knowledge (van Dijk, 1972) and domain knowledge represented in the text. Especially, it has

been widely accepted that readers use their domain knowledge about causality to understand a text, that is, construct a coherent text representation. In story comprehension, domain knowledge can be described in terms of (a) *causal schemas,* which include the concept of causality and the naive theories about the physical and social worlds possessed by nonexpert readers; (b) *scripts*, which are stereotyped representations of actions or familiar event sequences; and (c) *plans* that use intentional relations to organize sequences of actions as a hierarchical structure of goals. In L1 research, the predictable parts of a text are shown to be read faster than unpredictable parts, due to the reader's use of causal knowledge including scripts and plans. Similarly in Experiment 5, the time to read the conclusion sentence of the story is expected to become shorter if EFL readers uses the analogical information based on structural similarity and generates thematic inferences.

In addition, Johnson and Seifert (1992) indicated that the cue to search the knowledge structure in thematic inference generation is the goal information in the text. They used cued recall tasks to examine TAU-based story information retrieval. Three types of cues were used: cues that contain all the information on goal, planning and conclusion (complete cue), cues that contain information on goal and planning (partial cue), and cues that contain information on the conclusion only (conclusion cue). As a result, it was discovered that a story with a common theme is retrieved with any type of cue, and that a partial cue leads to more retrieval than a conclusion cue. In place of a conclusion cue, Hano and Chihara (1996) used a predictive cue (goal, planning, plus one sentence that suggests the conclusion) and conducted cued recall tasks with Japanese participants. The results demonstrated that complete cues and predictive cues led to the retrieval of more stories with a thematic match. This indicates that readers predict conclusion information based on the goal and planning information as the cue to construct the overall theme.

Previous studies examined individual cases of thematic generation in between-texts-level comprehension and the type of information that provides cues to access knowledge, however, the examinations were not comprehensive. Therefore, Experiment 5 used both on-line and off-line measurement methods to comprehensively examine EFL learners' thematic comprehension process and knowledge information activation. As the on-line measurement, the time to read the conclusion sentence in the target was measured. As the off-line measurement, a cued recall task was conducted as in Experiment 3. This measured the amount of target text information retrieved based on the cue text. In addition, theme identification tasks were also conducted to examine the quality of the theme that EFL learners generate from the text. The research questions were as follows:

RQ5-1: Do EFL learners activate thematic inferences in between-texts-level comprehension?

RQ5-2: Is thematic inference activation influenced by structural similarity, type of cue information and L2 reading proficiency?

RQ5-3: Is the level of understanding of the text influenced by structural similarity, type of cue information and L2 reading proficiency?

RQ5-4: Is the understanding of the theme influenced by structural similarity, type of cue information and L2 reading proficiency?

5.3.1 Method

5.3.1.1 Participants

Participants in Experiment 5 were 51 Japanese students, studying at national universities in the first to the fourth year. Their specialties were varied, such as international relations, humanities, science and engineering and information technology. Participants were recruited in the general English class, and they were paid a small amount of remuneration. Participants were

divided into two groups according to the LD2 comprehension test score. Two participants scored less than 80% in the questions which assess the understanding of the experiment. These two participants were regarded as lacking sufficient understanding of the comprehension passage, and therefore excluded from the analysis (see 4.3.1.4 Scoring and Data Analysis for details). As a result, 49 participants were analyzed.

5.3.1.2 Materials

(1) *L2 Reading Proficiency Test*

The L2 reading proficiency test was the same as Experiment 1, 2 and 4. All of the test items were adopted from the Society for Testing English Proficiency (STEP) pre-Grade 1 (14 items), Grade 2 (17 items), and pre-Grade 2 (3 items) tests (2007). The item numbers were 34, and only the reading section was used.

(2) *Reading materials*

Three sets of texts that were used in Wharton et al. (1996) and Hano and Chihara (1996) were modified and used as the reading materials. One set comprised five passages in total as shown in Table 5.10 (the rest of materials are listed in Appendix 5). Each set contained two targets: (a) same-theme target and (b) different-theme target. As for the content of the target, the same-theme target contained the same theme as the cue text. The different-theme target contained similar goal planning, but the ending is different, and therefore the theme is different.

Table 5.10

An Example of Target Passages and Cues Used in Experiment 5 [Set A]

Same-Theme Target	*Different-Theme Target*
[Betrayal of Deal] Once there was an eagle named Karla. She saw a hunter armed with bows and arrows coming after her. Karla noticed that some of the arrows had no feathers. She thought that she might be able to make a deal. Karla flew down and donated a few of her tail feathers to the hunter. The hunter promised never to attack eagles again. <u>However, when she was nesting on a cliff, the arrows shot her with her tail feathers on them.</u>	[Making a deal to avoid a bad situation] One day, lying on a big rock, was a snake called Elrod. He saw a farmer coming with a blow gun tracking him. The farmer shot at him. Elrod realized that the darts had no poison so he decided that he might be able to make a deal. He crawled over and gave some poison to the farmer so that the darts now had his poison on them. <u>The farmer agreed never to kill snakes again, so Elrod spent the rest of his life safely.</u>
Complete Cue Salam was a small country. One day, Salam realized that its warlike neighbor, Bolon was arming against them. The Salamians realized that all of the Bolonian missiles were badly made and decided that an alliance could be formed. The Salam ambassador gave useful missile components to the Bolonian army. Bolonians promised that they would never fight against Salam from now on. <u>However, Bolon declared war against Salam, and attacked them with the new rockets made of Salamian missile components.</u>	
Partial Cue Salam was a small country. One day, Salam realized that its warlike neighbor, Bolon was arming against them. The Salamians realized that all of the Bolonian missiles were badly made and decided that an alliance could be formed. The Salam ambassador gave useful missile components to the Bolonian army. Bolonians promised that they would never fight against Salam from now on.	
Predictive Cue Salam was a small country. One day, Salam realized that its warlike neighbor, Bolon was arming against them. The Salamians realized that all of the Bolonian missiles were badly made and decided that an alliance could be formed. The Salam ambassador gave useful missile components to the Bolonian army. Bolonians promised that they would never fight against Salam. <u>At that time, Salamians were not aware of the danger of forming an alliance with Bolonians.</u>	

Note. The underlined sentence in each Target and in Complete Cue is the conclusion sentence; the sentence with the dotted underline in Predictive Cue is the sentence that predicts the conclusion.

For example, the story "An eagle makes a deal with the hunter where the eagle gives the hunter a feather in exchange for the hunter not shooting the eagle with an arrow, but in the end, the promise is betrayed," is a same-theme target. On the other hand, the story "A snake gives

the farmer its poison in return for the promise that the farmer will not shoot the snake, and the promise is kept," is a different-theme target. In short, the goal planning is common between the two targets, and the superficial structures of the stories are extremely similar. However, because the ending is different, the theme of the former is "betrayal of deal" and the theme of the latter is "making a deal to avoid a bad situation."

The cues had three conditions: (a) complete cues that contain information on all of the goal, planning and conclusion, (b) partial cues that contain information on goal and planning only, and (c) predictive cues that contain information on goal, planning, plus one sentence that indicates the conclusion. The cue takes the form of analogy of the target, therefore the story has low superficial similarity but high situational similarity. For example, a story with (a) Complete Cue is: "A country called Salam forms an alliance with its warlike neighbor, Bolon, where Salam provides Bolon with missile components and Bolon will not fight against Salam, but the promise is eventually broken."

Here, the characters' attributes and superficial characteristics are different from the target, but the theme matches with the same-theme target in terms of the causal relationship of the events. In Partial Cue (b), the conclusion that "the neighboring country Bolon breaks its promise" is omitted. On the other hand, (c) includes a predictive sentence that "Salamians were not aware of the danger of forming an alliance with Bolonians," instead of the conclusion sentence. Figure 5.3 illustrates the components of analogical texts used in Experiment 5.

The target theme conditions (same-theme target, different-theme target) are the factors between test subjects, and the cue conditions (complete cue, partial cue and predictive cue) are the factors within each test subject. Therefore, it was arranged in the experiment so that a participant reads three same-theme targets, or three different-theme targets consistently, paired with three types of cues.

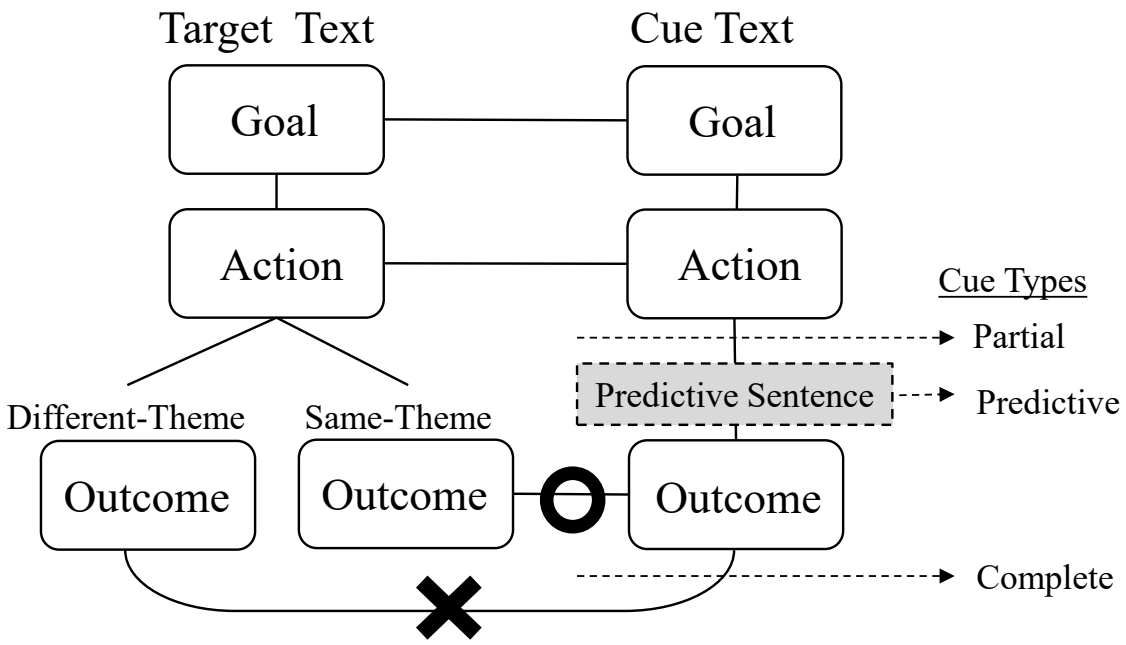


Figure 5.3. Components of analogical texts used in Experiment 5.

The time to read the conclusion sentence in the target was measured in the experiment. If the reader generates common thematic inference between the cue and the target, the reading time for same-theme target is expected to be shorter than for the different-theme target. On the other hand, if the reader does not generate thematic inference during reading, the reading time for same-theme target and for different-theme target is expected to be similar.

In terms of the difficulty level of the text, words that are regarded Level 5 or above in JACET 8000 (JACET, 2003) were replaced with simpler words. In order to measure the understanding of the content, two comprehension questions were prepared for each target and cue. Table 5.11 shows the number of words, sentences and readability. Please refer to Appendix 5 for the complete materials.

Table 5.11

Number of Words, Sentences, and the Readability in Target Texts

	Same-Theme Target			Different-Theme Target		
Set	(A)	(B)	(C)	(A)	(B)	(C)
Words	82	91	83	85	91	94.
Sentences	7	7	5	6	7	5
FKG level	5.2	5.7	6.9	5.6	6.3	6.4

Note. The FKG (Flesch-Kincaid Grade) Level was calculated with Microsoft Word 2010.

5.3.1.3 Procedure

The experiments were carried out individually. Each session was comprised of three stages. Firstly, the L2 reading proficiency test was carried out within the time limit of 30 minutes. Each participant was given a booklet, and wrote the answers on the answer sheet.

Then the on-line comprehension task was carried out. Each participant read a practice text on the PC and practiced how to use SuperLab 4.0, which is the device used to measure the experiment data. Thereafter, each participant read the main text at their own pace. Each participant read six comprehension texts, comprised of three target texts and three cue texts. A cue text and the target text were paired into three pairs. The three pairs were randomly presented for counterbalance. Text was displayed on the computer screen one by one, and the reading time was measured. This is because the reading time for the target conclusion sentence was also analyzed.

Participants were notified in advance that they will be given two Yes/No comprehension questions after reading each text, and they were instructed to understand the content while reading. Yes/No comprehension questions were displayed at the end of reading each text after "???" was displayed for 500ms. Participants were asked to press the *Yes* key if they thought the question was correct, and *No* when incorrect. “Karla noticed that the hunter had some arrows with no feathers.” Answer: *Yes*)

Finally, after completing the reading task on-line, cued recall tasks and theme identification tasks were carried out. For the cued recall task, participants were given only the cue on a sheet of paper, and asked to write the paired text that can be recalled from the cue as much as possible in Japanese. For the theme identifying task, participants were given only the target text on a sheet of paper. Participants were then asked to write the theme that was derived from the text in one sentence or so in Japanese.

5.3.1.4 Scoring and Data Analysis

First of all, the two participants who scored less than 80% in the 12 Yes/No questions were excluded for not understanding the comprehension text fully. The L2 reading proficiency test scores of the remaining 49 participants were analyzed, using one-way factorial analysis of variance: 2 (L2 reading proficiency: upper, lower) x 2 (theme: same, different).

Secondly, the reading time for the target text conclusion sentence was divided by the number of syllables to calculate the reading time per syllable. This is because the target conclusion sentence was different for the same-theme condition and for the different-theme condition, so it was necessary to control the difference in the number of syllables before comparing the reading time. In addition, data over ±2.5*SD* were excluded as outliers. Firstly, two-way factorial analysis of variance was used to analyze 3 targets x 2 (theme: same, different), and this confirmed that the results were not due to an effect specific to the text. F (2, 94) = .035, p = .966, η_p^2 = .001. Then the reading time data of the conclusion sentence was analyzed using three-way ANOVA: 2 (theme: same, different) x 2 (L2 proficiency: upper, lower) x 3 (cue: complete, partial, predictive).

Thirdly, the six target texts were categorized by IUs in marking the cued recall task data. This categorization was carried out by two raters according to Ikeno's (1996) criteria. Inter-

rater agreement was 93.12%. Disagreement was resolved by consultation. In the analysis, the ratio of the exact match of the target recalled production protocol to the paired cue was calculated. The aim was to examine the extent of identifying common theme between target and cue in the strategic process of recalling target from cues. Based on this, the recalled production rate for the matched recall protocol was calculated. Then the recalled production rate was analyzed using three-way ANOVA: 2 (theme: same, different) x 2 (L2 proficiency: upper, lower) x 3 (cue: complete, partial, predictive).

Finally, the theme identification task data was analyzed. This task measured whether the participants were able to specify the theme that encompasses goal, planning and conclusion from the target text. The response to the open-ended question was given 2 points if the response provided the themes below, 1 point if the response provided themes that partially reflected the content, and zero points in the case of the wrong theme or no response. Table 5.12 below shows each theme for same-theme condition and different-theme condition.

Two raters marked 30% of the data (inter-rater agreement reliability r = 79.16), followed by the remaining 70% (inter-rater reliability r = 86.82). The marked data was analyzed using the chi-square test.

Table 5.12
Themes of Same-Theme/Different-Theme Target Texts

	Same-Theme Target	Different-Theme Target
Set A	Betrayal of a deal	Making a deal to avoid a bad situation
Set B	Sour grapes	Self-blame
Set C	Taking appropriate action too late	Reconciliation and learning from experience

5.3.2 Results

5.3.2.1 L2 Reading Proficiency Test

L2 reading proficiency data was analyzed using one-way analysis of variance, x 2 (L2 proficiency: upper, lower) x 2 (theme: same, different). Table 5.13 shows the L2 reading proficiency scores by group. The result showed a significant difference between the upper and lower groups, F (1, 45) = 956.18, p = .000 η_p^2 = .709. This result confirmed that the upper and lower grouping was appropriate according to the level of comprehension proficiency.

Table 5.13
Descriptive Statistics for L2 Reading Proficiency Test

Group	Target	*n*	*M*	*SD*
Upper	Same-Theme	13	27.92	1.70
	Different-Theme	10	27.60	1.71
Lower	Same-Theme	12	19.50	3.89
	Different-Theme	14	18.21	3.55
Total	Same-Theme	25	23.88	5.18
	Different-Theme	24	22.12	5.53

Note. Full mark for the test was 34.

On the other hand, no significant difference was observed between thematic conditions, F (1, 45) = 7.81, p = .894 η_p^2 = .019, indicating a similar level of proficiency between thematic conditions, which are between-subjects effects. Cronbach's alpha for the L2 reading proficiency test was .82.

5.3.2.2 Reading Time of Conclusion Sentence

The reading time of conclusion sentence was analyzed to examine thematic inference activation by the readers while reading the text. Analysis was made based on the following logic: If the EFL learner activates theme while reading based on the information obtained from

the cue, the reader can predict the conclusion of the target of the same theme. Therefore the reading time for conclusion becomes shorter. Table 5.14 shows descriptive statistics for reading time of conclusion sentences. Figure 5.4 shows the mean reading time of conclusion sentences.

Table 5.14
Descriptive Statistics for Reading Time of Conclusion Sentences

	Same-Theme Target			Different-Theme Target		
	COM	PAR	PRE	COM	PAR	PRE
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Upper	337.05	377.87	340.48	482.20	419.17	512.54
(n = 25)	(105.52)	(175.86)	(201.73)	(181.50)	(194.33)	(162.95)
Lower	435.21	410.26	333.22	439.97	485.46	519.99
(n = 24)	(205.20)	(173.76)	(177.69)	(155.74)	(150.84)	(180.88)
Total	384.17	393.42	337.00	457.56	452.31	516.88
(N = 49)	(165.44)	(17197)	(186.63)	(164.47)	(179.86)	(169.99)

Note. COM = complete cue; PAR = partial cue; PRE = predictive cue.

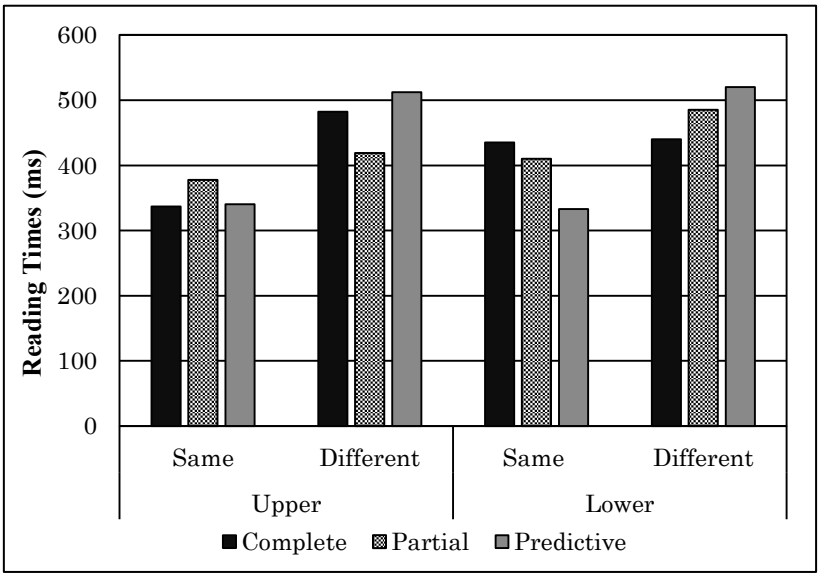


Figure 5.4. Reading times of conclusion sentences.

Reading time data was analyzed using three-way ANOVA: 2 (theme: same, different) x 2 (L2 proficiency: upper, lower) x 3 (cue: complete, partial, predictive) (See ANOVA Table

5.15). As a result, the main effect of theme F (1, 45) = 14.36, $p < .001$, $\eta_p^2 = .242$, and cue x theme x proficiency two-way interaction F (2, 90) = 4.39, $p = .042$ $\eta_p^2 = .089$ was observed. On the other hand, the main effect of cue, F (1, 45) = .14, $p = .867$, $\eta_p^2 = .001$, and the main effect of proficiency F (1, 45) = 1.51, $p = .227$ $\eta_p^2 = .032$, and the first-order interaction were not significant.

Table 5.15

Results of Three-Way ANOVA for Reading Time of Conclusion Sentences

	SS	*df*	*MS*	*F*	*p*	η_p^2
Within-Subjects Effects						
Cue	8046.31	2	4023.15	.14	.867	.003
Cue×Theme	75052.11	2	37526.05	1.33	.268	.029
Cue×Proficiency	44431.19	2	22215.59	.79	.456	.017
Cue×Theme×Proficiency	92152.71	2	46076.35	4.38	.042	.089
Between-Subjects Effects						
Theme	485756.55	1	485756.54	14.36	.000	.242
Proficiency	50776.39	1	50776.39	1.51	.227	.032
Theme×Proficiency	476.13	1	476.13	.01	.906	.000

Note. η2 = effect size; 0.010 ≤ η2 (small) < 0.059, 0.059 ≤ η2 (medium) < 0.138, 0.138 ≤ η2 (large; Cohen, 1988, pp. 284-287); Effect size of 0.009 and below is regarded as having no effect.

As a result of multiple comparison with Bonferroni correction, the upper group showed a significantly shorter reading time for same-theme target than different-theme target ($p = .041$ and $p = .030$, respectively) when they were given a complete cue or predictive cue. However, no difference was observed with the partial cue between the two targets ($p = .572$). The lower group, on the other hand, showed shorter reading time for same-theme target than different-theme target ($p = .013$) only when they were given a predictive cue. In other words, no difference was observed with the complete cue and partial cue ($p = .941$, $p = .382$, respectively). This means that the upper group generated thematic inferences when given a complete cue and

predictive cue, while the lower group did so only with the predictive cue. However, when a partial cue was provided, thematic inference generation was inhibited. In addition, the upper group's reading time for same-theme target was shorter than the lower group under complete cue condition.

It was interesting to discover that both the upper and lower groups activated thematic inference generation when they were given a predictive cue. Under the predictive cue condition, all the cue information of goal, planning, and conclusion was not provided as in the complete cue, but an additional sentence that predicts the conclusion was provided. (e.g., *At that time, Salamians were not aware of the danger of forming an alliance with Bolonians.→ Bolon betrays the deal and attacks Salam.*) As a result, the readers are thought to have generated predictive inferences that activated possible conclusions in their minds. It is thought that the generated predictive inferences were re-activated when reading the conclusion sentence of the text that has the same theme, and therefore shortening the reading time.

In addition, under the complete cue condition, where all information on goal, planning and conclusion was provided, only the upper group generated thematic inferences. In other words, more proficient readers were able to combine the goal, planning and conclusion information and activate the comprehensive theme of the story. In contrast, the lower group found it hard to identify comprehensive theme without predictive inferences that strongly activate strong conclusion information. Under the partial cue condition, where information only on goal and planning was provided, neither the upper group nor the lower group showed any difference in the reading time for same-theme target and different-theme target. Based on the above, it is thought that a cue with just goal and planning information does not help the readers to grasp the theme that includes the conclusion and that activation does not occur.

5.3.2.3 Cued Recall Task

Table 5.16 shows the descriptive statistics for cued recall rates. Figure 5.5 shows average recalled production. Recalled production protocol data was analyzed using three-way ANOVA: 2 (theme: same, different) x 2 (L2 proficiency: upper, lower) x 3 (cue: complete, partial and predictive) (see ANOVA Table 5.17).

Table 5.16
Descriptive Statistics for Cued Recall Rates

	Same-Theme Target			Different-Theme Target		
	COM	PAR	PRE	COM	PAR	PRE
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Upper	.62	.49	.57	.45	.47	.49
(n = 25)	(.29)	(.20)	(.18)	(.19)	(.21)	(.27)
Lower	.41	.41	.53	.38	.48	.33
(n = 24)	(.22)	(.21)	(.20)	(.19)	(.20)	(.21)
Total	.53	.45	.55	.41	.48	.41
(N = 49)	(.20)	(.21)	(.19)	(.19)	(.20)	(.25)

Note: COM = complete cue, PAR = partial cue, PRE = predictive cue.

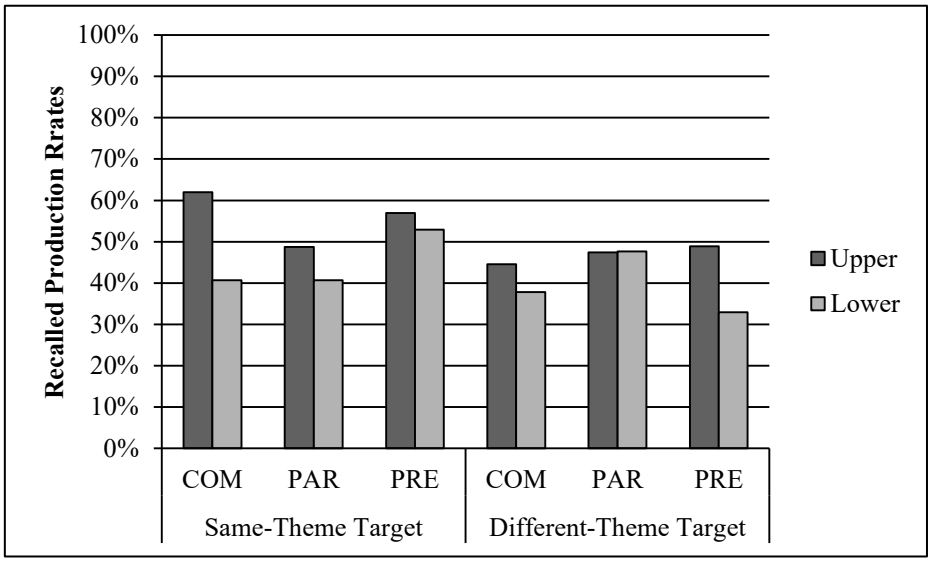


Figure 5.5. Cued recall rates for each condition.

Table 5.17
Results of Three-Way ANOVA for Cued Recall Protocols

	SS	*df*	*MS*	*F*	*p*	η_p^2
Within-Subjects Effects						
Cue	.01	2	.01	.17	.847	.004
Cue×Theme	.06	2	.03	1.08	.346	.023
Cue×Proficiency	.19	2	.09	3.22	.045	.067
Cue×Theme×Proficiency	.12	2	.06	201	.140	.043
Between-Subjects Effects						
Theme	.31	1	.31	5.03	.030	.101
Proficiency	.18	1	.18	2.95	.093	.061
Theme× Proficiency	.01	1	.01	.19	.663	.004

Note. η2 = effect size; 0.010 ≤ η2 (small) < 0.059, 0.059 ≤ η2 (medium) < 0.138, 0.138 ≤ η2 (large; Cohen, 1988, pp. 284-287); Effect size of 0.009 and below is regarded as having no effect.

As a result, the main effect of theme F (1, 45) = 5.03, p = .030, η_p^2 = .101, and cue x proficiency one-way interaction F (2, 90) = 3.22, p = .045 η_p^2 = .067 was observed. On the other hand, the main effect of cue, F (2, 90) = .166, p = .847, η_p^2 = .004, and the main effect of proficiency F (1, 45) = 2.95, p = .093 η_p^2 = .061, interaction with the theme, and the two-way interaction were not significant. It was found from the above that participants recalled and produced more same-theme target information than different-theme target information. Furthermore, cue × proficiency interaction was analyzed by multiple comparison with Bonferroni correction. The result shows the upper group recalled and produced more information than the lower group under the complete cue condition and the predictive cue condition (p = .019, p = .028, respectively).

No interaction was observed with the theme conditions. The main differences between the upper and lower groups were observed for same-theme target under complete cue condition, and for different -theme target under predictive cue condition. In other words, proficient readers recalled and produced the most amount of information of the same-theme target, when they

were given a complete cue containing goal, planning and conclusion information. On the other hand, the predictive cue is thought to have facilitated the recall and production of same-theme information for both the upper and lower groups. This effect of the predictive cue was not observed in the reading time data.

5.3.2.4 Theme Identification Task

Table 5.18 shows the scores for theme identification tasks. Table 5.19 is descriptive statistics for theme identification tasks. Figure 5.6 shows the mean score of theme identification tasks.

Table 5.18
Scores for Theme Identification Task

Target	Cue	Proficiency	Points		
			0	1	2
Same-Theme	COM	Upper	3	3	6
		Lower	4	4	4
	PAR	Upper	3	7	2
		Lower	5	6	1
	PRE	Upper	4	6	2
		Lower	1	3	8
Different-Theme	COM	Upper	3	6	2
		Lower	11	3	0
	PAR	Upper	5	4	2
		Lower	7	5	2
	PRE	Upper	4	6	1
		Lower	11	3	0

Note. 2 points were given to the correct theme, 1 point for the partial theme, and zero points for the wrong theme or no response.

Table 5.19

Descriptive Statistics for Theme Identification Task

	Same-Theme Target			Different-Theme Target		
	COM	PAR	PRE	COM	PAR	PRE
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Upper	1.25	0.83	0.92	0.80	0.60	0.60
(n = 25)	(0.87)	(0.72)	(0.67)	(0.63)	(0.70)	(0.52)
Lower	1.00	0.67	1.58	0.21	0.64	0.21
(n = 24)	(0.85)	(0.65)	(0.67)	(0.43)	(0.74)	(0.43)
Total	1.13	0.75	1.25	0.46	0.63	0.38
(N = 49)	(0.85)	(0.66)	(0.78)	(0.59)	(0.71)	(0.49)

Note. COM = complete cue, PAR = partial cue, PRE = predictive cue.

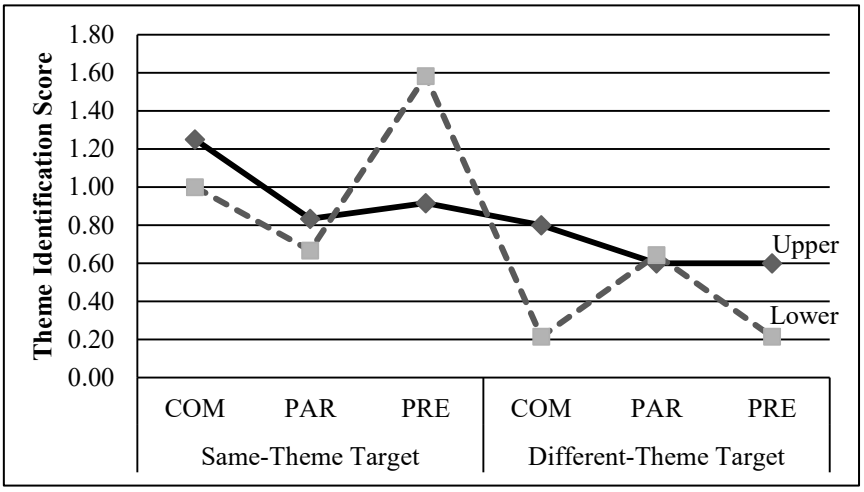


Figure 5.6. Mean score of theme identification task (full score = 2).

The scores were analyzed using three way-ANOVA of 2 (theme: same, different) x 2 (L2 proficiency: upper, lower) x 3 (cue: complete, partial and predictive) (see ANOVA Table 5.20). As a result, the main effect of theme F (1, 45) = 23.83, p < 001, η_p^2 = .351, and cue x theme x proficiency two-way interaction F (2, 90) = 3.31, p = .041 η_p^2 = .070 was observed. On the other hand, the main effect of cue, F (1, 45) = 1.09, p = .303, η_p^2 = .024, and the main effect of proficiency F (1, 45) = 3.28, p = .077 η_p^2 = .069, and the first-order interaction were not

significant.

Participants were able to write down the theme more accurately in same-theme target than the different-theme target. Multiple comparisons with Bonferroni correction identified differences in the lower group in theme identification tasks. The lower group showed a higher level of thematic understanding in same-theme target than different-theme target under the complete cue condition ($p = .007$) and predictive cue condition ($p < .001$).

Table 5.20

Results of Three-Way ANOVA for Theme Identification Task

	SS	*df*	*MS*	*F*	*p*	η_p^2
Within-Subjects Effects						
Cue	.35	2	.18	.38	.683	.009
Cue×Theme	2.49	2	1.25	2.72	.072	.058
Cue×Proficiency	2.13	2	1.07	2.32	.104	.050
Cue×Theme×Proficiency	3.03	2	1.52	3.31	.041	.070
Between-Subjects Effects						
Theme	9.96	1	9.96	23.83	.000	.351
Proficiency	.45	1	.45	1.09	.303	.024
Theme× Proficiency	1.37	1	1.37	3.28	.077	.069

Note. η2 = effect size; 0.010 ≤ η2 (small) < 0.059, 0.059 ≤ η2 (medium) < 0.138, 0.138 ≤ η2 (large; Cohen, 1988, pp. 284-287); Effect size of 0.009 and below is regarded as having no effect.

The level of thematic understanding was especially high under predictive cue condition, which was significantly high even in comparison with the upper group ($p = .003$). On the other hand, the upper group showed a high level of thematic understanding in same-theme target, but no difference was observed between different types of cues.

Below is an example concerning the effectiveness of the predictive cue in the lower group's thematic understanding. For example, in the predictive cue in Set A (see 5.3.1.2), the addition of a sentence, "they were not aware of the danger of trusting the neighbor," is thought

to have strongly activated the conclusion, “they are going to be betrayed by the neighbor and destroyed,” in the reader's mind. The activated information was combined with the target structure and encouraged thematic understanding. In fact, many responses by the lower group for same-theme target contained story conclusion and abstract theme such as “betrayal of deal” and “return kindness with hostility.” On the other hand, under partial cue condition without such additional sentence, responses tended to be just the themes derived from part of the text, without including conclusions such as “negotiation with the other party” and “mutual interest.”

5.3.3 Discussion for Experiment 5

This experiment examined what story component (goal, planning and conclusion) the readers use as a cue to access their knowledge structure in understanding the theme of a text. Reading time was measured to observe the process of thematic inference generation while reading, whereas cued recall tasks were carried out to examine whether the theme and the text are understood. In the theme identifying tasks, the theme that is consciously generated by the reader was qualitatively analyzed to examine the impact of the similarity between the text and the cue on the reader's thematic understanding.

In this section, the discoveries in Experiment 5 are described in the form of answering the research questions. Firstly, answers to RQ 5-1 (Do EFL learners activate thematic inferences in between-texts-level comprehension?) and RQ 5-2 (Is thematic inference activation influenced by structural similarity, type of cue information and L2 reading proficiency?) are given below. The result of the reading time showed that EFL learners activate thematic inferences during between-texts-level comprehension. However, it was also found that thematic inference activation is not triggered by goal and planning information alone, but it is triggered by the information that predicts the conclusion. When the cue contains goal, planning and

conclusion information which are all similar, only proficient readers activated thematic inferences. In other words, less proficient readers use predictive features that predict potential conclusion, rather than complete and similar information, as an index to access their knowledge information. On the other hand, a partial cue that contains goal and planning information only did not activate thematic inferences, and only when conclusion information was provided, was previous information searched. This coincides with the concept of case-based reasoning models (Johnson & Seifert, 1992; Schank, 1982), which suggests that people activate relevant knowledge that is relevant to the goal or based on predictive features to understand new information.

Secondly, answers to RQ5-3 (Is the level of understanding of the text influenced by structural similarity, type of cue information and L2 reading proficiency?) are given below. The result of cued recall tasks showed that more proficient readers recalled and produced more information when they were given a complete cue. A predictive cue did not result in a different amount of information recalled and produced between the upper and lower groups when the same theme was shared, but a difference was observed between the groups when the theme was not shared. In other words, as is the case with the reading time, a complete cue was effective only for the upper group, while a predictive cue was effective for both the upper and lower groups in recalling and producing information based on the theme.

Answers to RQ 5-4 (Is the understanding of the theme influenced by structual similarity, type of cue information and L2 reading proficiency?) are given below. Qualitative analysis of the theme identifying tasks showed that the upper group understood the theme correctly when the same theme was shared, but this was not particularly influenced by the type of cue information. In contrast, when the lower group was given a predictive cue, they understood the theme most comprehensively, including the goal, planning and conclusion information.

To summarize the results so far, it was found that EFL learners understand text that centers on the theme. Especially, proficient readers can generate thematic inferences when they are given the cues that share the same goal, planning and conclusion information. In contrast, predictive features that predict a story's conclusion strongly activate predictive inferences even for less proficient readers, and the understanding of the text with common theme also improves. In other words, EFL learners access knowledge structure using not only the goal and planning information, but also the series of causal associations as search cues.

5.4 Summary of Study 3

Study 3 (Chapter 4) was investigated how the text features work as access cues to the readers' knowledge structure using analogical texts. Analogical texts as advance reading organizer are used so as to strategically activate readers' prior knowledge. in between-text-level reading. Consequently, three main findings were indicated as follows:

1. Knowledge-based inferences caused by the transfer of analogical information were generated during reading comprehension. The proficient readers were not affected by the similarity types or the strength of causality, while the less proficient readers were affected by causality. Put differently, it is thought that the speedy access of information during reading was unrelated to the quality of similar information, and that similar information was activated associatively. However, for less proficient readers, the transfer of information among texts became more difficult when causality was low, and inferences were not generated as readily. At the same time, for recall of the text after reading, sematic similarity functioned as a strong cue for both proficient and less proficient EFL readers, though structural similarity functioned effectively for recall by only proficient readers.
2. The presentation of complete cues in which the causal structure in terms of the characters'

goals, actions, and outcome match across analogies triggered the transfer of information when learners subsequently read related texts, which shows that it facilitates the text deciphering process and the reproduction of information. This effect, however, was limited to proficient readers. On the other hand, it was shown that with only partial cues consisting of goals and actions, the activation of theme-related knowledge did not occur. From this, we can conclude that, for the understanding thematic structure (structural similarity), it is important for the causal relationship of goal-action-outcome information to be consistent with a reader's predictions.

3. Even when predictive cues were less complete analogies than complete cues, they facilitated the generation of thematic inferences most effectively for both proficient and less proficient readers. It is thought that this is because when reading an analogical text, predictive inferences are generated by predictive cues and strongly encoded as mental representation in the minds of the readers. The effect of facilitating the explicit understanding of texts was seen among less proficient readers in particular, which suggests that for EFL reading of narrative texts it is important for readers to make their predictions during reading.

Chapter 6

General Discussion and Conclusion

The present chapter overviews the actual condition of L2 reading proficiency of EFL learners who participated in this study, as well as the results of the five experiments reported in Chapters 3, 4, and 5 above, and comprehensively discusses the generation of knowledge-based inferences in Japanese EFL learners' reading comprehension.

The current research investigated the process of the knowledge-based inference generation which are drawn by connecting textual ideas and readers' prior knowledge in Japanese EFL learners' reading. In particular, it is focused on how (a) contextual constraints (CC), which is one of the text factors that strengthen the semantic relationship between textual information and readers' prior knowledge, and (b) EFL learners' L2 reading proficiency, which is one of the learner factors that have an impact on inference generation and encoding into mental representation. Also, in Experiments 4 and 5, analogical texts are used as advance reading organizers so as to strategically activate readers' prior knowledge. Showing an analogy activates the readers' schema related to sentence information, and understanding is then promoted by assimilating new information into the schema (Taniguchi, 1988). However, it was still unclear what kind of text information triggered activation of knowledge-based inferences in EFL reading. Thus, the present study focused on the effects of (c) similarity types (surface similarity and structural similarity) of analogy on knowledge-based inference generation.

In sum, the present study formulated three research questions (RQs):

RQ1: How do the text and learner factors affect EFL readers' knowledge-based inference generation?

RQ2: How does the presentation of analogy with either surface or structural similarities affect EFL readers' knowledge-based inference generation?

RQ3: How do the text and learner factors affect EFL readers' explicit text comprehension after reading?

In order to clarify these points, three studies were conducted. RQ 1 was investigated through Studies 1 and 2 (corresponding to Experiments 1, 2 and 3). The Study 1 examined sentence-level reading and Study 2 dealt with text-level reading. RQ 2 was investigated through Study 3 (Experiments 4 and 5). Study 3 looked at comparisons of texts-level reading using analogical texts. Moreover, RQ 3 was explored over all Studies 1 to 3 (Experiments 1 to 5).

6.1 Actual L2 Reading Proficiency of EFL Learners who Participated in This Study

In this section, the nature of the study participants' L2 reading proficiency is reviewed. The participants' proficiency was measured by using a common L2 reading proficiency test in four of the five experiments. A total of 34 test items were adopted from the Society for Testing English Proficiency (STEP) pre-Grade 1 (14 items), Grade 2 (17 items), and pre-Grade 2 (3 items) tests (2007). The test comprised of 14 fill-in-the-blank questions and four reading passages with comprehension questions from the reading section of the STEP tests.

Table 6.1 summarizes the descriptive statistics of the proficiency test in each of the experiments. Throughout Experiments 1, 2, 4 and 5, the mean score of the upper groups (i.e., proficient readers) ranged between 24.79 and 27.76 [with a difference of 2.97], whereas the mean score of the lower groups (i.e., less proficient readers) varied between 18.04 and 20.35

[with a difference of 2.31]. There was not much variation in the scores across different experiments.

Table 6.1

Summary of Descriptive Statistics From the Proficiency Test in Experiments 1, 2, 4 and 5

	Upper Groups			Lower Groups		
Experiment	*n*	*M*	*SD*	*n*	*M*	*SD*
1	20	26.95	1.99	19	20.35	2.71
2	20	25.20	1.82	22	19.36	2.49
4	21	24.79	1.68	23	18.04	2.07
5	23	27.76	1.71	26	18.86	3.72

Note. A full score for the test was 34.

Judging from these mean scores and the examiner's impressions, the upper groups answered almost all of the Grade 2 and pre-Grade 2 questions correctly, and approximately 40% of the pre-Grade 1 questions correctly. In contrast, the lower groups answered only a small percentage of pre-Grade 1 questions correctly although they answered Grade 2 and pre-Grade 2 questions correctly in most instances. Based on these findings, it is speculated that many of the upper group participants in the current study had L2 reading proficiency at the pre-Grade 1 or Grade 2 level, whereas many in the lower groups were at the Grade 2 level of the L2 reading proficiency.

The EIKEN Can-do list (STEP, 2008) includes the following Can-do statements for pre-Grade 1 competence in reading:

1. Can read different kinds of texts appropriately, adapting reading styles according to the type of text and purpose of reading (e.g., quickly skimming a newspaper, reading a critical essay in detail, reading a novel for pleasure).

2. Can understand the main points of lengthy texts (e.g., required readings and materials for lectures and training courses).

In contrast, Can-do statements (for reading skill) for Grade 2 competency include the following:

1. Can understand expository texts written for a general audience (e.g., guidebooks for travelers).
2. Can understand the main points of lengthy texts, provided the content is simple (e.g., required readings and materials for lectures and training courses).

In other words, if we are to define more accurately the actual proficiency levels, the readers in the upper groups of the current study can be described to have been *upper-intermediate learners* who had the ability to flexibly modify their reading styles and understand the main points according to the type of text and purpose of reading. The readers in the lower groups, on the other hand, can be described to have been intermediate learners, who can understand the main points as long as the text content is general and not too difficult.

The L2 reading proficiency test was not conducted in Experiment 3, due to certain constraints on the test implementation. For this reason, Experiment 3 was the only experiment in which no proficiency-based comparison was performed. Nevertheless, it is possible to estimate the approximate proficiency level of the participants in Experiment 3, based on their scores in the reading part of the TOEIC IP test. The reading section of the TOEIC IP test is determined by the number of correct answers, which is converted to a scaled score between 5 and 495, and a CEFR level. The mean score of Experiment 3 participants in the reading part was M = 196.35, with a SD =55.33 [score range: 110-295]. This score falls between the Grade

2 (TOEIC reading part conversion score = 229) and the Pre-Grade 2 (TOEIC reading part conversion score = 165), based on the score conversion table between TOEIC and Eiken, and it is at the A2 level, based on CEFR's assessment criteria.

Mapping the TOEIC® Listening and Reading onto the CEFR (Educational Testing Service; ETS, 2016) offers the following explanations about the characteristics of learners in these score ranges:

1. They can make simple inferences based on a limited amount of text.
2. They can locate the correct answer to a factual question when the language of the text matches the information that is required. They can sometimes answer a factual question when the answer is a simple paraphrase of the information in the text.
3. They can sometimes connect information within one or two sentences.

Based on these results, it is believed that the participants in Experiment 3 had reading proficiency that was either comparable to or somewhat lower than that of the participants in the lower groups in the other experiments. Such speculations, however, lack accuracy and for this reason no examination was performed in Experiment 3, including that on the factors of proficiency.

6.2 Summary of Findings

In this section, the author consolidates the findings of the five experiments and discusses the answer to RQs 1, 2 and 3 holistically.

6.2.1 Effects of Text and Learner Factors on EFL Readers' Knowledge-Based Inference Generation

The primary objective of the present study was to reveal the effects of specific text factor (CC: contextual constraints) and learner factor (L2 reading proficiency) on the knowledge-based inferences (thematic and predictive inferences) drawn by EFL learners from what they are reading. The effects of each factor exerted upon inference generation and the interactions among factors are discussed here.

In Study 1, the generation of thematic-based inferences in sentence-level comprehension was examined. The results revealed that participants responded to a thematic word (e.g., earthquake) at a significantly faster pace than to a non-thematic word (e.g., stranger) when reading the test passage (*The townspeople were amazed to find that all the buildings had collapsed except the bank. Obviously, it had been built to be ready for natural disasters*). High/low conditions of contextual constrains were established for these test sentences, based on the ease of drawing inferences. Additionally, participants were divided into two groups (upper/lower) based on their L2 reading proficiency. However, neither factor had any effect on the outcome. In other words, Japanese EFL learners generated thematic inferences under all conditions when comprehending short passages. This generation did occur regardless of the strength of contextual constraints and the level of L2 reading proficiency. The finding was unexpected in view of earlier studies. The following table compares the outcome of the earlier studies with that of the current study. Table 6.2 shows the comparison of findings from earlier studies and those of Study 1. In L1 studies, Long et al. (1994, 1997), as well as Hannon and Daneman (1998), conducted tests by using a lexical decision task as was done in Study 1 of the current study.

Table 6.2

Comparison of Findings from Earlier Studies and Present Study 1

	Activation of Thematic Inferences	
	Upper (Proficient readers)	Lower (Less proficient readers)
Long et al. (1994, 1997)	Activated	Not activated
Hannon and Daneman (1998)	Activated	Conditionally activated (Only when the text incorporated a question inviting the inferences and text presentation speed was slower (400ms/word))
Present Study 1	Activated	Activated

Both studies showed evidence that only more proficient readers generated thematic inferences during comprehension. Long et al. (1994, 1997) compared good and poor readers on a lexical decision task to words that were relevant and irrelevant to the topic of a discourse. The researchers found that high-skill readers executed a lexical decision task more quickly to topic-relevant words than topic-irrelevant words when the stimulus onset asynchrony (SOA) between the end of the passage and the target word was 500 ms. In contrast, low-skill readers showed no difference in lexical decision time to relevant and irrelevant topic words even with an SOA of 1000 ms, indicating that low-skill readers had much slower online access to topic level information. This finding is consistent with Kintsch's (1998) argument that high-skill readers are able to access context-based inferences automatically whereas low-skill readers do so in a controlled manner (p. 283). Furthermore, Hannon and Daneman (1998) conducted a follow-up experiment, which revealed that even less proficient readers were able to generate thematic inferences when given appropriate textual support and adequate text presentation time (400ms per word) during comprehension. In their test, the insertion of a question inviting the inference

in the test text (e.g., "The old woman awoke and said, 'Why is there a sound downstairs?' She reached into her purse and found only a file.") promoted the response to theme-appropriate targets (e.g., burglar).

Why is it, then, that in the current study the generation of thematic inferences was unaffected by proficiency? One possible reason is that the content of the text used in the test material was inductive to the generation of thematic inferences. In the previous studies by Long et al., (1994) and Hannon and Daneman (1998), both of which used the same text as that used in the current study, a word in the test passage was replaced with its homograph for the purpose of examining whether comprehension failures had been caused either by the absence of appropriate construction of representations or by the lack of thematic inference generation. For example, the following passage was presented: "The townspeople were amazed to find that all the buildings had collapsed except the *mint*. Obviously, it had been built to be ready for natural disasters." The word *mint* that appears in the passage has two semantic meanings: (1) a small plant with green leaves that have a fresh smell and taste and which is used in cooking, and (2) a place where coins are officially made. However, only the second meaning is appropriate in this context. In their test, an appropriate (i.e., money) or inappropriate (i.e., candy) association of the homograph targets was presented to determine whether or not appropriate representation construction occurred, based on the response time to these targets. The test participants in the earlier studies thus had to undergo a process of identifying the meaning of homographs while reading. In the present study, words in the test text were replaced with their synonyms rather than their homographs, which have dual meanings. The reason is that the objective here was to clearly reveal the thematic inference generation processes. At the same time, there was a possibility that the test participants, who were EFL learners, were unfamiliar with the multifarious meanings of homographs, especially secondary meanings. In other words,

processing of homographs was not required in the current study, thereby reducing the recognition processing load during reading. This may have contributed to the activation of thematic inferences.

Additionally, a pilot study was conducted during the material preparation phase in the current study. As a result, multiple passages were found that were deemed to be too difficult for Japanese participants to generate inferences, due to their different cultural backgrounds (e.g., "The rabbi looked for something to use in putting up his announcement. Finally, he used his shoe to tack it to the door of the temple." The inference target is *Church*). Elimination of such passages resulted in the eventual reduction of the total number of passages from 56 to 28. Therefore, there is a possibility that the activation of thematic inferences was promoted as texts that did not lend themselves well to inferences were removed at the outset.

A second reason for the ease of thematic inference generation by EFL learners is the different way in which test passages were presented. In earlier studies, a rapid serial visual presentation (RSVP) format was used to present the test passages on a PC screen one word at a time. Hannon and Daneman (1998) showed that even less proficient learners formed thematic inferences at a presentation rate of 400ms per word but not at 300ms per word. In other words, a slower presentation rate in combination with a text manipulation, such as the integration of a question inviting the inference, was found to be necessary for inference generation by less proficient readers. Another point that should be noted is the fact that the RSVP format creates a peculiar reading condition although it is effective in controlling comprehension conditions, as mentioned in their studies. The present study adopted a self-paced presentation format to proximate a more natural reading situation. Using the self-paced presentation format, participants read one sentence at a time at their own pace, and pressed a button to move forward. This allowed participants to rehearse the already-read content and also go back to re-read

sentences. It is believed that this enabled readers to retain text information accurately and generate thematic inferences more readily.

To summarize, the study found that Japanese EFL learners generated thematic inferences during sentence-level comprehension irrespective of coherence and proficiency factors as long as they did not need to divert part of their cognitive processing to the processing of homographs; that there was adequate information to generate inferences; and that the reading conditions were natural.

Is there, then, any influence of text and learner factors on text-level comprehension as opposed to sentence-level comprehension? In sentence-level reading, processes are necessary to form structures that represent the syntactic and conceptual relationships among words in phrases or clauses. These processes help in encoding propositions, abstract units that represent the meaning of a sentence. In text-level reading, processes become more complex and are required to form connections among successive propositions in a text. The processes mentioned above involve accessing not only knowledge about the language (e.g., word meaning, syntax), but also wider knowledge about the world. In Study 2, the formation of predictive inferences in text-level reading was examined. Unlike the short passage used in Study 1, which consisted of two sentences, the material used in Study 2 was a text consisting of between 15 and 17 sentences (containing between 201 and 212 words) and including background information about characters and scenes. Sufficiently-restricted predictable outcomes and the readers' having adequate prior knowledge about the content are prerequisites for predictive inferences to be drawn online (e.g., Calvo et al., 1999; Graesser et al., 1994). Accordingly, high and low conditions were established for contextual constraints (CC). In the high CC condition case, contextual information was added to help predictive inferences to be readily generated. For example, Brad's story, which is one of the passages used in the text, narrates that "Brad, the

main character, wants to buy a birthday present for his wife but he does not have much money. He happens to notice a jewelry store showcase and approaches it." In the high CC condition case, a sentence that reads, "*He had been fired from his job three months ago and he could not afford to buy anything nice*," was inserted, manipulating the ready formation of a predictive inference, which is that the main character will steal jewelry. In contrast, a sentence that read, "*He had just started a new job but had not received his first salary yet and he was not sure if he could buy anything nice*," was inserted at the same point in the low CC condition text, manipulating so that the ease of forming a predictive inference was lessened. In this test, three types of probes – inference, control, and verbatim repetition – were used to perform the probe recognition task. The results showed that participants in the upper proficiency group had a significantly higher rate of a "Yes" response to the inference probe than the lower proficiency group did under the condition of high coherence. In contrast, no differences were found with respect to responses to the control probe or the verbatim probe between the two proficiency groups, both of which responded accurately. Two points were thus clarified: (i) participants constructed an accurate representation for information that was written explicitly, and (ii) only the upper group generated predictive inferences.

In Experiment 3, the same material was used to examine the specificity of the predictive inference content. The results showed that the greater the contextual coherence, the more likely is the identification of a specific predictable outcome (e.g., Brad steals a ring) while weak coherence led to more divergent answers (e.g., Brad buys a present, or steals something) and ambiguous inferences. The findings of Experiment 3 clarified the important role that contextual coherence plays in the generation of predictive inferences that do not contribute to the preservation of textual consistency. They also show the need for a certain minimum level of L2 reading proficiency for this generation.

In sum, the results of Study 1 suggested there were no differences in the two groups' ability to generate knowledge-based inferences. Both proficient and less proficient EFL readers were sensitive to the thematic inference items. After examining their representations of simple sentences, the author found remarkable similarities in their representation. Both groups constructed representations that were consistent with the context-appropriate senses of inference target words. However, in Study 2 (Experiments 2 and 3), which dealt with text-level reading, there were large differences in the extent to which proficient and less proficient readers generated knowledge-based inferences, particularly with respect to their ability to elaborate their representations with predictive inference information. These differences occurred in spite of the fact that both levels of readers had adequate knowledge about the texts and constructed reasonably accurate sentence-level representation.

This result is similar to other previous L1 research that have examined a variety of reader characteristics. Cognitive researchers have documented many differences between proficient and less proficient readers that may be relevant to the generation of predictive inferences. For example, Murray and Burke (2003) explored whether high- and low-proficient readers differ in their tendencies to activate and encode predictive inferences. Their results showed that only high-proficient readers showed evidence of automatic activation, but all readers, regardless of skill, showed evidence of inference encoding. These findings support the conclusion that reading skill differences are likely to be manifest at the level of activation and not at the level of encoding. In addition, Calvo (2001) and Linderholm (2002) have found that high working memory (WM) span readers may be more likely to activate predictive inferences than low WM span readers. Furthermore, Calvo, Estevez, and Dowens (2003) found that readers with high vocabulary knowledge are more likely to immediately activate predictive inferences. They have argued that low-skill readers utilize lexical representation of poor quality, which leads to

difficulties in word identification. These difficulties, in turn, can affect higher-level inferential processes.

In the present study, why did less proficient readers fail to generate knowledge-based inferences in text-level reading? One possibility is that they failed to integrate information from different parts of a text at a global level. Although there is a large consensus on the idea that readers connect each current processing unit to the immediately preceding ones (i.e., local coherence), several theories of comprehension have evoked certain conditions under which distant units in the surface text structure are related (i.e., global coherence) and then incorporated into the reader's mental representation. For instance, according to the minimalist view, such connections are more likely to be made when participants are engaged in special comprehension strategies (McKoon & Ratcliff, 1992). The experimental situation, including the texts and the nature of the task performed, may encourage readers to access distant information and build a global representation. A break in local coherence may also increase the likelihood that readers will seek to establish global coherence. From the same minimalist view (McKoon & Ratcliff, 1992), inferences are drawn only when they are either needed to maintain the local coherence of the text (i.e., automatic inferences), or readers have easily available information. Therefore global coherence depends on non-automatic strategic processing and needs to be established only when local coherence fails. Conversely, the memory-based view of reading comprehension holds that inferences are activated automatically due to the spread of activation from textual ideas that are currently active in working memory to associated ideas in background knowledge via long-term memory networks (e.g., Myers & O'Brien, 1998; O'Brien & Myers, 1999; van den Broek, et al., 2005). In favor of this view, a large body of research has shown that elaborative inferences such as predictive inferences are more reliably activated when they constitute strong semantic associations with specific textual ideas (e.g., Calvo & Castillo,

1996; Keefe & McDaniel, 1993; McKoon & Ratcliff, 1986; Murray et al., 1993; O'Brien & Albrecht, 1992; O'Brien, Shank, Myers, & Rayner, 1988). For example, when reading about a delicate porcelain vase that is being thrown against a hard surface, readers routinely infer the strongly associated outcome of the vase breaking, although this predictive inference is not necessary for understanding the inference-evoking information. More evidence supporting the memory-based view comes from studies that have demonstrated the contributions of textual elaboration (e.g., Albrecht & Myers, 1998; Rizzella & O'Brien, 1996), textual cues (e.g., Albrecht & Myers, 1998) and feature overlaps of textual ideas (e.g., Duffy & Rayner, 1990) to the reactivation of prior textual ideas in working memory.

Furthermore, in recent years, the integrative reading models that examine the interaction of associative and strategic processes have proposed (e.g., Gerrig, & O'Brien, 2005; Long & Lea, 2005; van den Broek et al. 2005). That is, it has been suggested that associative processes are the default primary processes that take place continuously during reading without control or effort on the part of the reader, whereas strategic processes are initiated by the reader when the products of the associative processes are not sufficient to attain the desired coherence and comprehension (e.g., van den Broek, Bohn-Gettler, Kendeou, Carlson, & White, 2011; van den Broek, Lorch, Linderholm, & Gustafson, 2001; van den Broek et al., 2005). The findings of Studies 1 and 2 are interpreted in accordance with these integrative models as follows:

In sentence-level reading (Study 1), readers generated inferences associatively so as to maintain local coherence. In text-level reading (Studies 2 & 3), on the other hand, global coherence needed to be constructed strategically across sentences. Inferences are generated by overlapping textual information in the participants' working memory with their prior knowledge through the network in their long-term memory. Essentially, such informational overlapping occurs automatically and unconsciously. When the text volume increases or the

causality of textual information is weak, however, the possibility of strategic processing grows.

Furthermore, the fact that the predictive inferences were of the type where contextual coherence strongly influenced their generation is believed to have influenced the outcome. In such an event, the stronger the causal relationship (the "goal → action → outcome information") of the characters appearing in the text, easier is the generation of predictive inferences. Multiple earlier studies that attempted to determine the factors causing activation of knowledge-based inferences by using Latent Semantic Analysis (LSA) identified textual semantic constraints to be a major factor that induced the activation of inferences (Yeari & van den Broek, 2015). Text manipulations of contextual coherence performed in Studies 1, 2 and 3 were made so as to compare reading conditions with strong causality and those with weak causality. It is believed that learners in the upper group tended to respond with heightened sensitivity to the contextual coherence and strategically generate predictive inferences.

6.2.2 Effects of the Presentation of Analogy with Either Surface or Structural Similarities on EFL Readers' Knowledge-Based Inference Generation

The pair of Experiments 4 and 5 (Study 3) was designed to clarify the answer to RQ 2: How does the presentation of analogy with either surface or structural similarities affect the generation of knowledge-based inferences? Analogical texts were used as experimental materials in order to verify the role of analogical texts as organizers of advance reading. Previously acquired knowledge—the source analog can be mapped onto another domain—the target is referred to as *information transfer* (Gentner, Rattermann, & Forbus, 1993). With this information transfer, there is a possibility that knowledge-based inferences are generated on the

basis of reactivation the information of the source analog. It is important to understand how inferences are generated between-text levels because it is a powerful mechanism for the acquisition of new knowledge and conceptual change (Blanchette & Dunbar, 2002). Therefore, the present research attempted to clarify what kinds of text information in analogy promote inference generation.

Analogical texts in this study were pairs of texts in which two types of features (i.e., surface and structural similarities) are shared across different texts. Furthermore, these experimental stories were based on thematic abstraction units (TAUs), which contained shared relationships between concepts, such as interactions between a character's goal, action and outcome in a storyline (Dyer, 1983). TAUs are based on abstract interactions of goal, action and outcome as reflected in familiar cultural adages (e.g., *counting chickens before they have hatched*) so as to investigate whether thematic text features lead to useful access to readers' knowledge structure.

After performing experiments using recognition tasks and cued recall tasks, it was found that knowledge-based inferences relying on information transfer via analogy were generated during reading comprehension. The results showed that, among the experiment participants, the upper proficiency group was not affected by the type of similarity or the strength-weakness of the cause-effect relationship, but the lower proficiency group was influenced by the cause-effect relationship. In other words, it is possible to conclude that, in accessing information for quick comprehension, regardless of the *quality* of similar information, it is the *associative* property of similar information that becomes activated. With readers of lower proficiency, however,

when the degree of causality decreases, it became more difficult to make this information transfer, resulting in fewer inferences generated. At the same time, with regard to the recall of textual information after reading, surface similarity functioned as a major key in both high- and low-proficiency groups, but structural similarity functioned effectively in only the high-proficiency group. Many of the previous L1 studies reviewed acknowledge that lexical-level processing generates some inferences. Graesser et al. (1994) viewed vocabulary as the starting point in activation of background knowledge. The recognition of the role that explicit content words, combination of content words, and interpreted text constitutes is crucial to knowledge-based inferences. Furthermore, Experiment 4 also found that causal information and a reader's L2 reading proficiency level are strongly related in identifying structural similarities between text information. Also, the quality of recalled information represented in the upper group tended to include more important information when the causality was stronger under structural similarity conditions. This effect, which did not influence the lower groups' production, occurred because proficient readers were more aware of the causal relations underlying the story they were reading. This promoted the notion of analogical transfer of causally relevant information to similar text structures. To summarize these points, analogical transfer on the basis of surface similarity occurred easily, irrespective of L2 reading proficiency. On the other hand, analogical transfer based on structural similarity occurred with a certain level of L2 reading proficiency and causality between two analogical texts as the key factors.

To summarize, it was found that providing analogous information prior to the reading of texts could promote the generation of knowledge-based inferences during the reading of a

related text. That generation process was not found to be influenced by the nature of the similarity of the text (surface or structural) nor the reader's L2 reading proficiency level. In other words, the reactivation of similar information during the reading of related texts occurred associatively and automatically. This is consistent with the process posited by the memory-based text processing view in which existing knowledge is activated, i.e. that if it is information that shares many similar characteristics with the text information, it is all activated, even if it is not appropriate to the current process (e.g. information that is related superficially but is irrelevant situationally) (O'Brien & Myers, 1999).In this process, when new concepts are encoded into working memory, a signal is sent from these concepts to all of long-term memory in parallel (both the episodic memory trace and general world knowledge). Concepts in long-term memory that share features in common with the contents of working memory will "resonate" passively in response and those concepts that resonate the most are also most likely to be incorporated into working memory. O'Brien, Cook, and Peracchi (2004) provided the evidence that information from earlier portions of a text are reactivated when the reader encounters related information; this occurs independent of whether that reactivated information facilitates or hinders the processing of current information. Readers' prior knowledge appears to influence processing at the feature level. Although "features" have never been specifically defined in the discourse processing literature, many theories agree that they are an important element in reading (e.g., construction-integration [CI] model Kintsch, 1988). Features involve more basic units of meaning than can be captured by propositions or lexical items. That is, concepts are made up of clusters of features and features can overlap among concepts.

At the same time, in the context of the activation of textual information after reading, there was interaction observed between text factors and learner factors. A text's surface similarity was an effective means of information activation for both the hither and lower proficiency groups, but structural similarities used with the higher proficiency group produced the most instances of information activation, and a significant difference was seen between this and the lower proficiency group. Structural similarity is the removal of individual features of a text and its reduction to abstract information as in terms of causality between propositions. Based on this, the search of long-term memory should be an advanced process that can only be performed by advanced learners. In other words, the ability to perceive cause-and-effect relations among texts differs from reader to reader, and in EFL reading in particular is related to reading skills.

In the area of reading comprehension, L1 studies have sought to examine the standards of coherence among how multiple readers perceive the relations among the information within a single text. Standards of coherence to "the types and strengths of coherence that the reader aims to maintain during reading" (van den Broek et al., 2001, p.124). These standards consist of a set of implicit or explicit criteria that the reader adopts for the particular reading situation, reflecting the desired level of understanding. They vary between individuals as well as within an individual from one reading situation to the next. The standards influence the dynamic pattern of automatic and strategic cognitive processes that take place during reading. They are influenced by various aspects of the reader, the text, and the task. The role and properties of standards of coherence are best understood in the context of current models of discourse processing.

A lot of prior research has shown that readers are sensitive to causal relations between sentences. In addition, the extent to which readers put weight on causal relations seems to depend on their reading skill. For example, Wittwer and Ihme (2014) reports an experiment that examined the influence of semantic similarity and causal specificity on the perceived comprehensibility and goodness of causal explanations. The results showed that semantic similarity in the form of an overlap of nouns between adjacent sentences exerted more influence on less proficient readers' perceptions of the comprehensibility of explanations. Conversely, causal specificity in the form of verbs indicating causation with a particular result more strongly impacted more-proficient readers' perceptions of the goodness of explanations. The results suggest that, depending on their reading skill, readers have different standards of coherence that influence their judgments of different aspects of coherence. The findings are discussed in the context of validating the epistemic status of science-related explanations.

The standards of coherence mentioned above involve the maintenance of a consistency or coherence within a single text. The standards of coherence, however, may be similarly applied to the processes that link different texts together. Hence, in this study, surface similarities (semantic similarity) between the analogies were recognized with ease even by less proficient readers, but structural similarities (causal specificity) were perceived more easily by more advanced readers, and this result may be related to a standard of coherence that differs from individual to individual.

We should also note that Forbus, Gentner, & Law (1994) proposed what they called the *MAC/FAC model* for processing analogies. The term is an acronym for “many are called, but

few are chosen." In this model, a reader selects from a vast wealth of candidates those works that are analogically based, which later leads to analogical/deductive reasoning. This is a model that is strikes a good balance between high-speed searches based on surface similarities and evaluations and responses to structural similarities. As a result of such searches, most of the information utilized from memory at search time is useless, and a rough search based on surface similarities at the MAC level begins. At this point, evaluations based on structural similarities at the FAC stage are resolved with the respective target. In other words, the company still maintains the advantages in terms of superficial superiority upon querying, as well as the advantage of structural superiority when evaluating the base value of the projects.

The purpose of the subsequently performed Experiment 5 was to study in more depth the function of structural similarity examined in Experiment 4 as a cue. Using chiefly the main themes from literary tales, it was extracted the attributes of characters according to goals, actions, and outcome (Zhang & Hoosain, 2005). Here, a structural similarity refers to a similarity in the relationship among the structural elements. It was examined whether the addition of an analogy with structural similarity facilitated the generation of knowledge-based inferences. As a result it was found that the presentation of a complete cue—conveying that all of the elements, namely the goal, action, and outcome of the main character coincide—caused information transfer when reading a related text later, hence facilitating the deciphering process and the reuse of information.However, this effect was limited to proficient readers. On the other hand, it was shown that when there were only partial cues such as the goal or action, knowledge related to the themes was not activated. In previous studies as well, thematic inferences during

reading comprehension by EFL learners were generated when information on the goal and outcome of the main character coincided (Zhang & Hoosain, 2005), suggesting that the goal-action-outcome information being consistent with the reader's predictions is important for the grasping of thematic structure (structural similarity).

Furthermore, even though it is not as easy to establish themes with predictive cues as with complete cues, they functioned most effectively for both high- and low-proficiency groups. This is probably because, when deciphering cues, predictive inferences are generated and encoded into representations, and then reactivated during target comprehension. And because facilitative effects were seen in the low-proficiency group in particular, it is highly likely that predictive inferential information is strongly encoded as mental representations. This coincides with the concept of case-based reasoning models (Johnson & Seifert, 1992; Schank, 1982), which suggests that people activate relevant knowledge that is related to the goal or predictive features to understand new information.

Even with incomplete cues, if there is predictive information given as cues, information access did occur. At the same time, with just partial cues, information was not accessed as efficiently. From this, it is clear that even before the thematic structure is input completely into the mind of the reader, based on predictive information, outcome information is generated through predictive inferences and can be accessed in memory. Meanwhile, with only partial cues, because information was not accessed effectively, it is clear only when the reader perceives the cause-effect relations of the goals, actions, and outcome in the story line can similar examples be searched. From this, it is possible that when the reader is understanding

new circumstances, a certain planning knowledge, which can be characterized as cause-effect features in terms of what should (or shouldn't) be done, is being activated.

This study showed the importance of understanding goal planning information, in the searching done on the part of the reader during reading comprehension. When EFL learners search their base of knowledge for examples during reading comprehension, rather than having certain information on planning given to them predictively, they base their searches on complete cues that are more trustworthy.

In conclusion, information transfer that occurs across texts in Japanese EFL learners takes place as follows. Fundamentally, the semantic similarities of propositions cause the activation of similar examples that exist in the memory of the reader, but it became clear that more proficient readers can use similarities in cause-effect relations across texts to search for examples. In addition, for such searches, it is important for all cue information—in terms of what kinds of goals the characters have, what actions they take, and the outcome of those actions—to consistently match. However, searches are performed most effectively when there is predictive information in the text and the reader is able to predict conclusions spontaneously.

6.2.3 Effects of Text and Learner Factors on EFL Readers' Explicit Text Comprehension After Reading

To measure EFL readers' explicit text comprehension after reading, cued recall tasks using inferential information and analogical information as cues were performed in Experiments 1-5. It was examined how text factors, such as the strength of contextual

constraints and types of analogical information, and learner factors, such as degree of L2 reading proficiency, affect EFL readers' explicit text comprehension after reading. As a premise, it was assumed that as long as analogical information generated by certain text factors is encoded EFL learners' mental representation, the performance of information reproduction through cued recall task using analogical information will improve (Klin, 1995; McKoon & Ratcliff, 1986; Murray & Burke, 2003).

As a result, it was found that, regardless of the unit of reading (sentence or text) and inference type (thematic inference or predictive inference), contextual constraints were a major factor in the rates of reproduction of explicit textual information. The results of the experiments are summarized in the following. In Experiment 1, which looked at thematic inferences in reading comprehension at the sentence level, contextual constraint factors did not affect the generation of thematic inferences during reading. However, in recall tasks after reading, in high contextual constraint conditions for links between the text and target words, more textual information was recalled (recall production rates in high CC condition: M = 65%, low CC condition: M = 50%). In Experiment 2, which examined predictive inferences in reading comprehension at the text level, in high contextual constraint conditions, regardless of proficiency level, about half of the information was reproduced (recall production rates by upper group: M = 52%, lower group: M = 45%), while in low contextual constraint conditions, there were differences based on proficiency levels. In other words, text representations encoded into the long-term memory of EFL learners was structured around causal connections. The purpose of the manipulation of contextual constraints is to more strongly link the goals and

actions of the characters in the story with the outcome and thereby generate inferences, which is the same as increasing the causal coherence of the text. EFL learners build text representations based on such causal coherence, and in high contextual constraint conditions, more textual information, including predictive information, is encoded into mental representations. For this reason, the coherency of the mental representation in the mind of the reader increased in proportion to the strength of contextual constraints, and explicit text comprehension was facilitated. In sum, contextual constraints do not necessarily affect the activation process of predictive and thematic inferences during reading, but do necessarily affect the retrieval process of text information from long-term memory. It also became clear that this influence is more prominent in longer texts, where the causal relationships among the components of the story are more complex, than in shorter ones, and also more prominent when the L2 reading proficiency of the learner is lower.

Furthermore, the result of the inference question task in Experiment 3 showed that as contextual constraints increase, the specificity of the activated inferences also increases. EFL readers tended to generate predictive inferences with a more definitive and specific image when the contextual constraints were high, while they generated a more general and ambiguous image about the story developments when the contextual constraints were low. In L1 research, it has been shown that high contextual constraints contribute to the improvement of text coherence and generation of specific inferences, whereas low contextual constraints are likely to give rise to general inferences (Lassonde & O'Brien, 2009). Moreover, Sanford and Garrod (2005) proposed the granularity hypothesis, where level of granularity refers to the

specificity of the semantic representation. At a “crude” level of granularity, specificity is low; activated inferences would tend to be more general and could be reflected in the activation of several different lexical items. On the other hand, at a “finer” level of granularity, specificity of the semantic representation increases; inferences would also tend to become more clearly defined and activation would be captured by a smaller set of lexical items.

Interesting results were obtained via cued recall tasks. In Experiment 3, two subtypes of predictive inferences: “consequence-” and “motivational” predictive inferences were examined. Consequence-predictive inferences are simply the consequence of the events describing the story. When reading Jimmy’s story (see Appendix 3: Text A: Jimmy’s story), readers may generate the inference that the door of the new car is going to be damaged automatically. This is due to a build up of the activation of the inferred concept; damage is related to the text information: ‘the new car was made of soft metal’ and ‘Jimmy begins playing a game with the other children that involve throwing rocks at a target for points’. So ‘Jimmy accidentally damaged the door of the new car’ is a highly probable consequence of the rock being thrown, and it serves to elaborate on the learners’ mental representations. On the other hand, motivational predictive inferences are generated from an understanding of the causal relationship between the character’s motivation and the consequence. When reading Brad’s story (see Appendix 3: Text B: Brad’s story), it is not only the predicted action of stealing the ring. The inferred consequence of Brad walking quietly over to the counter is also the motivation or the cause. Brad is quietly making his way to the counter because he intends to steal the ring. Without the inference, the second sentence creates a break in the causal coherence

of the passage; there is no cause for Brad's actions. Because these predictive inferences are needed to maintain the coherence of the text, it has high possibility of being generated by the readers.

Analysis of cued recall tasks using inference-related sentence cues suggested that participants recalled more text information when the inferences were motivational and strongly constrained by the context. In contrast, with consequence-inferences, there was no difference of recall production rates between the strength of the contextual constraints. This result showed that readers are more likely to generate motivational inferences than consequence-inferences and encode the inferred information into their text representation in long-term memory. Why did motivational inferences promote the production of explicit text information?

One possible answer is that EFL readers are more likely to generate predictive inferences when they are related to narrative characters' goals or motivations and their understanding of stories in long-term memory are organized mainly based on causal connections. The expected inference in Text B: Brad's story (i.e., stealing the ring) is not only a possible outcome of the described event, but also the motivation or cause of the described character's actions (i.e., approaching the display and opening his bag). In other words, the inference explains why the character performed the intentional actions described in the passage. Thus, this inference is necessary to maintain the local coherence of the text and readers reactivate information from long-term memory to draw causal inferences when there is a break in causal coherence.

In Section 6.1.2, The author address in detail the findings of the examinations using analogies. In Experiment 4, it was shown that information transfer occurs easily across

analogous texts with causal relations, and that much explicit text information is recalled. At the same time, however, when two texts had different domain shares underlying structural features but not surface features, only proficient readers were able to recall much text information. Experiment 5 showed that text information with predictive features facilitates information transfer across texts and increases the production rates of recalled text information where themes are similar.

To summarize the findings so far, text representations of stories for EFL learners are formed mainly based on causal relations among the story's components, namely the goals and actions of the characters and the outcome. Raising contextual constraints does not necessarily affect the activation of knowledge-based inferences. The production of predictive inferences requires not only that the outcomes be limited to those that can occur due to context, but also that the reader has a certain level of L2 reading proficiency. At the same time, when retrieving text information after reading, contextual constraints had a strong influence regardless of proficiency. The text factors of contextual constraint were an increase in coherence related to goal-planning on the part of the characters, which led to stronger text representation. It became clear that the motivation or cause of the described character's actions in particular tended to stick in the memory of the reader.

Furthermore, the things learned about the process of reactivation of information from one text when the reader is reading another text are as follows. Fundamentally, when reading, information in the network that shares words and premises that are in the same domain is reactivated in the mind. However, with proficient readers, even if the text information at first

appears to be of a different domain, as long as the causal relations at the root are shared, those structurally similar elements of information can be reactivated. In other words, it is possible that, as EFL learners grow in proficiency, the processes by which they encode overarching causal relations and thematic structures into mental representations, and transfer that information to other texts, occur much more naturally and frequently. Also, among different texts that share thematic structures based on goal planning, only EFL learners that are highly proficient can reactivate information that is structurally similar. However, when there is information in a text of the kind that would lead a reader to predict the outcome, even readers of low proficiency are able to generate predictive inferences, which then facilitates an understanding of the thematic structure. From these findings, it was shown that the inclusion of predictive information about the outcome within the text of stories and the asking of questions that urge the reader to make predictions could aid in text comprehension.

6.3 Limitations of This Study and Suggestions for Further Research

Five experiments were conducted in the current study to examine the process of knowledge-based inferences in EFL reading comprehension. In this section, the limitations of these experiments are described, and suggestions for future research are provided.

First, throughout all of the experiments, after the experiment texts were read, cued recall tasks were performed. In addition to measuring explicit text understanding, one of the purposes of these tasks was to measure how the strength of the inference information activated by the readers was encoded into a mental representation in their long-term memories. If the inferred information becomes part of the mental representation, then, in a recall task, participants should

have difficulty distinguishing between propositions that were explicitly presented and those that were inferred, leading to intrusions in recall protocols. (Klin et al., 1999). However, all of the experiments for this study were conducted in one session, so under the conditions of immediate recall after reading, such inference intrusions were hardly seen. For this reason, in the future, it is necessary to examine whether inference information is being encoded into a mental representation in their long-term memories by using delayed recall tasks performed a week later.

Second, with the exception of the short text used for Experiment 1, all of the texts used for the experiments were narrative texts. The discourse structure differs between narrative texts and expository texts, and the reader must be aware of these differences. With narrative texts, it is possible to make predictions based on coherence in terms of cause and effect with regards to the intentions, goals, and actions of the characters in the story. Expository texts, on the other hand, are more varied in structure, and it becomes particularly important for readers to grasp the relationships among various pieces of information (cause-effect, etc.). On the effect of text genre on the L2/EFL reading process, Grabe (2000) suggested that the skill to recognize text genres and text types is important in comprehend the text successfully. Horiba (2000) also clarified that L2 readers are affected by text types. Contextual constraints and the effects of presenting analogies examined in this study may require different conclusions in the context of expository text reading. Going forward, it is necessary to compare how texts that differ in structure, such as narrative texts vs. expository texts, are processed.

Third, for the texts used in the experiments in this research, to ensure that the passages were sufficiently understandable by EFL learners, I replaced difficult words and expressions with simpler ones. It is a well-known fact that there are limitations in using materials that are only easily understandable for learners in actual EFL classroom situations. When teaching authentic materials such as newspapers and magazines, we often simplify the sentences or

vocabulary so they will be suitable for learners. However, such manipulations of simplification of the text could be a heavy burden on the part of teachers, in terms of time and the additional work required. Also, this may cause the problem of narrowing the range of language expressions that can be learned from text. One of the important skills on the part of the learners is how to deal with linguistic expressions that they are not familiar with. In L2 reading research, Koda (2005) presents the following three effective methods when the learners have encountered difficult text: (a) Obtain a supporting method which makes the reading more understandable; (b) Learn strategies for reading; and (c) Learn to monitor one's own comprehension ability. Given these limitations, it is necessary in the future to study the generation process of knowledge-based inferences when learners attempt to read more authentic and difficult texts.

Finally, this research focused exclusively on the characteristics of text information. I was not able to examine strategic processing in text comprehension via the provision of reading goals and task instructions. Some L1 studies have shown that readers generate different patterns of inferences depending on the reading goal (i.e., read to explain, predict, or understand) (Magliano et al., 1999), and that readers process and recall a text differently depending on its relevance to the perspective (given as task instructions) (Kaakinen & Hyona, 2005; McCrudden, Schraw, & Kambe, 2005). Concerning L2 reading, Horiba (2013) suggested that, even with the same text, different modes of L2 text processing and different contributions of L2 proficiency and general comprehension skill to L2 reading may occur when different task instructions are given or when students set different reading goals. Accordingly, it is necessary to look at strategic processing in EFL text comprehension in a way that also considers what kinds of reading goals EFL learners have and what kinds of tasks they are performing as a function of

reading instructions.

6.4 Pedagogical Implications and Concluding Remarks

The findings of the current study suggest the following four pedagogical implications. First, knowledge-based inferences (e.g. predictive inferences, thematic inferences), which are generated when text information and the background knowledge of the reader intersect, provide the opportunity to construct a mental representation of the text in the mind of the reader, and helps text information to be retained in long-term memory (Fincher-Kiefer, 1996). However, in L2/EFL reading comprehension, less proficient readers cannot assign their cognitive resources to their inferential process because of the insufficiency of their lower-level text processing, and they cannot control their inference generation flexibly according to the text characteristics. Study 1 found that thematic inferences from text information from a short passage were generated during reading, affected by neither the level of L2 reading proficiency nor the strength of contextual constraints. For Study 2, narrative texts were used, texts that had multiple elements such as background setting for story, the personalities of the characters, their goals, and actions. The results showed that proficient EFL readers activated predictive inferences during reading when the necessity and sufficiency of the inference activation was high due to contextual constraint manipulation, whereas less proficient EFL readers did not activate them. That is, less proficient readers may be processing the literal meaning of the text in local discourse (e.g., relations between adjacent sentences), but they are not making the necessary connections between the text and appropriate background information in terms of global coherence (e.g., relation between larger chunks of text). Hence, more cognitive resources are necessary for the higher-level text processes that integrate the multiple propositions that connect sentences and yield inferences from the cause-effect information that has been constructed.

Proficient readers, who are able to use these processes more fluently, need fewer working memory resources to derive propositional meaning during reading and hence are left with more cognitive resource for higher-level text processing and inferential processes. Therefore, the first implication is that, in order for Japanese EFL readers to allocate their cognitive resources to the inferential process, it is necessary to improve the efficiency of lower-level language processes such as word recognition, syntactic parsing, and the construction of the meaning of a text. The straightforward ways to accomplish these goals are to reduce the reader's burden in regards to word recognition by learning the meaning of unknown words that appear in the text, and also to reduce the reader's burden for lower-level text processing by using a text with a high topic familiarity or one with a simple syntactic structure. Such techniques, however, only offer one-time help to enable students to accurately comprehend a particular text, and they cannot be said to be effective at allowing long-term development of students' comprehension. In contrast, teachers directing students' attention to discourse markers (such as *therefore*, *because* and *however*), which indicate the relationship between sentences, is believed to spontaneously encourage students to predict the content of the English sentence that follows and to understand the causal structure of the text. In reading English, it is important that a reader actively constructs the meaning of the text data rather than passively receiving the meaning. Using text data as clues to predict what lies ahead and understand the overall theme of the text is an important skill for experienced readers to have. In addition to this, as one of techniques, Nassaji (2003) suggested that less proficient readers should read slowly or re-read the text so that they can have enough cognitive resources for integrating meanings with prior knowledge and constructing a coherent mental representation of the text.

The second implication involves the type of texts that should be used for EFL learners' reading comprehension. In Studies 1 and 2, it was found that raising the contextual constraints

facilitated explicit understand of the text on the part of EFL readers. This facilitative effect was found to be present regardless of differences in proficiency levels. As stated previously, constraint factors did not necessarily cause inferences to be produced in all EFL readers while reading, but during the process of reproducing text information in memory after reading, if they functioned as useful cues. Moreover, Experiment 3 of Study 2 showed that a specific type of predictive inference referred to as a motivational predictive inference, which serves as a cause or a motivation, was more strongly encoded than the simple sequence of the events describing the story. Reinforcing contextual constraints is the same thing as strengthening the causal links between text information and specific inferential content. From this, it could conclude that two effects have been achieved: (a) raising the causal coherence among texts made it easier for readers to construct coherent text representations; and (b) when inferential information has been strongly encoded into EFL readers' mental representations and that inferential information is used as a cue, it became easier to access information in memory. Thus, these findings imply that it is important for EFL classroom teachers to provide their students (especially less proficient readers) with text that is highly constrained contextually. In doing so, the causal relations of the events in the story become easier for students to understand, and it will become easier to draw thematic and predictive inferences. As a result, it will aid in the accurate understanding of the text and in the building of elaborative mental representations.

The third implication deals with the use of analogies in reading comprehension activities. Study 3 shed some light on how the reading comprehension of EFL learners can be facilitated through the use of analogical information. It was demonstrated that providing an analogical text with surface similarities as an advance organizer to EFL learners made it easier for them to activate knowledge-based inferences and to understand the learning text. Recent L1 research utilizing Latent Asemantic Analysis (LSA) has also suggested the similar fact that stronger

associations between textual information and information in background knowledge result in a greater likelihood that the inferred idea will be activated (Yeari & van den Broek, 2015). Moreover, as L2/EFL proficiency levels increase, it would likely be effective to provide learners with analogical texts that have structural similarities. For example, when having someone read a passage about *hierarchically organized memory mechanisms*, we can imagine the example of presenting that person with an analogy—either in written or graphic form—of library shelves properly arranged according to topic or category, thereby getting them to imagine the relationship between the two. For less proficient EFL learners, on the other hand, it might be effective to issue instructions or tasks designed to get them to think about the common things between two texts, thereby getting them to notice the similarities in the causal structure between the texts. It is educationally valuable to develop analogical transfer skills based on structural similarity, particularly because the process is similar to the way we use schemas in our daily lives. Analogical transfer and schemas have much in common, in that existing knowledge may be used to inform situations that are less well understood, and that both rely on structured representations to do so (Day & Gentner, 2007). Teaching and nurturing the ability to transfer analogically means cultivating strategic ways in which learners can apply their knowledge to new problems, concepts or situations that they need to understand.

However, the current research only covers narrative text and thus the impact of analogies on expository text reading is unknown, and the interpretation of the outcome is limited. Many L1 reading studies claimed that analogies aid the readers' comprehension and the process of acquiring new knowledge (Blanchette & Dunbar, 2002; Gentner & Markman, 1997; Holyoak & Koh, 1987). On the other hand, some L2 reading studies showed that analogies inhibit comprehension (Brantmeier, 2005; Hammadou, 1990, 2000). In these L2 studies, the addition of analogies to scientific texts (e.g., eye/camera analogy) did not have a positive effect on either

L1 or L2 comprehension as measured via recall, sentence completion, and multiple choice questions. Expository writing often contains text that explains new concepts and knowledge, and describes complex structures. For this reason, greater dependence on the readers' knowledge of the topics is required than with story text. In the future, the effect of analogies on expository text reading needs to be examined based on not only readers' L2 proficiency but also the readers' familiarity with topics and their amount of knowledge.

Lastly, one important key to narrative reading comprehension by EFL learners is to make inferences about the causal relationships among the goals and actions of the characters, and even among the motives behind their actions. As mentioned in the discussion about the constructionist view of reading, readers naturally attempt to explain why actions, events, and states are described in a text, and become actively engaged in the inferential process in order to construct a coherent mental representation. However, since L2/EFL learners are less likely to generate inferences by themselves, teachers need to help these immature students to generate inferences strategically. One effective technique is questioning-answering. In Experiment 3, EFL readers' strategic generation of predictive inferences was prompted by asking them, "What will happen next?" The provision of this question was effective for: (a) inducing the generation of a more specific inference, and (b) improving recall performance of motivational inference text information that had a strong connection between the protagonist's motivational goal and the outcome of the story. In other words, *what-happens-next questions* are effective for directing students' focus on causal consequences and improving their awareness toward causal coherence to construct a story by reflecting back on the stories they have read. Other than the above type of question, when used by teachers, *why-questions* can encourage less proficient readers to generate inferences. *How-questions* expose subordinate goals and actions and causal antecedent events (Graesser et al.,1994). Moreover, teachers should ask students a few

questions to confirm their understanding, for example, questions such as, “What do you think the main character felt like?”, “Why do you think that happened?” and “What do you think will happen next?” The presentation of these questions will help students to activate their general knowledge concerning the story and facilitate the generation of inferences, which allow more coherent mental representations to be established.

To date, a number of empirical studies have investigated inference generation among Japanese EFL readers (e.g., Horiba et al. 1993; Muramoto, 2000; Yoshida, 2003). However, few studies have paid attention to how text and learner variables affect knowledge-based inference generation among Japanese EFL readers and how this inference generation is related to their reading comprehension processes. The current study shed some light on the mechanism of knowledge-based L2/EFL text comprehension and the effects of text variables (i.e., the strength of contextual constraint, similarity information provided through analogical text) and L2 reading proficiency on it. The study primarily clarified (a) the condition under which EFL readers activate and encode knowledge-based inferences, (b) the significant roles of making prediction with high contextual constraint in constructing a more stable long-term text memory, and (c) the contribution of L2 reading proficiency to use knowledge of causal and structural text features in text comprehension.

Moreover, these findings are important in order to teach EFL learners how to develop a deeper understanding of text. In Japan, English teachers sometimes seem to pay too much attention to students’ superficial errors, such as spelling and grammar, or understanding the meaning of the text at the surface and textbase level. However, if we view the final goal of L2/EFL reading comprehension as the construction of a coherent mental representation, it should be more focused on the process of readers connecting the text information with their

prior knowledge. Deeper comprehension is achieved when readers infer the global message of the text, and the ability to generate inferences is the cornerstone of reading competence. Finally, the author hopes this dissertation will help in the development of inference research for L2/EFL learners, and will contribute to development in the field of English language education in Japan.

References

Albrecht, J. E., & Myers, J. L. (1995). Role of context in accessing distant information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1459-1468.

Albrecht, J. E., & Myers, J. L. (1998). Accessing distant text information during reading: Effects of contextual cues. *Discourse Processes, 26*, 87-108.

Allbritton, D. (2004). Strategic production of predictive inferences during comprehension. *Discourse Processes, 38*, 309-322.

Barnes, M. A., Dennis, M., & Haefele-Kalvatis, J. (1996). The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferencing in children from six to fifteen years of age. *Journal of Experimental Child Psychology, 61*, 216-241.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Blanchette, I., & Dunbar, K. (2002). Representation change and analogy: How analogical inferences alter target representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 672-685.

Bohn-Gettler, C. M., Rapp, D. N., van den Broek, P., Kendeou, P., & White, M. J. (2011). Adults' and children's monitoring of story events in the service of comprehension. *Memory & Cognition, 39*, 992–1011.

Braasch, J. L. G., & Goldman, R. S. (2010). The role of prior knowledge in learning from analogies in science texts. *Discourse Processes, 47*, 447-479.

Brantmeier, C. (2005). Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension in Spanish. *The Modern Language Journal, 89*, 37-53.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal*

Behavior, 11, 717-726.

Cain, K., Oakhill**,** J. V., Barnes, M. A., & Bryant**,** P. E. (2001). Comprehension skill, inference making ability and their relation to knowledge. *Memory & Cognition, 29*, 850-859.

Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. *Memory, 9*, 365-381.

Calvo, M. G., Castillo, M. D., & Estevez, A. (1999). On-line predictive inferences in reading: Processing time *during* versus *after* the priming context. *Memory & Cognition, 27*, 834-884.

Calvo, M. G., Estevez, A., & Dowens, M. G. (2003). Time course of elaborative inferences in reading as function of prior vocabulary knowledge. *Learning and Instruction, 13*, 611-631.

Calvo, Meseguer, and Carreiras (2001). Inferences about predictable events: Eye movements during reading. *Psychological Research, 65*, 158-169.

Campion, N., & Rossi, J. P. (1999). Inference and text comprehension. *L'Année Psychologique, 99*, 493–527.

Campion, N. (2004). Predictive inferences are represented as hypothetical facts. *Journal of Memory and Language, 50*, 149-164.

Carrell, P. L. (1984a). Evidence of a formal schema in second language comprehension. *Language Learning, 34*, 87-112.

Carrell, P.L., Devine, J. & Eskey, D.E. (1988). I*nteractive Approaches to Second Language Reading,* Cambridge: CUP.

Casteel, M. A. (2007). Contextual support and predictive inferences: What do readers generate and keep available for use? *Discourse Processes, 44*, 51-72.

Catrambone, R. (2002). The effect of surface and structural feature matches on the access of

story analogs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 318-334.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Collins, S., & Tajika, H. (1996). Do EFL learners make instrumental inferences when reading? Some evidence from implicit memory types. *JALT Journal (Japan Association for Language Teaching Journal), 18*, 27-39.

Cook, A. E., Limber, J. E., & O'Brien, E. J. (2001). Situation-based context and the availability of predictive inferences. *Journal of Memory and Language, 44*, 220-234.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge University Press.

Day, S. & Gentner, D. (2007). Nonintentional analogical inference in text comprehension. *Memory & Cognition, 35*, 39-49.

Duffy, S. A. (1986). Role of expectations in sentence integration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 208-219.

Duffy, S. A., Henderson, J. M, & Morris, R. K. (2003). Semantic facilitation of lexical access during sentence processing. *Psychological Science, 26*, 289- 291.

Dyer, M. G. (1983). *In-Depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension.* Cambridge, MA, MIT Press.

Educational Testing Service [ETS]. (2016). *Mapping the TOEIC® Listening and Reading onto the CEFR (PDF)* Retrieved from https://www.etsglobal.org/Global/Eng/content/download/768/12037/version/8/file/TOEIC+L%26R+Descriptors-MAR089-LR.pdf

Fillmore, C. J. (1976). Frame semantics and the nature of language. *In Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech, 280*, 20–32

Fincher-Kiefer, R. (1993). The role of predictive inferences in situation model construction. *Discourse Processes, 16*, 99-124.

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

Fincher-Kiefer, R. (1996). Encoding differences between bridging and predictive inferences *Discourse Processes, 22*, 225-246.

Forbus. K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19*, 141-205.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95*, 393-408.

Gentner, D., Ratterman, M.J., & Forbus, K. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology, 25*, 524-575.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 45-56.

Gerrig, R., & O'Brien, E. (2005). The scope of memory based processing. *Discourse Processes, 39*, 225–242.

Grabe, W. (2000). Reading research and its implications for reading assessment. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment*. Cambridge University Press.

Grabe, W., & Stoller, F. (2002). *Teaching and researching reading*. Harlow: Longman.

Graesser, A.C. & Clark, L.F. (1985). *Structures and procedures of implicit knowledge.* Norwood, NJ: Ablex.

Graesser, A. C., & Kreuz, R. J. (1993). A theory of inference generation during text comprehension. *Discourse Processes, 16*, 146-160.

Graesser, A. C., Millis, K. K., & Zwaan. R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48*, 163-189.

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371-395.

Graesser, A. C., & Wiemer-Hastings, K. (1999). Situation models and concepts in story comprehension. In S. R. Goldman, A. C. Graesser & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essay in honor of Tom Trabasso.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hammadou, J. (2000). The impact of analogy and content knowledge on reading comprehension: What helps, what hurts. *The Modern Language Journal, 84*, 38-50.

Hano, Y. & Chihara, T. (1996) Anarogi katei ni okeru ninchi konponento [Cognitive Components in Analogical Reasoning Process]. *The Faculty of Education Shiga University, 46*, 203-220.

Hannon, B., & Daneman, M. (1998). Facilitating knowledge-based inferences in less- skilled readers. *Contemporary Educational Psychology, 23*, 149-172.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15*, 332-340.

Horiba, Y. (1993). The role of causal reasoning and language competence in narrative comprehension. *Studies in Second Language Acquisition, 15*, 459-472.

Horiba, Y. (1996a). Comprehension processes in L2 reading: Language competence, textual

coherence, and inferences. *Studies in Second Language Acquisition, 18*, 433-473.

Horiba, Y. (1996b). The role of elaborations in L2 comprehension: The effect of encoding task on recall. *The Modern Language Journal, 80*, 151–164.

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

Horiba, Y. (2013). Task-induced strategic processing in L2 text comprehension. *Reading in a Foreign Language, 25*, 98-125.

Horiba, Y., & Fukaya, K. (2015). Reading and learning from L2 text: Effects of reading goal, topic familiarity, and language proficiency. *Reading in a Foreign Language, 27*, 22–46.

Horiba, Y., van den Broek, P. W., & Fletcher, C. R. (1993). Second language readers' memory for narrative texts: Evidence for structure-preserving top-down processing. *Language Learning, 43,* 345-372.

Ikeno, O. (1996). The effects of text-structure-guiding questions on comprehension of texts with varying linguistic difficulties. *Japan Association of College English Teachers (JACET) Bulletin, 27*, 51-68.

Iseki, R., & Kaiho, H. (2002). Is this inference on-line?: Methodological and conceptual considerations of on-line inferences in discourse comprehension. *Tsukuba Psychological Research, 24*, 83-97.

Iseki, R. (2003). Text rikai ni okeru on-line shori mechanism: Joukyou model kouchiku katei ni kansuru rironteki gaikan [On-line processing mechanisms in text comprehension: A theoretical review on constructing situation models]. *The Japanese Journal of Psychology, 75*, 442-458.

Japan Association of College English Teachers (JACET) Basic words revision committee (Ed.). (2003). *JACET List of 8000 Basic Words*. Tokyo: JACET.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122-149.

Kawasaki, E. (2005). Bunsyourikai to kioku no moderu [The model of reading comprehension and memory] kotoba no jikkenshitsu (pp. 133-161). Tokyo: Brain Shuppan.

Keane, M. (1987). On retrieving analogues when solving problems. *Journal of Experimental Psychology, 39*, 29-41.

Keefe, D. E., & McDaniel, M. A. (1993). The time course and durability of predictive inferences. *Journal of Memory and Language, 32*, 446-463.

Keenan, J. M., Potts, G. R., Golding, J. M., & Jennings, T. (1990). Which elaborative inferences are drawn during reading?: A question of methodologies. In D. A. Balota, G. B. Flores d'Arcais, & K. Raynor (Eds.), *Comprehension processes in reading* (pp. 377-403). Hillsdale, NJ: Erlbaum.

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

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*, 294-303.

Kintsch, W. & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.

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

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

Kim, C. M., & Myers, J. L. (1993). Reinstatement of causal information during reading. *Journal*

of Experimental Psychology: Learning, Memory, and Cognition, 19, 554-560.

Kaakinen, J. K., & Hyönä, J. (2005). Perspective effects on expository text comprehension: Evidence from think-aloud protocols, eyetracking, and recall. *Discourse Processes, 40*, 239-257.

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

Langston, M. C., & Trabasso, T. (1999). Modeling causal integration and availability of information during comprehension of narrative texts. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 29–69). Mahwah, NJ: Erlbaum.

Lassonde, K. A., & O'Brien, E. J. (2009) Contextual specificity in the activation of predictive inferences. *Discourse Processes, 46,* 426-438.

Lehman, M.T., & Tompkins, C.A. (2000). Inferencing in adults with right hemisphere brain damage: An analysis of conflicting results. *Aphasiology, 14*, 485-499.

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. *Discourse Processes, 34*, 259-280.

Long, D.L., & Golding, J.M. (1993). Superordinate goal inferences: Are they automatically generated during comprehension? *Discourse Processes, 16*, 55–73.

Long, D. L., Golding, J. M., & Graesser, A. C. (1992). A test of the on-line status of goal-related inferences. *Journal of Memory and Language, 31*, 634-647.

Long, D. L., & Lea, R. B. (2005). Have we been searching for meaning in all the wrong places? Defining the "search after meaning" principle in comprehension. *Discourse Processes, 39,* 279–298.

Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual differences in the time course of

inferential processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1456-1470.

Long, D. L., Oppy, B. J., & Seely, M. R. (1997). Individual differences in sentence-level and inferential processing. *Journal of Memory and Language, 36*, 129-145.

Lorch, R. F., & O'Brien, E. J. (1995). *Sources of coherence in reading*. Hillsdale, NJ: Lawrence Erlbaum.

Magliano, J. P., Baggett, W. B., Johnson, B. K., & Graesser, A. C. (1993). The time course of generating causal antecedent and causal consequence inferences. *Discourse Processes, 16*, 35-53.

Magliano, J. P., Zwaan, R. A., & Graesser, A. (1999). The role of situational continuity in narrative understanding. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 219-245). Mahwah, NJ: Lawrence Erlbaum Associates.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 82-91.

McKoon, G. & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*, 440-466.

McCrudden, M. T., Schraw, G., & Kambe, G. (2005). The effect of relevance instructions on reading time and learning. *Journal of Educational Psychology, 97*, 88-102.

McNamara, D. S. (1995). Effects of prior knowledge on the generation advantage: Calculators versus calculation to learn simple multiplication. *Journal of Educational Psychology, 87*, 307-318.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*, 1-30.

McNamara, D.S., & Kintsch, W. (1996). Learning from texts: Effect of prior knowledge and text coherence. *Discourse Processes, 22*, 247-288.

McNamara, D. S., O'Reilly, T., Best, R., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research, 34*, 147-171.

Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language, 26*, 453-465.

Minaminosono, H. (1997). Dokkai sutoratejii no shiyoo to dokkairyoku to no kankei ni kansuru choosakenkyuu: gaikokugo to shite no nihongo tekisuto dokkai no baai. [A study on the relationship between reading strategy and reading comprehension ability: Comprehending Japanese texts as a foreign language] *Sekai no Nihongo Kyooiku 7,* 31-44.

Muramoto, T. (2000). Dai-ni gengo no bunshou rikai katei ni oyobosu shujukudo no eikyou: Bun sainin kadai niyoru jikken [The effects of second-language proficiency on text comprehension]. *The Science of Reading, 44*, 43-50.

Murray, J. D., & Burke, K. A. (2003). Activation and encoding of predictive inferences: The role of reading skill. *Discourse Processes, 35*, 81-102.

Murray, J. D., Klin, C. M., & Myers, J. L. (1993). Forward inferences in narrative text. *Journal of Memory and Language, 32*, 464-473.

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes, 26*, 131-157.

Nassaji, H. (2003). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *Modern Language Journal, 87*, 261-276.

Negishi, M. (2011). CEFR-J Kaihatsu no Keii [The Development Process of the CEFR-J].

Action Research Center for Language Education (ARCLE) Review, 5, 37-52.

Noordman, L. G. M., & Vonk, W. (2015). Inferences in discourse, psychology of. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences (2nd ed.), 12*, 37-44. Amsterdam: Elsevier.

Oakhill, J. V. (1993). Children's difficulties in reading comprehension. *Educational Psychology, 5*, 223-237.

Oakhill, J. V., Yuill, N. M., and Parkin, A.J. (1986). On the nature of the difference between skilled and less-skilled comprehenders. *Journal of Research in Reading, 9*, 80–91.

O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 777–784.

O'Brien, E. J., Cook, A. E., & Peracchi, K. A. (2004). Updating situation models: Reply to Zwaan and Madden (2004). J*ournal of Experimental Psychology: Learning, Memory, and Cognition,* 30, 289-291.

O'Brien, E. J., & Myers, J. L. (1999). Text comprehension: A view from the bottom up. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.) *Narrative comprehension, causality, and coherence: Essays In honor of Tom Trabasso* (pp. 35-53). Mahwah, NJ: Lawrence Etlbaum Associates, Inc.

O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1200-1210.

O'Brien, E. J., Shank, D. M., Myers, J. L., & Rayner, K. (1988). Elaborative inferences during reading: do they occur on-line? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 410-420.

O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. *American Educational Research Journal, 44*, 161-196.

Osaka, M., & Osaka, N. (1992). Language independent working memory as measured by Japanese and English reading span test. *Bulletin of the Psychonomic Society, 30*, 287-289.

Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & Mcnamara, D. S. (2007). Influence of Question Format and Text Availability on the Assessment of Expository Text Comprehension. *Cognition and Instruction, 25*, 399-438.

Peracchi, K. A., & O'Brien, E. J. (2004). Character profiles and the activation of predictive inferences. *Memory & Cognition, 32*, 1044-1052.

Perfetti, C. A. (1997). Sentences, individual difference, and multiple texts: Three issues in text comprehension. *Discourse Processes, 23*, 337-355.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language, 27*, 399-415.

Rizzella, M. L., & O'Brien, E. J. (1996). Accessing global causes during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1208-1218.

Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sanford, A.J., & Garrod, S. C. (2005). Memory-based approaches and beyond. *Discourse Processes, 39*, 205-224.

Sanford, A. J., & Graesser, A. C. (2006). Shallow processing and under specification. *Discourse Processes, 42*, 99-108.

Seifert, C. M., McKoon, G., Abelson, R. P., & Ratcliff, R. (1986). Memory connections between thematically similar episodes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 220-231.

Schank, R.C. (1986). *Explanation Patterns: Understanding Mechanically and Creatively.* Hillsdale , NJ: Erlbaum.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Shears, C., & Chiarello, C. (2004). Knowledge-Based Inferences Are Not General. *Discourse Processes, 38,* 31-55.

Shimizu, H. (2011). *Generation of Japanese EFL Readers' Bridging Inferences: The Effects of Text Characteristics and L2 Reading Proficiency.* (Doctoral dissertation).

Shimizu, M. (2006). *Inference Generation Processes of Japanese EFL Learners: Effects of Questioning on Their Reading Comprehension.* (Doctoral dissertation). Retrieved from https://tsukuba.repo.nii.ac.jp/index.php?active_action=repository_view_main_item_detail&page_id=13&block_id=83&item_id=20706&item_no=1

Singer, M., & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Learning and Verbal Behavior, 22*, 437-448.

Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992). Validation of causal bridging inferences in discourse understanding. *Journal of Memory and Language, 31*, 507-524.

Society for Testing English Proficiency [STEP]. (2007). *The EIKEN Test in Practical English Proficiency Pre Grade 1*. Tokyo: Obunsha.

Society for Testing English Proficiency [STEP]. (2007). *The EIKEN Test in Practical English Proficiency Grade 2*. Tokyo: Obunsha.

Society for Testing English Proficiency [STEP]. (2007). *The EIKEN Test in Practical English Proficiency Pre-Grade 2*. Tokyo: Obunsha.

Society for Testing English Proficiency [STEP]. (2008).*The EIKEN can-do list.* Tokyo, Japan: The Society for Testing English Proficiency.

Taniguchi, A. (1988). Effect of inserted analogies on the retention of a text. *Japanese Association of Educational Psychology, 36*, 282-286.

Taniguchi, A. (1999). Tekisuto no rikai to kioku wo sokushin suru gutaikajyouhou. [The concrete information facilitating text comprehension and memorization] Kazamashobou.

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

Trabasso, T., & Magliano, J. P. (1996). Conscious understanding during comprehension. *Discourse Processes, 21*, 255-287.

TUFS Tonolab. (2013). CEFR based framework for ELT in Japan. Retrieved from www.tufs.ac.jp/ts/personal/tonolab/cefr-j

van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P. (2005) Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443–467.

van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science, 328*, 453456.

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

van den Broek, P., & Gustafson, M. (1999). Comprehension and memory for texts: three generations of reading research. In Goldman, S. R., Graesser, A. C., van den Broek, P

(Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (p. 17). Mahwah, NJ; Lawrence Erlbaum.

van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition, 29*, 1081-1087.

van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes, 39,* 299–316.

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the on-line construction of a memory representation. In H. van Oostendorp & S. R. Goldman (Eds.),The construction of mental representations during Reading (pp. 71–98). Mahwah, NJ: Erlbaum.

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

Wharton, C. M., Holyoak, K. J., & Lange, T. E. (1996). Remote analogical reminding. *Memory and Cognition, 24*, 629–643.

Wittwer, J., & Ihme, N. (2014). Reading skill moderates the impact of semantic similarity and causal specificity on the coherence of explanations. *Discourse Processes, 51*, 143–166.

Yeari, M. P. & van den Broek. (2015).The role of textual semantic constraints in knowledge-based inference generation during reading comprehension: A computational approach. *Memory, 23*, 1193-1214,

Yoshida, M. (2003). Working memory capacity and the use of inference in L2 reading. *JACET Bulletin, 36*, 1-17.

Zhang, H., & Hoosain, R. (2005). Activation of themes during narrative reading. *Discourse*

Processes, 40, 57-82.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*, 292-297.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 386-397.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162-185.

Appendices

Appendix 1

Experimental Passages and Thematic Inference Targets Used in Experiment 1

No.	Passages	Thematic Target		Syllables	Frequ-ency	Familiarity	Contextual Constraint
		Appropriate	Inappropriate				
No. 1-7: High Contextual Constraint condition							
1	The townspeople were amazed to find that all the buildings had collapsed except the bank. Obviously, it had been built to get ready for natural disasters.	earthquake	stranger	2	3	6.6	6.8
2	The waitress smiled and said she would get off work soon. The young man decided to wait outside the restaurant.	date	wind	1	1	6.2	6.8
3	The millionaire jumped from the window when he heard about the new money rates. His entire fortune was at stake.	suicide	wisdom	2	3	6.2	6.6
4	The big moment came and the boy was very excited. He blew out the candles and then bit into more cake than he could possibly eat.	birthday	breakfast	2	2	6.4	6.6
5	The parents helped the little boy toss food through the fence in the direction of monkeys. Then they all walked on to see the other animals.	zoo	dust	1	3	6.8	5.6
6	Thinking of the amount of garlic in his dinner, the guest asked for a mint gum. He soon felt more comfortable socializing with the others.	breath	weather	1	2	6.2	5.4
7	For the third time, the worried baseball player swung but missed the ball. He knew what the coach would say.	out	car	1	1	5.4	5.4

No.	Passages	Thematic Target Word		Syllables	Frequ-ency	Familiarity	Contextual Constraint
		Appropriate	Inappropriate				
No. 8-14: Low Contextual Constraint condition							
8	The chemistry student knew that this was not a good time to forget how to calculate volume. Again, she tried to recall the formulas.	test	meal	1	1	7	4.6
9	Round after round, the player tried to find his opponent's weakness. When the fencing instructor brew his whistle, the player reluctantly lowered his sword.	defeat	affair	2	2	4.4	4.4
10	The jockey was happy to receive the trophy. However, he said the secret was his use of a new kind of harness.	race	role	1	1	5.8	4.4
11	The little girl was very happy with the new doll from her grandmother. She reached up to hug her and give her little kiss.	love	rise	1	1	6.2	4
12	The worker was struck by a gigantic, falling steel ball. Everyone was seriously upset by the accident.	dead	year	1	1	6	3.8
13	The bird loved to compete with the others. But no matter how hard he worked, he could not keep up with the swallow.	flying	moment	2	1	5.6	3.2
14	Danny had a lot to learn about riding bicycles. Going too fast and looking the other way, he crashed into the wall.	pain	sleep	1	1	6	3.2
No. 15-21: Filler condition							
15	The scribbling on the paper was hard to read but was apparently of great importance. It was carried to the King by a young knight.	test	sweim (nonword)	2	1	6.8	5.4

No.	Passages	Thematic Target Word		Syllables	Frequ-ency	Familiarity	Contextual Constraint
		Appropriate	Inappropriate				
16	The doctor became very nervous as he watched the patient's eye. He had seen this kind of problem only once or twice before.	breath	blorlt (nonword)	1	2	6.4	5.2
17	All afternoon the chef monitored the food preparations for the party. He tasted the soup, then added a little bit.	race	geuked (nonword)	1	2	6.6	5
18	When the maid turned away from the laundry, the baby grabbed the iron. Later that day, the maid started looking for a new job.	love	fribbs (nonword)	1	2	6	5
19	The waiter left quickly as he saw the angry customer tear up the bill. He did not want to risk getting in a fight.	dead	volms (nonword)	1	1	6.4	5
20	The teammates heard the loud crack. That was the last time anyone would be able to use that bat.	flying	scincs (nonword)	2	2	5.2	4.8
21	After desperately holding the rope for hours, the climber felt his arms begin to go weak. Still he had hopes that he would be rescued.	pain	rhinde (nonword)	1	1	6.6	4.8

Note. Word frequency level was calculated by Jacet 8000 Word Level Checker; Nonwords were presented in filler condition No. 15~21 passages for lexical decision task.

Appendix 2

Answer Sheets for Appropriateness Rating Task Used in Experiment 1

実験回答用紙　Part 1　　氏名＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

■ 以下に先ほど PC 上で読んだのと同じ文章のうちのいくつかが書かれています。英単語 1 語が挙げられているので、これが文章内容から連想される語として、どの程度適切だと感じるか 7 段階で評価してください。

(1)	The townspeople were amazed to find that all the buildings had collapsed except the bank. Obviously, it had been built to ready for natural disasters.
earthquake	1　2　3　4　5　6　7 ←全く適切でない，あまり，　普通，　まあまあ，　とても適切である→
(2)	The waitress smiled and said she would get off work soon. The young man decided to wait outside the restaurant.
date	1　2　3　4　5　6　7 ←全く適切でない，あまり，　普通，　まあまあ，　とても適切である→
(3)	The millionaire jumped from the window when he heard about the new money rates. His entire fortune was at stake.
suicide	1　2　3　4　5　6　7 ←全く適切でない，あまり，　普通，　まあまあ，　とても適切である→
(4)	The big moment came and the boy was very excited. He blew out the candles and then bit into more cake than he could possibly eat.
birthday	1　2　3　4　5　6　7 ←全く適切でない，あまり，　普通，　まあまあ，　とても適切である→
(5)	The parents helped the little boy toss food through the fence in the direction of monkeys. Then they all walked on to see the other animals.
dust	1　2　3　4　5　6　7 ←全く適切でない，あまり，　普通，　まあまあ，　とても適切である→
(6)	For the third time, the worried baseball player swung but missed the ball. He knew what the coach would say.
car	1　2　3　4　5　6　7 ←全く適切でない，あまり，　普通，　まあまあ，　とても適切である→
(7)	Round after round, the player tried to find his opponent's weakness. When the fencing instructor brew his whistle, the player reluctantly lowered his sword.
affair	1　2　3　4　5　6　7 ←全く適切でない，あまり，　普通，　まあまあ，　とても適切である→
(8)	The chemistry student knew that this was not a good time to forget how to calculate volume. Again, she tried to recall the formulas.
test	1　2　3　4　5　6　7 ←全く適切でない，あまり，　普通，　まあまあ，　とても適切である→

(9)	Thinking of the amount of garlic in his dinner, the guest asked for a mint gum. He soon felt more comfortable socializing with the others.
breath	1 2 3 4 5 6 7 ←全く適切でない，あまり， 普通， まあまあ， とても適切である→
(10)	The jockey was happy to receive the trophy. However, he said the secret was his use of a new kind of harness.
race	1 2 3 4 5 6 7 ←全く適切でない，あまり， 普通， まあまあ， とても適切である→
(11)	The little girl was very happy with the new doll from her grandmother. She reached up to hug her and give her little kiss.
rise	1 2 3 4 5 6 7 ←全く適切でない，あまり， 普通， まあまあ， とても適切である→
(12)	The worker was struck by a gigantic, falling steel ball. Everyone was seriously upset by the accident.
year	1 2 3 4 5 6 7 ←全く適切でない，あまり， 普通， まあまあ， とても適切である→
(13)	The bird loved to compete with the others. But no matter how hard he worked, he could not keep up with the swallow.
moment	1 2 3 4 5 6 7 ←全く適切でない，あまり， 普通， まあまあ， とても適切である→
(14)	Danny had a lot to learn about riding bicycles. Going too fast and looking the other way, he crashed into the wall.
sleep	1 2 3 4 5 6 7 ←全く適切でない，あまり， 普通， まあまあ， とても適切である→

実験回答用紙　Part 2　　氏名＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

■　キーワード語を手がかりに、各文章について思い出せる限り全ての内容を、できるだけ本文に忠実なかたちで、日本語ですべて書き出してください。

(1)	
earthquake	
(2)	
date	
(3)	
suicide	
(4)	
birthday	
(5)	
dust	
(6)	
car	
(7)	
affair	
(8)	
test	

(9)	
breath	
(10)	
race	
(11)	
rise	
(12)	
year	
(13)	
moment	
(14)	
sleep	

Appendix 3 Materials Used in Experiment 2 and 3

Text A : Jimmy's story

Introduction
Jimmy was the new kid on the block. Although his parents urged him to go meet the other kids in the neighborhood, he was shy and hadn't made any new friends. One Saturday morning, his mom asked him to go to the store for her. While he was walking back home, Jimmy ran into some of the kids from the neighborhood. They asked him if he wanted to play with them. Jimmy was delighted and ran across the street to play with them. They showed their family's brand new car that didn't have any scratches.
High contextual constraint version
Neighbor kids said his father was so proud of his new car. They had to be careful when playing near it. That was because the body of the car was made of very soft metal. They knew it would be damaged without much pressure. They taught him a fun game that involved throwing *rocks* at a target to get points. Jimmy and his friends were having a great time. Jimmy even won the game once or twice. He stepped up to take his turn and aimed at the target.
Low contextual constraint version
They had to be careful when playing near it. Then a man who lived in the neighborhood came by and blamed children. That was because he was angry at the fact that his window glass had been broken by other children just 3 days ago. They taught him a fun game that involved throwing *sponge ball* at a target to get points. Jimmy and his friends were having a great time. Jimmy even won the game once or twice. He stepped up to take his turn and aimed at the target.
Inference-evoking sentence
However he missed and he accidentally hit the door of a new car.

Probe sentences for text A

(a) Explicit probe [His mom asked him to go to the store for her.]

(b) Inference probe [He accidentally damaged the door of a new car.]

(c) Control probe [Jimmy hit the target each time during the game.]

Text B: Brad's story

Introduction
Brad and his wife had been married for 20 years. They had met at their high school and had gotten married when they were 21. Although they have no children, they are happy and understand each other well. He had worked very hard for his wife. But he couldn't make enough money to lead good lives. One day, Brad was wandering through a department store because he was looking for a present for his wife's birthday. He wanted to find something special for her.
High contextual constraint version
He had been fired from his job three months ago. He wasn't sure if he could buy anything nice. In the jewelry department, he saw a beautiful ruby ring sitting in a display at the counter. However there was no way he could pay for it.
Low contextual constraint version
He had just started a new job but had not received his first salary yet. He wasn't sure if he could buy anything nice. In the jewelry department, he saw a beautiful ruby ring sitting in the display on the counter. However, there was no way he could pay for it. When Brad look at the next showcase, there was a relatively low-priced accessory displayed.
Inference-evoking sentence
Seeing no salesman and customers around, he quietly made his way closer to the counter.

Probe sentences for text B

(a) Explicit probe [He wanted to find something special for her.]

(b) Inference probe [Brad stolen the beautiful ruby ring.]

(c) Control probe [He promised to buy it one day.]

Appendix 4

(1) Materials Used in Experiment 4 [Version A]

Target	Karla, an old hawk, lived at the top of a tall oak tree. One afternoon, she saw a hunter on the ground with a bow and some crude arrows that had no feathers. The hunter took aim and shot at the hawk but missed. Karla knew the hunter wanted her feathers so she glided down to the hunter and offered to give him a few. The hunter was so grateful to Karla. (He promised never to shoot at hawk again.) He went off and shot deer instead.
Surface Strong	Once there was an eagle named Salam who nested on a rocky cliff. One day she saw a sportsman coming with a crossbow and some arrows that had no feathers. The sportsman attacked her but the arrow missed. Salam realized that the sportsman wanted her tail-feathers so she flew down and provided a few of her tail-feathers to the sportsman. The sportsman was pleased. He promised never to attack eagles again. (No. of matches: Entity 3, FOR 6, HOR 1)
Surface Weak	Once there was an eagle named Salam. She provided a few of her tail-feathers to a sportsman. So he promised never to attack eagles. One day Salam was nesting high on a rocky cliff when she saw the sportsman coming with a crossbow. Salam flew down to meet the man, but he aimed and shot her with a single arrow. Falling down to the ground, Salam realized that the arrow had her own tail-feathers on it. (No. of matches: Entity 3, FOR 3, HOR 0)
Structural Strong	Once there was a small country called Salam. It developed the world's smartest computer. One day Salam was attacked by its war-like neighbor country, Bolon. But the missiles were badly aimed and the attack failed. The Salam government realized that Bolon wanted smart computers. So they offered to sell some of their computers to the country. The government of Bolon was very pleased. Bolon promised never to attack Salam again. (No. of matches: Entity 0, FOR 5, HOR 1)
Structural Weak	Once there was a small country called Salam. It developed the world's smartest computer. Salam sold one of its supercomputers to the neighbor country, Bolon. So Bolon promised never to attack Salam. But one day, Salam was overwhelmed by a surprise attack from Bolon. Upon surrendering, the defeated people of Salam realized that the attackers' missiles had been guided by Salam's supercomputers. (No. of matches: Entity 0, FOR 1, HOR 0)

Note. The underlined sentence in the target was actually omitted in the experiment and presented as probe statements. Surface/ Structural refers to the similarity types; Strong/ Weak refers to the causality conditions.

Recognition Test Items

(a) Explicit Statement	"Karla knew the hunter wanted her feathers."
(b) Inference Statement	"The hunter promised never to shoot at the hawk again."
(c) Control Statement	"The hunter tried to capture a live brown bear."

(2) Materials Used in Experiment 4 [Version B]

Target	Percy, a skylark, spent the whole warm season singing and flying. When it began to get colder Percy visited a squirrel and sang a song for her, expecting to get some of the squirrel's sunflower seeds in return. However, the squirrel was very disappointed with him. "You are a terrible singer!" she yelled. "Your song is not worth what I would pay for my seeds! (Percy could not get any seeds from the squirrel.) Tears rolled down Percy's cheek, and he vowed to give up singing for good.
Surface Strong	A magpie named Sam sang all summer. When winter came, he visited a mouse and performed a ballad for her, hoping she would give him some nuts in return. However, the mouse was not at all pleased." Your song was terrible, you don't deserve any of my nuts." she exclaimed. Sam could not get any nuts from the mouse. He was so disappointed. (No. of matches: Entity 5, FOR 6, HOR 1)
Surface Weak	A magpie named Sam sang all summer. When winter came, he visited a mouse he often visited, hoping she would give him some nuts. The mouse was always glad to meet him and had always given some nuts to him. However, this year she said, "You have wasted the summer while I have been working hard. I will not give you any of my nuts anymore." Sam could not get any nuts from the mouse. He was so disappointed. (No. of matches: Entity 5, FOR 3, HOR 0)
Structural Strong	Sam travelled all over the world buying beautiful things. When he ran out of money he visited his mother and gave her a gift he bought while he was in Tibet, hoping she would lend him some money in return. However, his mother was not at all pleased. "This is a piece of junk! You don't deserve any of my money." she exclaimed. Sam could not get any money from his mother. He was so disappointed. (No. of matches: Entity 0, FOR 6, HOR 1)
Structural Weak	Sam travelled all over the world buying beautiful things. When he ran out of money he visited his mother. She was pleased to meet him and always lent him money when he asked. However, one day she said "While I have been working hard, you have been wasting your time. I will not give you any of my hard-earned money. Against his hope, Sam could not get any money from his mother. He had no other choice but to find a job and earn by himself. (No. of matches: Entity 0, FOR 3, HOR 0)

Note. The underlined sentence in the target was actually omitted in the experiment and presented as probe statements. Surface/ Structural refers to the similarity types; Strong/ Weak refers to the causality conditions.

Recognition Test Items

(a) Explicit Statement	"The squirrel was very disappointed with him."
(b) Inference Statement	"Percy could not get any seeds from the squirrel."
(c) Control Statement	"Percy's song was so noisy that it woke a nearby squirrel."

(3) Materials Used in Experiment 4 [Version C]

Target	Mr. Johnson had been found a cancer in his stomach and it had to be treated. The doctor had an available kind of ray that could be used to destroy the cancer. If the ray reached the cancer at a sufficiently high intensity, the cancer would be destroyed. However, at this intensity the healthy parts of the body would be badly hurt if the ray passed through on the way to the cancer. The doctor decided to use multiple ray machines to send low-intensity rays to the cancer simultaneously from different directions. (The combined rays could destroy the cancer without hurting the surrounding body parts.) In this way, the doctor succeeded the operation of removing Mr. Johnson's cancer.
Surface Strong	The dentist wanted to extract a bad tooth from his patient. He was going to use an ultrasound device that would make the tooth to break up so that it could be painlessly removed. However, if the ultrasound was used at a high level, it would damage the gums near the tooth. The dentist chose to use several ultrasound devices, each at a low-level. The ultrasound rays were shot at the tooth from several directions at the same time. The combined ultrasound rays could destroy the bad tooth without hurting the surrounding gums. (No. of matches: Entity 6, FOR 5, HOR 1)
Surface Weak	The dentist wanted to protect his patient's tooth while he did some other needed treatment of the gums near the tooth. He was going to use a device that would send out an ultrasound ray that would temporarily strengthen the tooth. The dentist used several ultrasound devices at a reduced setting and directed the ultrasound rays at the tooth at the same time. The ultrasound rays hitting the tooth would be sufficient to strengthen it without affecting the gums. (No. of matches: Entity 4, FOR 2, HOR 0)
Structural Strong	The leader of firefighters rushed to the building where a fire had broken out. He ordered his men to spray a large amount of water from one window so as to stop the fire at once. However, the owner of the building complained that the window frame would be broken if the water were to be sprayed so strongly from a single window. Therefore, the leader decided to surround the building with several fire engines and spray water from different directions gradually. Thus, a sufficient amount of water sprayed from different directions put out the fire instantly without breaking any window frame of the building. (No. of matches: Entity 0, FOR 4, HOR 1)
Structural Weak	The leader of firefighters rushed to the building where a fire had broken out. He ordered his men to spray a large amount of water from one window so as to stop the fire at once. However, the owner of the building complained that the window frame would be broken if the water were to be sprayed so strongly from a single window. Nevertheless, the leader judged it was important to stop the fire as quickly as possible, so a large amount of water was sprayed from a single window. Fortunately, the fire was put out without breaking any window frame of the building. (No. of matches: Entity 0, FOR 2, HOR 0)

Note. The underlined sentence in the target was actually omitted in the experiment and presented as probe statements. Surface/ Structural refers to the similarity types; Strong/ Weak refers to the causality conditions.

Recognition Test Items

(a) Explicit Statement	"The doctor had an available kind of ray that could be used to destroy the cancer."
(b) Inference Statement	"The combined rays could destroy the cancer without hurting the healthy parts of the body."
(c) Control Statement	"The doctor was interested in an article about a new medical technology."

(4) Materials Used in Experiment 4 [Version D]

Target	King Otto was fascinated with wars to be a powerful nation. One year, he decided to start rebuilding all the bridges of the kingdom. Otto's counselor, the minister complained that he was spending excessive money on the bridges and not enough on actual military activities. (King Otto was convinced that the counselor was right and he abandoned the bridge project to plan an invasion of a neighboring country.) Then, one day when King Otto was travelling through his kingdom, his carriage crashed through an incomplete bridge. He was severely injured, but he learned the need to balance his military and domestic expenditures.
Surface Strong	Cornelius was an astronomer who worked for the royal family as a war forecaster. He was always thinking about stars and galaxies; gradually he was remodeling his home into an observatory. But before he was finished, his advisor objected that he was using too much time renovating the house and an insufficient amount of time on his job as a royal astronomer. Cornelius figured that his advisor was correct and he went back to thinking about nothing but astronomy. One evening when Cornelius was walking through his house, a temporary walkway collapsed under his weight. The accident showed him that he should try to better coordinate his time between astronomy and the maintenance of his home. (No. of matches: Entity 5, FOR 6, HOR 1)
Surface Weak	Cornelius, an astronomer, worked for the royal family as a war forecaster. He was always thinking about stars and galaxies. Gradually he was remodeling his home into an observatory. But before he finished it, his advisor said that he was using too much time renovating the house and an insufficient amount of time on his job as a royal astronomer. Cornelius figured that his advisor was correct and reduced the time he spent on remodeling his house. Afterwards, he dedicated himself to his astronomic profession, which made his country a powerful nation. (No. of matches: Entity 4, FOR 4, HOR 0)
Structural Strong	A painter named Cornelius was working for a wealthy man as an exclusive painter. One day, he fell in love with a beautiful woman and became obsessed with painting her. He met her every day, and she also was pleased with his painting. However, the secretary of the wealthy man told Cornelius that his master will be angry if the ordered painting was not finished. According to his advice, Cornelius started to paint the picture for the wealthy man again. One day, he forgot the promise to meet the woman he was in love with because of his work. She got angry at him for that, but Cornelius learned the need to balance his love affair and work. (No. of matches: Entity 0, FOR 4, HOR 1)
Structural Weak	A painter named Cornelius was working for a wealthy man as an exclusive painter. One day, he fell in love with a beautiful woman and became obsessed with painting her. He met her every day, and she also was pleased with his painting. However, the secretary of the wealthy man told Cornelius that his master will be angry if the ordered painting was not finished. According to his advice, Cornelius started to paint the picture for the

	wealthy man again. He earned a reputation as a great painter, and the portrait of the woman later became famous as his masterpiece.
	(No. of matches: Entity 0, FOR 3, HOR 0)

Note. The underlined sentence in the target was actually omitted in the experiment and presented as probe statements. Surface/ Structural refers to the similarity types; Strong/ Weak refers to the causality conditions.

Recognition Test Items

(a) Explicit Statement	"The minister complained that he was spending excessive money on the bridges."
(b) Inference Statement	"King Otto was convinced that the minister's opinion was right."
(c) Control Statement	"People loved King Otto because he enriched their lives."

Appendix 5

(1) Target and Cue Passages Used in Experiment 5 [Set A]

Same-Theme Target	*Different-Theme Target*
[Betray of Deal]	[Making a deal to avoid a bad situation]
Once there was an eagle named Karla. She saw a hunter armed with bows and arrows coming after her. Karla noticed that some of the arrows had no feathers. She thought that she might be able to make a deal. Karla flew down and donated a few of her tail feathers to the hunter. The hunter promised never to attack eagles again. However, when she was nesting on a cliff, the arrows shot her with her tail feathers on them.	One day, lying on a big rock, was a snake called Elrod. He saw a farmer coming with a blow gun tracking him. The farmer shot at him. Elrod realized that the darts had no poison so he decided that he might be able to make a deal. He crawled over and gave some poison to the farmer so that the darts now had his poison on them. The farmer agreed never to kill snakes again, so Elrod spent the rest of his life safely.

Complete Cue

Salam was a small country. One day, Salam realized that its warlike neighbor, Bolon was arming against them. The Salamians realized that all of the Bolonian missiles were badly made and decided that an alliance could be formed. The Salam ambassador gave useful missile components to the Bolonian army. Bolonians promised that they would never fight against Salam from now on. However, Bolon declared war against Salam, and attacked them with the new rockets made of Salamian missile components.

Partial Cue

Salam was a small country. One day, Salam realized that its warlike neighbor, Bolon was arming against them. The Salamians realized that all of the Bolonian missiles were badly made and decided that an alliance could be formed. The Salam ambassador gave useful missile components to the Bolonian army. Bolonians promised that they would never fight against Salam from now on.

Predictive Cue

Salam was a small country. One day, Salam realized that its warlike neighbor, Bolon was arming against them. The Salamians realized that all of the Bolonian missiles were badly made and decided that an alliance could be formed. The Salam ambassador gave useful missile components to the Bolonian army. Bolonians promised that they would never fight against Salam. At that time, Salamians were not aware of the danger of forming an alliance with Bolonians.

Note. The underlined sentence in each Target and in Complete Cue is the conclusion sentence. The sentence with the dotted underline in Predictive Cue is the sentence that suggests the conclusion.

(2) Target and Cue Passages Used in Experiment 5 [Set B]

Same-Theme Target	*Different-Theme Target*
[Sour Grapes]	[Self-Blame]
John was very confident about himself. He did a lot of homework in order to get good marks. John had all A grades except for only one B+ in his first year in high school. He was sure he could do better. Earlier, a counselor had arranged for him to meet with the recruiter from Yale. When he got home from class, he opened the thin rejection letter from Yale. That night he mentioned to his father how he believed that people from Ivy League schools think too highly of themselves.	Derrick had failed to enter the gymnastics team last fall. He practiced a lot in order to enter the team. He wanted to try again. Derrick was positive he had a lot of potential. His Physical Education teacher had gotten him take a test for entering the team with the gymnastics team coach. The gymnastics team coach watched him perform and then told his teacher that he didn't want him on the team. Derrick confessed to his teacher that the coach undoubtedly thought he, Derrick, didn't have the talent for gymnastics.

Complete Cue

Lisa spent long hours trying to make her corporation successful. She was very sure about herself. Lisa had broken up with her fiancé a year ago. She wanted to meet someone new. A co-worker set her up for a date with an investment banker he knew well. She waited at the fancy restaurant until 8:30 p.m. and then left without ordering dinner. She told her friend that she thought that the man probably wasn't that handsome and that investment bankers are really boring, anyway.

Partial Cue

Lisa spent long hours trying to make her corporation successful. She was very sure about herself. Lisa had broken up with her fiancé a year ago. She wanted to meet someone new. A co-worker set her up for a date with an investment banker he knew well. She waited at the fancy restaurant until 8:30 p.m. and then left without ordering dinner.

Predictive Cue

Lisa spent long hours trying to make her corporation successful. She was very sure about herself. Lisa had broken up with her fiancé a year ago. She wanted to meet someone new. A co-worker set her up for a date with an investment banker he knew well. She waited at the fancy restaurant until 8:30 p.m. and then left without ordering dinner. Her pride was badly hurt because the man did not show up.

Note. The underlined sentence in each Target and in Complete Cue is the conclusion sentence. The sentence with the dotted underline in Predictive Cue is the sentence that suggests the conclusion.

(3) Target and Cue Passages Used in Experiment 5 [Set C]

Same-Theme Target	*Different-Theme Target*
[Taking appropriate action too late] Phil was in love with his secretary. However, he was afraid of taking his responsibility, so he kept dating others and made up excuses to postpone the proposal. Finally, his secretary got fed up, began dating another man, and fell in love with a doctor. When Phil found it out, he went to her and asked her to marry him, showing her the ring he had bought for her. <u>However, his secretary was already planning her honeymoon with the doctor by that time.</u>	[Reconciliation and learning from experience] Larry was madly in love with Wendy but did not ask her to go out with him. Wendy got frustrated, began going out with others, and became close to a football player. Unfortunately, the football player didn't want to be tied down, so he started seeing other girls, and made up reasons why he didn't want to see her regularly. Larry heard about this, so he brought her flowers, and asked if she would go to a party with him. <u>To his surprise, Wendy had been hoping he would ask her to do so.</u>

Complete Cue

Dr. Jones depended on his graduate student, Sandy. However, he kept finding reasons not to upgrade his research facilities because he was very cheap. After many useless meetings, Sandy began looking for a new job and was accepted to work at a corporate laboratory. When Dr. Jones heard the rumors, he offered Sandy lots of new research equipments suddenly. <u>But by that time, Sandy had decided to work on her new job at the corporate laboratory.</u>

Partial Cue

Dr. Jones depended on his graduate student, Sandy. However, he kept finding reasons not to upgrade his research facilities because he was very cheap. After many useless meetings, Sandy began looking for a new job and was accepted to work at a corporate laboratory. When Dr. Jones heard the rumors, he offered Sandy lots of new research equipments suddenly.

Predictive Cue

Dr. Jones depended on his graduate student, Sandy. However, he kept finding reasons not to upgrade his research facilities because he was very cheap. After many useless meetings, Sandy began looking for a new job and was accepted to work at a corporate laboratory. When Dr. Jones heard the rumors, he offered Sandy lots of new research equipments suddenly. <u>However, Sandy did not trust his words at all from his regular behaviors.</u>

Note. The underlined sentence in each Target and in Complete Cue is the conclusion sentence. The sentence with the dotted underline in Predictive Cue is the sentence that suggests the conclusion.