Reading in a Foreign Language April 2017, Volume 29, No. 1
ISSN 1539-0578 pp. 86–112

Standards of coherence in second language reading: Sentence connectivity and reading proficiency

Shingo Nahatame
Kyoei University
Japan

Abstract

Standards of coherence are one of the major factors that influence reading comprehension. This study investigated the standards of coherence that second language (L2) learners employ when reading. In a pair of experiments, Japanese learners of English read two-sentence texts with varying causal and semantic relatedness between sentences and then judged the coherence of the texts on a 5-point scale. Analysis of the judgment ratings indicated that both causal and semantic relatedness affected readers' perceived text coherence. The texts high in causal or semantic relatedness were judged as more coherent. Furthermore, when making judgments, lower proficiency readers were likely to place greater emphasis on semantic relatedness, whereas higher proficiency readers were likely to place greater emphasis on causal relatedness. These results suggest that L2 readers generally base standards of coherence on causal and semantic relatedness between sentences; however, readers of varying L2 skill levels may adopt different standards of coherence.

Keywords: coherence, discourse, causal relatedness, semantic relatedness, latent semantic analysis, reading proficiency, Japanese learners of English

It is widely accepted that a reader's comprehension of a text is influenced by both the text's characteristics (e.g., topic, vocabulary, and text structure) and the reader's cognitive traits (e.g., reading skills, working memory capacity, and background knowledge). In addition to these factors, some researchers suggest a third major factor, *standards of coherence*, or "the types and strengths of coherence the reader aims to maintain during reading" (van den Broek, Bohn-Gettler, Kendeou, Carlson, & White, 2011, p. 124). These standards play an important role in the reader's textual comprehension because the process and product of comprehension vary depending on what standards the reader holds (van den Broek, Rapp, & Kendeou, 2005; van den Broek, Risden, & Husebye-Hartmann, 1995). For example, readers who adopt a high standard would examine the coherence of a text more strictly than those who adopt a low standard.

Thus far, many studies on second language (L2) reading have examined the coherence of a text or of the memory representation that L2 learners construct from a text. However, few empirical studies have directly investigated what standards of coherence L2 learners employ while reading, even though these standards could closely relate to their process and product of comprehension.

Investigating standards of coherence adopted by L2 readers could further illustrate how they construct coherent text comprehension, the insights of which could suggest a basis for educational interventions in L2 reading. Accordingly, this study explored standards of coherence in L2 reading, specifically, by examining how L2 readers' perceptions of text coherence are affected by the characteristics of a text (i.e., two types of connectivity between sentences) and their reading proficiency. The following section provides a theoretical framework for the construct of standards of coherence as defined by the comprehension model, and reviews empirical findings concerning them. It also describes some previous studies on L2 reading closely related to this study.

Theoretical Framework for Standards of Coherence

Standards of coherence refer to the types and strengths of coherence that a reader tries to attain during text processing (van den Broek et al., 2011; van den Broek et al., 2005). A reader's standards of coherence can be either implicit (unconscious) or explicit (conscious) to the reader. These standards can vary in the type of coherence that is pursued, including but not limited to referential, causal, spatial, temporal, and logical coherence. They are also different in desired strength of coherence, such as a moderate or strong degree, mostly depending on the reader's goal (e.g., strong coherence is more required for reading to study than reading for entertainment). The standards are also influenced by reading skills and text types; consequently, these standards vary inter-individually and intra-individually.

In order to better understand what standards of coherence are, it seems important to differentiate two commonly confused terms: *coherence* and *cohesion*. Although coherence and cohesion are related concepts, there are several ways to discriminate between these terms. Generally, they are regarded as distinct aspects of text organization (Li & D'Angelo, 2015). Coherence in a text refers to the way in which text elements, such as sentences or phrases, belong together in terms of their meaning (i.e., they make sense in a discourse context), whereas cohesion refers to the way in which they belong together in terms of grammatical devices used for textual linkage, such as reference and conjunction (Warren, 2012). Cohesion is one of the factors that contribute to textual coherence; however, a text can be highly cohesive but not sufficiently coherent (Traxler, 2011). For example, consider the following texts: (a) "Laura got a lot of mail today. Therefore, her palms were sweaty"; (b) "Mary's exam was about to start. The palms were sweaty" (Ferstl & von Cramon, 2001, p. 328). In text (a), the two sentences are connected by cohesive ties, such as *therefore* and *her*, but the relationship between the two events described in these sentences is less clear (i.e., cohesive but pragmatically incoherent). On the other hand, text (b) appears more logical despite the absence of linguistic cohesion.

In recent discourse processing research (e.g., Graesser & McNamara, 2011; Graesser, McNamara, Louwerse, & Cai, 2004; McNamara, Graesser, McCarthy, & Cai, 2014), coherence is described as a characteristic of the mental representation of the text, whereas cohesion is described as a characteristic of the text itself. Specifically, coherence refers to the types of connectedness that readers make among propositions in discourse representations, whereas cohesion consists of explicit words, phrases, and sentences that help readers mentally connect ideas in the text. In this context, coherence can be derived from both the text information and the reader's prior

knowledge. For example, text (b) above appears coherent without cohesive devices because the connection between the sentences can be established based on the reader's prior knowledge about a test situation. Hence, cohesion should not be equated with coherence; rather, cohesion is a textual factor that may affect coherence. Nevertheless, when "we put the spotlight on the text as an object of investigation, coherence can be defined as characteristics of the text (i.e., aspects of cohesion) that are likely to contribute to the coherence of the mental representation" (Graesser & McNamara, 2011, p. 379).

To describe the process of establishing coherence in the understanding of discourse, psycholinguistic researchers have proposed several models of comprehension, including the construction-integration model, the structure-building framework, the resonance model, the event-indexing model, the causal network model, the constructionist theory, and the landscape model (see McNamara & Magliano, 2009, for a review). Standards of coherence are a central notion of *the landscape model* developed by van den Broek and colleagues (van den Broek et al., 2011; van den Broek et al., 2005; van den Broek et al., 1995). The landscape model is characterized according to its flexibility and emphasis on simulating individual differences in comprehension processes and skills. This model simulates the fluctuation of concept activation while reading, and illustrates the dynamic interaction of two types of processes that prior theories have already introduced: memory-based (i.e., passive and autonomous) and constructionist (i.e., active and strategic) processes. Specifically, it assumes that readers activate a set of concepts from the text itself during reading, as well as their prior knowledge, and then the activated concepts are evaluated per the readers' standards of coherence. If the activated concepts are sufficient for meeting those standards, then the connection between the concepts is made with ease and speed by passive, autonomous processes (called *cohort activation*). If, however, they are insufficient to satisfy the standards, then more effortful, strategic processes (called *coherence-based retrieval*) are initiated to establish a connection between the concepts and achieve the desired coherence. Thus, standards of coherence play a role in determining what processes readers need to engage in while reading.

A reader's standards of coherence can be estimated by examining the reader's perception of the coherence of various texts. It may also be important to distinguish between coherence of a text and a readers' perception of text coherence, although these concepts are not necessarily exclusive. As described above, coherence of a text is influenced or manipulated by textual constructs, such as the way a text conveys ideas, sentence connectives, and text organization. Text coherence has traditionally been evaluated through propositional analysis (Kintsch & van Dijk, 1978) and causal network analysis (Trabasso & van den Broek, 1985). On the other hand, perception of text coherence is assessed through the reader's judgments of whether the text is coherent and to what extent it is coherent (Todaro, Millis, & Dandotkar, 2010; Wittwer & Ihme, 2014). Although it might seem easy for readers to perceive the coherence of the text, such perception is "the outcome of an intricate interplay of cognitive processes that involve memory, background knowledge, textual constraints, processing limitations, and so on" (van den Broek et al., 1995, p. 353). Thus, perception of text coherence is likely to vary according to individual differences in the background knowledge and reading skills necessary to construct a text representation (Graesser et al., 2004; Todaro et al., 2010). Consequently, investigating how L2 readers perceive the coherence of various texts (i.e., estimating standards of coherence that the readers employ) may contribute to clarifying the individual differences in L2 reading.

Empirical Support for Standards of Coherence: Effects of Causal and Semantic Relatedness

In the area of first language (L1) reading research, there is ample evidence that a reader applies a set of standards of coherence to his or her text comprehension. Although there are various types of coherence as mentioned above, researchers have often examined causal and semantic coherence by measuring the effects of causal and semantic relatedness between sentences on comprehension (e.g., Todaro et al., 2010; Wittwer & Ihme, 2014; Wolfe, Magliano, & Larsen, 2005). Causal relatedness refers to cause-and-effect relations between events described in the sentences, whereas semantic relatedness refers to semantic overlap between propositions of the sentences.

Researchers have focused on causal and semantic relatedness for several reasons. First, several theories suggest that these types of relatedness contribute to text cohesiveness, perceived coherence, comprehension processes, and text production (e.g., Graesser et al., 2004; Graesser, Singer, & Trabasso, 1994; Kintsch & van Dijk, 1978). Indeed, empirical evidence has shown that when narrative statements have multiple causal connections to other statements, they are rated as more important and more frequently recalled (Trabasso & van den Broek, 1985; van den Broek, 1988). Similarly, semantic overlap between sentences has been found to provide an accurate measure of text cohesiveness and reader comprehension (Foltz, Kintsch, & Landauer, 1998). Second, causal and semantic relatedness correspond to the levels of mental representation of a text. Semantic relatedness is preserved at surface and textbase levels of comprehension which only contains explicit meanings conveyed by the text (Kintsch & van Dijk, 1978). Causal relatedness, on the other hand, is inferred by the reader and preserved at deeper levels of comprehension which contains information from the text and inferences generated by the reader (Graesser et al., 1994); this is the situation model. Hence, comparing the effects of causal and semantic relatedness helps reveal individual differences in the levels of text comprehension.

With respect to causal relatedness, certain researchers conducted experiments using two-sentence texts that varied in causal relatedness (Keenan, Baillet, & Brown, 1984; Myers, Shinjo, & Duffy, 1987). For instance, consider the following texts: (a) “Dick kicked the living room wall in his fury. A mirror fell down and shattered on the floor”; (b) “Dick came home late after his evening class. A mirror fell down and shattered on the floor” (Myers et al., 1987, pp. 461-462). These two texts share the second sentence, but the first sentence of (a) provides better causation for the second sentence than the first sentence of (b). This means text (a) has stronger causal relatedness than text (b). In the experiments, participants were asked to read these kinds of texts; then, they were provided either the first or second sentence as cue and were asked to recall the content of the paired sentence. The result showed that the stronger the causal relatedness the sentences had, the faster the reader would process the texts. Additionally, readers could better recall more causally related sentences than less causally related sentences.[1] These results suggest that L1 readers employ their standards of coherence, which are influenced by the degree of causal relatedness of sentences, to achieve their understanding of the meaning conveyed by a text. Consequently, the process and result of reading vary depending on whether the text meets these standards.

In Wolfe et al. (2005), the sentence pairs were manipulated in terms of semantic relatedness, in

addition to causal relatedness. For example, consider the following text: (c) "Dick felt the earthquake hit with sudden force. A mirror fell down and shattered on the floor." The causal relatedness of the text in (c) seems as strong as the text in (a) above; however, the first sentence of (a) has a stronger semantic relation to its second sentence than does the first sentence of (c), because there is more semantic overlap between the words (e.g., "room" and "wall" in the first sentence with "mirror" and "floor" in the second sentence) as well as the propositions conveyed by the sentences. Such semantic relatedness is measured by a technique called *latent semantic analysis (LSA)*, which estimates general semantic relatedness between text units based on direct and indirect relations among words in a large corpus (Landauer, Foltz, & Laham, 1998). Based on their own research, Wolfe et al. concluded that both semantic and causal relatedness had an impact on processing time and the amount of textual information recalled. That is, sentences high in causal or semantic relatedness were read faster and were better recalled, although the processing time was more strongly influenced by causal, rather than semantic relatedness. This suggests that L1 readers employ standards of coherence based not only on causal relatedness but also on semantic relatedness (as measured by LSA).

Todaro et al. (2010) specifically examined standards of coherence employed by L1 readers. In their study, participants were instructed to read the sentence pairs used in Wolfe et al. (2005) and to judge the coherence of the pairs using a "yes" or "no" (i.e., whether they perceived that the text was or was not coherent) forced choice. The results indicated that the participants more often responded "yes" to the sentence pairs high in causal or semantic relatedness. More importantly, there was an interaction between causal and semantic relatedness, such that the effect of semantic relatedness was greater when the texts had low causal relatedness whereas the effect of causal relatedness was greater when the text had low semantic relatedness. These findings provide clearer evidence that L1 readers base their standards of coherence on both causal and semantic relatedness, and suggest that these factors influence coherence perception in an interactive manner.

Furthermore, Todaro et al. (2010) showed that semantic and causal relatedness influenced coherence judgments differently for skilled and less skilled readers. The results indicated that causal relatedness had a greater impact on judgments made by skilled readers, while semantic relatedness had a greater impact on judgments made by less skilled readers. As described above, semantic relatedness is likely to be preserved at surface and textbase levels of comprehension, whereas causal relatedness is preserved at deeper levels of comprehension. Thus, this result suggests that less skilled readers tend to build comprehension of the text at surface- and textbase-levels, while skilled readers tend to build comprehension at deeper levels. Although these findings are informative with respect to the relationship between standards of coherence and reading skills, Todaro et al. pointed out that one limitation of their study is the use of a binary scale for coherence judgments.

Wittwer and Ihme (2014) conducted similar experiments as Todaro et al. (2010), but used longer passages that explained phenomena in physics. Nevertheless, in line with Todaro et al., the results showed that semantic similarity was more important for less skilled readers' judgments, whereas causal specificity was more important for skilled readers' judgments.

In sum, the above-mentioned studies provide some empirical support for standards of coherence

in L1 reading. First, L1 readers hold in their minds several types of standards of coherence, including causal and semantic coherence. Second, it matters in reading comprehension whether or not the text meets the employed standards of coherence. For example, reading processes and resulting text representations vary depending on the causal and semantic relatedness of the text. Third, readers have different standards of coherence depending on their reading skills. Particularly, less skilled readers likely have a standard of coherence that emphasizes semantic relatedness, whereas skilled readers tend to have a standard of coherence that emphasizes causal relatedness.

Building Coherent Text Representations in L2 Reading

Like the field of L1 reading research, many studies have examined how L2 readers construct coherent mental representations of a text. Previous studies have identified various text and reader factors that affect the coherence of L2 readers' text representations, such as text genre, topic, and readers' L2 proficiency (e.g., Horiba, 2000, 2013; Lee, 2009; Ushiro, 2010; Yoshida, 2012). Studies have also suggested that, compared to L1 reading, higher order comprehension processes are somewhat limited for L2 reading (e.g., Horiba, 1996a, 2000, 2013; Morishima, 2013; Nahatame, 2014; see also Li & D'Angelo, 2015, for a review). These higher level processes include exploiting text structure for comprehension, reactivating information previously described in the text, generating inferences while reading, and altering the reading process for given goals. L2 readers often face difficulties with these processes because much of their cognitive resources are allocated to lower comprehension processes, such as word recognition and syntactic parsing (Horiba, 2000).

Some L2 studies have conducted similar experiments as in L1 studies using two-sentence texts. Horiba (1996b) examined L2 readers' representations of two-sentence texts that had different degrees of causal relatedness. When L2 readers of Japanese were instructed to study and memorize each text, highly related sentences were recalled better than less related sentences. Shimizu (2009) conducted a similar experiment as Horiba, with Japanese university students learning English. She also measured participants' reading times in addition to recall performance. The results indicated the effects of causal relatedness on both reading time and recall measures. Sentences high in causal relatedness were processed faster and recalled better than those that were low in causal relatedness. These results are similar to the findings from L1 research, thus suggesting that L2 readers employ similar standards of causal coherence as L1 readers while reading two-sentence texts.

These previous studies suggest that the standards of coherence that L2 learners employ while reading paired sentences are based on the causal relatedness between the sentences. However, only indirect evidence, such as the analysis of written recall protocols and reading times, has suggested this view of standards of coherence in L2 reading. Few studies have directly examined the standards of coherence employed in L2 reading, for example, by asking readers to judge the coherence of sentence pairs.

Furthermore, most previous L2 studies manipulated sentence connectivity only by causal relatedness. Some L1 studies indicated the significant role of semantic relatedness in text

comprehension, which is measured using LSA (Foltz et al., 1998; Todaro et al., 2010; Wolfe et al., 2005). LSA has also been used in recent L2 studies, and these studies suggest that semantic relatedness between sentences computed by LSA is one of the indications of L2 text readability (e.g., Crossley, Allen, & McNamara, 2012; Crossley, Louwerse, McCarthy, & McNamara, 2007; Crossley & McNamara, 2008; Nahatame, 2012). Therefore, it is possible that semantic relatedness impacts L2 readers' text comprehension, for example, the perception of text coherence. However, little research has been conducted to examine this possibility.

Study

As reviewed above, the comprehension model suggests that standards of coherence are a major factor that influences text comprehension, and empirical evidence has supported the role of these standards in L1 reading. However, few studies have directly examined L2 readers' standards of coherence, although some studies have provided indirect evidence. Thus, the present study was conducted to bridge this research gap and explore the properties of standards of coherence employed in L2 reading.

The factors manipulated in this study were causal and semantic relatedness between sentences. Studies have suggested that these types of relatedness contribute to the process of establishing coherence in text comprehension and correspond to levels of comprehension (i.e., surface and textbase levels vs. situation-model level). Thus, examining standards of causal and semantic coherence among L2 readers will provide more detailed information about how L2 readers achieve coherent and deep text comprehension, which may, in turn, help improve the quality of educational interventions aimed at L2 readers. Furthermore, compared to causal relatedness, it remains unclear whether—and how—semantic relatedness plays a role in L2 readers' text comprehension.

In this study, participants were asked to judge the coherence of sentence pairs varying in causal and semantic relatedness. Coherence judgments were conducted using a 5-point scale because it enabled the participants to monitor more precise levels of coherence than a binary scale (cf. Todaro et al., 2010). The effects of causal and semantic relatedness on coherence judgments were examined in relation to the participants' L2 reading proficiency. Consequently, the following two research questions (RQs) were addressed in this study:

1. Do L2 readers judge text coherence differently according to the causal and semantic relatedness between sentences?
2. Do the patterns of L2 readers' coherence judgments differ according to their L2 reading proficiency?

Two experiments were conducted within the present study to address these questions. Experiment 1 explored the RQs by using the paired sentences and the coherence judgment task mentioned above. Experiment 2, which was an approximate replication of Experiment 1, was conducted to confirm and better understand the findings of Experiment 1. Experiment 2 was designed to help us determine whether the original finding might be an outcome of a specific condition setting or if it more likely reflects a general phenomenon.

Given earlier findings that L2 readers are sensitive to the causal relatedness of two-sentence texts (Horiba, 1996b; Shimizu, 2009), causal relatedness is assumed to significantly impact L2 readers' coherence judgments. On the other hand, the effect of semantic relatedness remains unclear. Nevertheless, the findings from L2 studies using LSA (e.g., Crossley et al., 2012; Crossley et al., 2007; Crossley & McNamara, 2008; Nahatame, 2012) have implied that semantic relatedness may potentially have effects on text coherence as perceived by L2 readers.

Furthermore, L2 readers of varying skill levels might place differing emphases on causal and semantic relatedness when judging coherence. L1 studies suggest that less skilled readers tend to build comprehension at the surface and textbase levels, which only contain information about word meaning and the explicit ideas conveyed by text; therefore, their coherence judgments are more likely influenced by semantic relatedness. In contrast, skilled readers usually build comprehension at the situation model level, which includes causality inferred by the reader; consequently, their coherence judgments are more likely affected by causal relatedness (Todaro et al., 2010; Wittwer & Ihme, 2014). If the same holds for L2 readers, less proficient readers may emphasize semantic relatedness more in their coherence judgments, whereas proficient L2 readers may emphasize causal relatedness more in their coherence judgments.

As described above, L2 reading proficiency was an important reader variable in this study. Although L2 reading proficiency is closely related to overall L2 proficiency and general comprehension skills, it was necessary to consider them separately here. L2 proficiency overall is often measured by a test assessment not only of L2 reading skills but also of other skills and knowledge in listening, writing, speaking, and grammar (e.g., Horiba, 2013; Horiba & Fukaya, 2015); L2 reading proficiency, in contrast, is assessed only by the reading subsection of a standardized test (e.g., Nahatame, 2014; Yoshida, 2012). General comprehension skills are usually tested via L1-based comprehension measures (e.g., recall of an L1 text as in Horiba, 2013). Empirical evidence for standards of coherence suggests that the standards in L1 reading vary according to reading skills as measured by a standardized reading test (Todaro et al., 2010; Wittwer & Ihme, 2014). Thus, given that participants of the present study were asked to read texts in their L2, L2 reading proficiency (as measured by an L2 reading test) was considered a more appropriate variable rather than overall L2 proficiency or general comprehension skills.

Experiment 1

Method

Participants

The participants were 49 university students in Japan (37 females and 12 males). Many of them were first-year students (n = 37) and the remaining participants were second- or third-year students. They were all native speakers of Japanese, and they had studied English as a foreign language for more than six years as part of their formal Japanese education. Their English proficiency could be assumed to be at beginning to intermediate levels (confirmed by testing). Before the experiment, participants received an explanation of the purpose and procedures of the

experiment. They were also told that participation was voluntary and that they could withdraw their participation at any time.

L2 reading proficiency measure

An L2 (English) reading proficiency test was created based on retired copies of the reading subsection of the Eiken Test (constructed by the Eiken Foundation, Japan). Although there are other standardized English tests available in Japan, the Eiken Test, supported by the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT), has been taken widely in Japan (more than 2 million examinees annually at 18,000 test sites) and for a long time (more than 50 years) to measure the English proficiency of Japanese students.[2] There are seven grades (levels) in the Eiken Test, and each grade has different test sets (with different difficulties) but with a similar test format. Because most Japanese students are familiar with the test, and it would allow us to choose a suitable test set according to the participants' English proficiency, this study selected part of the Eiken Test to measure the participants' English reading proficiency.

Some of the Eiken grades are recognized as benchmark for Japanese students at different levels from high school to tertiary; for example, Grade 3 is recognized as a benchmark for junior high school graduates, and Grades Pre-2 and 2 are recognized as benchmarks for high school graduates (these grades are also recognized as A1, A2, and B1 of the Common European Framework of Reference for Languages). Considering that most of the participants in this study were first-year university students and estimated to be beginning to intermediate level learners, the test passages were chosen from the sets of Grades 3, Pre-2, and 2. The present reading proficiency test included four passages (one from Grade 3, two from Grade Pre-2, and one from Grade 2) that were each paired with four or five multiple-choice questions, totaling 18 items.

Reading materials for coherence judgments

The reading materials used for coherence judgments were adopted from Wolfe et al. (2005). The original materials were sets of paired sentences written for L1 English-speaking readers. Although the original materials consisted of 32 sets of sentence pairs, the present study used only some of them because some pairs seemed difficult for Japanese learners of English to understand due to cultural and lexical reasons. Specifically, some pairs described situations closely related to American culture (including phrases such as "pastor's sermon" and "Republican in the state senate"), and other pairs included words that are context-specific (e.g., words such as "verdict" and "defendant" would appear in the legal court scene). To avoid the influence of specific cultural and lexical knowledge, the present study did not use such pairs. Consequently, 20 pairs, each of which seemed appropriate and comprehensible for participants of the present study, were selected from the original sets. Although reducing the number of materials might have decreased the generalizability of research findings, it allowed us to ease the participants' burden of reading and shorten the length of experimental sessions. These considerations were more necessary in the experiments with L2 readers compared to those with L1 readers.

Furthermore, the selected texts were simplified to make them more suitable for Japanese learners of English. Using the JACET 8,000 word frequency list (Committee of Revising the JACET Basic Words, 2003), which is a reliable frequency word list for Japanese learners of English,

low-frequency words (including words not listed) were replaced with high-frequency synonyms. This 8,000-item list of basic words is divided into Level 1 (the most frequent 1,000 words of the list) to Level 8 (the least frequent 1,000 words of the list). Consequently, this revision ensured that the texts mostly consisted of words at level 4 and below. Some words that are difficult to paraphrase were glossed (e.g., "weed" and "spoiled"). In addition, some sentence structures were simplified, with an effort to retain their main message (e.g., "She thought it hot in her room" → "it was hot in the room"; "met a guy that she really liked" → "met a handsome, tall guy"). Table 1 shows the mean number of words, readability, and vocabulary levels of the finalized texts. All the texts used in this and in the following experiment are shown in Appendix.

Table 1. *Number of words, readability, and vocabulary level of the texts used in experiments*

	M	95% CI	*SD*	*Max*	*Min*
			Number of Words		
First Sentence	8.43	[7.67, 9.18]	1.72	13	5
Second Sentence	7.40	[6.76, 8.04]	1.46	10	5
			Readability (Flesch-Kincaid Grade Level)		
First Sentence	4.17	[3.04, 5.29]	2.56	10.02	0.00
Second Sentence	3.38	[2.47, 4.28]	2.05	9.00	0.00
			Vocabulary Level (based on the JACET 8,000)		
First Sentence	1.44	[1.23, 1.64]	0.47	3.20	1.00
Second Sentence	1.47	[1.25, 1.68]	0.48	3.13	1.00

Note. Flesch-Kincaid Grade Level was calculated using Word 2010; Vocabulary Level is the mean of the JACET 8,000 level of each word included in the sentence.

Each sentence set had four types of the first (prime) sentence and one common second (target) sentence, as shown in Table 2. First sentences varied in terms of their causal and semantic relatedness to the second sentence (high or low), thus resulting in four variables: Causal Relatedness (CR)-High and Semantic Relatedness (SR)-High, CR-High and SR-Low, CR-Low and SR-High, and CR-Low and SR-Low. There was only a small difference in the length, readability, and vocabulary level of the first sentences between the four variables. The mean number of words ranged from 7.80 to 9.00 between the four variables, the mean readability ranged from 3.68 to 5.00, and the mean vocabulary level ranged from 1.33 to 1.55 (all Fs < 2.10, ps > .10).

Table 2. *Sample of experimental texts*

1a.	Mary could not find anything to read in the library.	(CR-High / SR-High)
1b.	Mary wanted to look for recipes for her dinner party.	(CR-High / SR-Low)
1c.	Mary went to the library to look for something to read.	(CR-Low / SR-High)
1d.	Mary was having a dinner party for her office.	(CR-Low / SR-Low)
2.	She went to the bookstore to get new books.	(Target sentence)

Note. CR = Causal Relatedness; SR = Semantic Relatedness.

Each sentence set had two prime sentences high in causal relatedness to the target sentence, and two sentences low in causal relatedness. For example, sentences 1a and 1b in Table 2, which were labeled CR-High, state an event (i.e., finding nothing to read/looking for recipes) that is likely to cause the event described in the target sentence (i.e., going to the bookstore to get new books). On the other hand, the events described in sentences 1c and 1d (i.e., going to the library/having an office dinner party), which were labeled CR-Low, are less likely to cause the

event described in the target sentence.

Similar to causal relatedness, two prime sentences of each sentence set had high semantic relatedness to the target sentence, and the other two sentences had low semantic relatedness. When semantic relatedness was high, the content words of the target sentences were more similar to the words in the prime sentences than when the relatedness was low. For example, sentences 1a and 1c in Table 2, labeled SR-High, include words such as "library" and "read" that are semantically related to "book" in the target sentence. On the other hand, sentences 1b and 1d, labeled SR-Low, do not include such semantically related words.

Because the present materials were slightly revised from the original ones, it was necessary to obtain data to verify the manipulation of the causal and semantic relatedness of the revised materials. Thus, following Wolfe et al.'s (2005) verification approach, causal manipulation was assessed in the preliminary study by asking participants to make subjective judgments, whereas semantic manipulation was ensured by employing LSA.

The preliminary study for ensuring causal manipulation was conducted with eight Japanese graduates studying English. Participants made goodness of answer judgments where the target sentence was presented as a why question (e.g., "Why did Mary go to the bookstore to get new books?"), and all the four prime sentences were given as answers to the question (e.g., "Because she could not find anything to read in the library"). Participants judged each answer on a 6-point scale ranging from 1 (*very bad answer*) to 6 (*very good answer*). The results showed that across all sentence pairs, sentences high in causal relatedness were rated higher (M = 5.25 [4.95, 5.55], SD = 0.74) than those low in causal relatedness (M = 1.96 [1.80, 2.12], SD = 0.72), $F(1, 19) = 400.37$, $p < .001$, $\eta_G^2 = .86$.

The manipulation of semantic relatedness was verified based on LSA similarities. It is important to note that semantic similarities estimated with LSA is characterized as a general property, rather than as mere repetition of specific words or even close synonyms. Following the procedure employed by Wolfe et al. (2005),[3] LSA cosines (i.e., semantic similarities approximately ranging from -0.1 to 1.0) were calculated between each of the four prime sentences and the target sentence. Across all sentence pairs, sentences high in sematic relatedness had significantly higher cosines (M = 0.29 [0.27, 0.31], SD = 0.11) than sentences low in semantic relatedness (M = 0.03 [0.02, 0.04], SD = 0.04), $F(1, 19) = 153.78$, $p < .001$, $\eta_G^2 = .71$. LSA cosines were not different as a function of causal relatedness, $F(1, 19) = 0.06$, $p = .816$, $\eta_G^2 = .00$, nor was there an interaction between causal and semantic relatedness, $F(1, 19) = 0.09$, $p = .767$, $\eta_G^2 = .00$.

Four lists of sentence pairs, each of which included 20 pairs, were constructed to counterbalance the four variables, ensuring that each sentence pair appears no more than once in a list, and all pairs appear in all four variables when all four lists are considered. One of the four lists was randomly assigned to each participant to ensure that every participant read five pairs in each of the four variables, and an almost equal number of participants read each pair in one of the four variables.

Procedure

Before the main experiment, participants completed the L2 reading proficiency test within 35 minutes each. In the main experiment, participants were randomly assigned one of the four lists of sentence pairs. Participants were told to read the sentence pairs for comprehension and that they would have a comprehension test about the sentences later. They were also instructed to judge the coherence of paired sentences after reading each pair. In the instructions, coherence was simply defined as "when two sentences are related to each other and make sense together" following Todaro et al.'s (2010, p.427) definition to avoid influencing the participants' judgments (these instructions were given in Japanese). The coherence judgment task was scored on a 5-point Likert-type scale ranging from 1 (*not coherent at all*) to 5 (*very coherent*). Participants were asked to circle the appropriate number on the scale below each sentence pair. They finished the task within 20 minutes each.

Results and Discussion

L2 reading proficiency test

The analysis identified four of 18 test items that had low discriminability (i.e., item-total correlation); if these items were eliminated, the reliability of the test would improve. Thus, these four items were dropped from the test sets, and the test consequently showed acceptable reliability (Cronbach's $\alpha = .73$ for the remaining 14 items). Based on a median split of test scores, the participants were classified into two proficiency groups: higher ($n = 25$, $M = 10.40$ [9.79, 11.01], $SD = 1.44$, $Max = 13$, $Min = 8$) and lower ($n = 24$, $M = 5.33$ [4.66, 6.01], $SD = 1.57$, $Max = 7$, $Min = 1$). There was a significant difference in test scores between these two groups, $t(47) = 11.52$, $p < .001$, $d = 3.29$.

Coherence judgment

Table 3 shows the means, 95% CIs, and standard deviations of coherence ratings to demonstrate the function of causal and semantic relatedness and the participants' L2 reading proficiency. Figure 1 represents the distribution of the coherence rating data in the form of box plots.

Table 3. *Means, 95%CIs, and standard deviations for the coherence judgment task in Experiment 1*

	SR-High			SR-Low		
Higher ($n = 25$)	*M*	95% CI	*SD*	*M*	95% CI	*SD*
CR-High	4.54	[4.34, 4.74]	0.49	3.99	[3.75, 4.23]	0.58
CR-Low	3.15	[2.91, 3.39]	0.58	1.96	[1.77, 2.14]	0.45
Lower ($n = 24$)						
CR-High	4.40	[4.21, 4.60]	0.46	3.55	[3.28, 3.82]	0.63
CR-Low	3.20	[2.97, 3.44]	0.55	1.73	[1.42, 2.03]	0.72
Total ($N = 49$)						
CR-High	4.47	[4.34, 4.61]	0.47	3.77	[3.59, 3.96]	0.64
CR-Low	3.18	[3.01, 3.34]	0.56	1.84	[1.67, 2.02]	0.60

Note. CR = Causal Relatedness; SR = Semantic Relatedness.

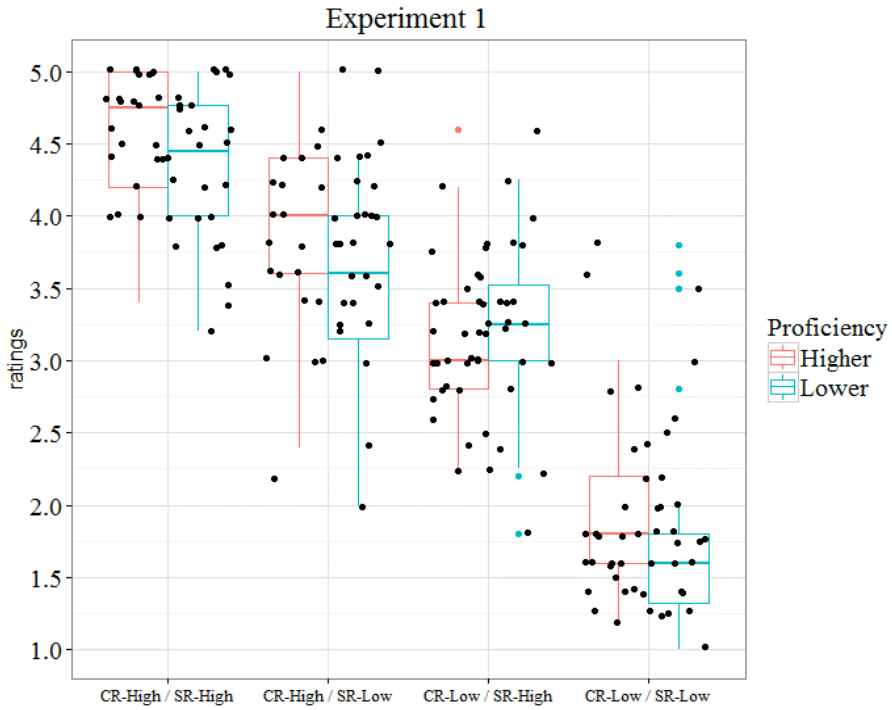

Figure 1. Box plots of the coherence judgment ratings for Experiment 1 as a function of causal and semantic relatedness and L2 reading proficiency. CR = Causal Relatedness; SR = Semantic Relatedness.

A mixed three-way analysis of variance (ANOVA) was conducted on the mean coherence ratings with causal and semantic relatedness as within-participants variables and L2 reading proficiency as a between-participants variable. The results revealed the significant main effects of both causal and semantic relatedness with higher ratings for sentences high in relatedness than of those low in relatedness (see Table 4). The main effect of L2 reading proficiency was also significant. Higher proficiency readers' ratings were slightly higher than were those of lower proficiency readers.

More importantly, a significant interaction between causal and semantic relatedness was found. Post-hoc comparisons found that the simple main effects of causal relatedness were significant for sentences that were both high and low in semantic relatedness, $F(1, 47) = 194.35$, $p < .001$, $\eta_G^2 = .62$; $F(1, 47) = 219.84$, $p < .001$, $\eta_G^2 = .73$. Similarly, the simple main effects of semantic relatedness were significant for sentences that were both high and low in causal relatedness, $F(1, 47) = 36.11$, $p < .001$, $\eta_G^2 = .30$; $F(1, 47) = 146.20$, $p < .001$, $\eta_G^2 = .58$. These results indicate that sentence pairs were judged as most coherent when both causal and semantic relatedness were high, and judged as least coherent when both causal and semantic relatedness were low. In addition, the sizes of causal relatedness effects were greater for sentences that were low in semantic relatedness than of those that were high in semantic relatedness ($\eta_G^2 = .73$ vs. .62; M_{diff} = 1.93 [1.68, 2.18] vs. 1.29 [1.08, 1.50]). Similarly, the sizes of semantic relatedness effects were greater for sentences low in causal relatedness than were those high in causal relatedness (η_G^2 = .58 vs. .30; M_{diff} = 1.34 [1.11, 1.57] vs. 0.70 [0.48, 0.93]). Thus, causal relatedness had greater effects on judgments of less semantically related sentences than more semantically related sentences, whereas semantic relatedness had greater effects on judgments of less causally related

sentences than more causally related sentences. These differences were likely the source of the interaction effect.

Table 4. *Summary table for three-way ANOVA of the effects of L2 reading proficiency, causal relatedness, and semantic relatedness on coherence ratings for Experiment 1*

Source	*SS*	*df*	*MS*	*F*	*p*	η_G^2
			Between-participants			
Proficiency (P)	1.76	1	1.76	5.85	.020	.03
Error (P)	14.17	47	0.30			
			Within-participants			
Causal Relatedness (C)	50.84	1	50.84	375.64	< .001	.68
C× P	0.51	1	0.51	1.51	.225	.01
Error (C× P)	15.92	47	0.34			
Semantic Relatedness (S)	50.84	1	50.84	147.02	< .001	.46
S × P	1.06	1	1.06	3.05	.087	.02
Error (S × P)	16.25	47	0.35			
C× S	4.92	1	4.92	17.11	< .001	.08
C × S × P	0.00	1	0.00	0.00	.967	.00
Error (C × S × P)	13.50	47	0.29			

In sum, the above results demonstrated that both causal and semantic relatedness have a significant impact on L2 readers' coherence judgments of two-sentence texts. L2 readers were likely to judge sentences that were high in causal or semantic relatedness to be more coherent than those low in the relatedness. In addition, these factors also influence coherence judgments in an interactive manner, such that the effects of causal relatedness were stronger when the semantic relatedness of texts was low, whereas the effects of semantic relatedness were more prominent when the causal relatedness was low. These results are consistent with those of Todaro et al. (2010) with L1 English-speaking readers.

Furthermore, the analysis indicated that the interaction effect between L2 proficiency and semantic relatedness approached a significant level (see Table 4). This raises the possibility that L2 reading proficiency intervenes in the influence of semantic relatedness on coherence judgments. One possible interpretation of this marginally significant interaction is that the interaction of L2 reading proficiency and semantic relatedness may actually have influence, but not in the context of this experiment. Therefore, to answer clearly the second RQ (i.e., whether the effects of causal and semantic relatedness differ for higher and lower proficiency readers), the following experiment further examined this issue with some changes in the experimental design. Additionally, it also aimed to replicate the findings of Experiment 1 regarding the first RQ, that is, the interaction effect of causal and semantic relatedness.

Experiment 2

The purpose of Experiment 2 was to confirm and extend the findings of Experiment 1. Specifically, Experiment 2 attempted to replicate the interaction between causal and semantic

relatedness on coherence judgments by L2 readers. It further addressed whether participants with different levels of L2 reading proficiency show different patterns of coherence judgments.

Note that Experiment 2 manipulated causal relatedness between participants rather than within each participant for the following reason. According to Wolfe et al. (2005), semantic relatedness can be a subtler manipulation for readers than causal relatedness; therefore, the focus of the readers is likely to be on causal relatedness. In other words, the effects of semantic relatedness might possibly be overwhelmed by those of causal relatedness, to some extent. This suggests that when causal relatedness is constant within each participant, readers may produce more focus on semantic relatedness. Thus, when the experiment manipulates causal relatedness between participants, the interaction effect of semantic relatedness and L2 proficiency, which was less clear in Experiment 1, may become more prominent. Experiment 2 examined this possibility. Furthermore, if the interaction effect between causal and semantic relatedness were found as in Experiment 1, this effect would be interpreted as not the consequence of a specific experimental design but rather a general phenomenon.

Method

Participants

Participants in this experiment included 104 university students in Japan (61 females and 43 males). They were all native speakers of Japanese, and none of them had participated in Experiment 1. Most of them were first-year students (n = 102) and the remaining participants were third-year students. They were assumed to possess beginning to intermediate levels of English proficiency and had studied English as a foreign language for more than six years as part of their formal Japanese education. Participants received a similar explanation of the experiment as in Experiment 1.

Materials and procedure

Before the main experiment, they completed the same L2 reading proficiency test as in Experiment 1. The experimental materials were the 20 sets of sentence pairs used in Experiment 1.

Participants were randomly assigned each of the two causal relatedness conditions (CR-High, CR-Low). Participants with the CR-High condition (n = 49) read sentence pairs that were consistently high in causal relatedness, while those with the CR-Low condition (n = 55) were provided sentence pairs that were consistently low in causal relatedness. In each condition, semantic relatedness was manipulated and counterbalanced similarly as in Experiment 1. The task procedure was identical to Experiment 1: Participants read the sentence pairs for comprehension and made coherence judgments on each sentence pair using a 5-point scale.

Results and Discussion

L2 reading proficiency test

As with the first experiment, the analysis identified four of 18 test items that had low discriminability. Thus, these four items were eliminated from the test sets (two of these items had also been eliminated in Experiment 1), resulting in the test's acceptable reliability (Cronbach's $\alpha = .72$ for the remaining 14 items).

Based on a median split of test scores, participants with the CR-High condition were categorized into higher ($n = 29$, $M = 10.93$ [10.30, 11.57], $SD = 1.67$, $Max = 14$, $Min = 9$) and lower ($n = 20$, $M = 6.00$ [5.21, 6.79], $SD = 1.69$, $Max = 8$, $Min = 2$) proficiency groups. Based on the same criteria, participants with the CR-Low condition were divided into higher ($n = 28$, $M = 10.43$ [9.96, 10.89], $SD = 1.20$, $Max = 13$, $Min = 9$) and lower ($n = 27$, $M = 5.63$ [4.95, 6.31], $SD = 1.71$, $Max = 8$, $Min = 2$) proficiency groups. When including proficiency groups and causal relatedness as between-participants factors, a two-way ANOVA on test scores only indicated the significant main effect of proficiency, $F(1, 100) = 243.75$, $p < .001$, $\eta_G^2 = .71$, showing that the higher proficiency group got higher scores than did the lower proficiency group. The main effect of causal relatedness and the interaction of proficiency and causal relatedness were not significant, $F(1, 100) = 4.85$, $p = .164$, $\eta_G^2 = .02$; $F(1, 100) = 0.04$, $p = .833$, $\eta_G^2 = .00$.

Coherence judgment

Table 5 shows the means, 95%CIs, and standard deviations of coherence ratings for Experiment 2. Figure 2 represents the distribution of the coherence ratings in the form of box plots.

Table 5. *Means, 95%CIs, and standard deviations for the coherence judgment task in Experiment 2*

	SR-High			SR-Low		
Higher ($n = 57$)	*M*	95% CI	*SD*	*M*	95% CI	*SD*
CR-High ($n = 29$)	4.53	[4.32, 4.74]	0.54	4.11	[3.83, 4.39]	0.74
CR-Low ($n = 28$)	3.28	[3.00, 3.56]	0.72	1.79	[1.58, 2.01]	0.55
Lower ($n = 47$)						
CR-High ($n = 20$)	4.62	[4.41, 4.83]	0.45	3.38	[2.90, 3.86]	1.04
CR-Low ($n = 27$)	3.55	[3.29, 3.81]	0.66	1.93	[1.70, 2.17]	0.60
Total ($N = 104$)						
CR-High ($n = 49$)	4.56	[4.42, 4.71]	0.50	3.81	[3.54, 4.08]	0.94
CR-Low ($n = 55$)	3.41	[3.22, 3.60]	0.70	1.86	[1.71, 2.02]	0.58

Note. CR = Causal Relatedness; SR = Semantic Relatedness.

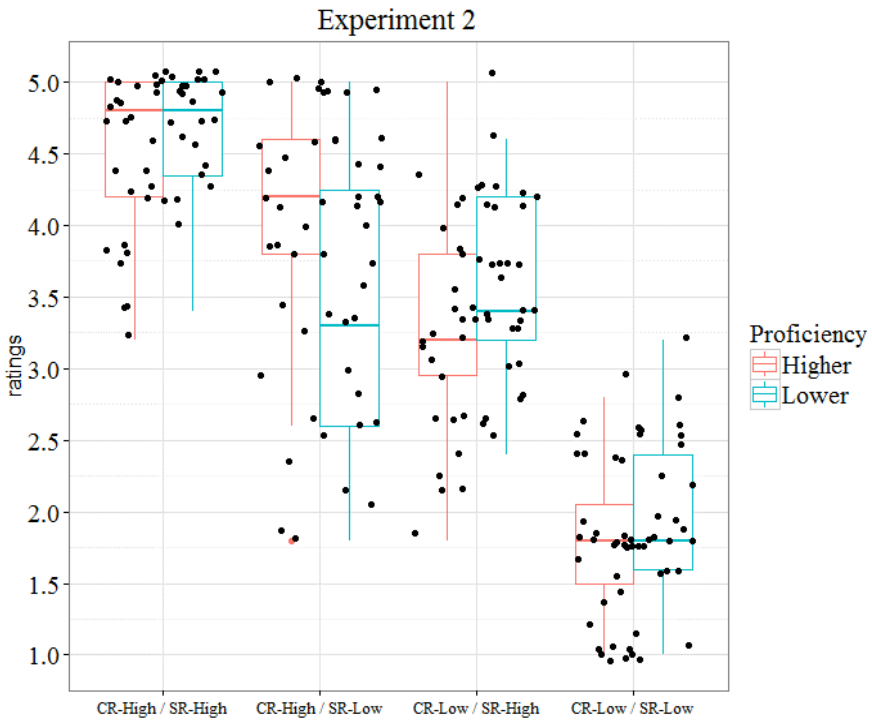

Figure 2. Box plots of the coherence judgment ratings for Experiment 2 as a function of causal relatedness, semantic relatedness, and L2 reading proficiency. CR = Causal Relatedness; SR = Semantic Relatedness.

A mixed three-way ANOVA was conducted on the mean coherence ratings in Experiment 2, with semantic relatedness as a within-participants variable and causal relatedness and L2 reading proficiency as between-participants variables. The results revealed the significant main effects of causal relatedness and semantic relatedness with higher ratings for sentences high in relatedness than those low in relatedness (see Table 6). More importantly, the interaction between causal and semantic relatedness was significant, indicating that causal relatedness effects were significant for sentences both high and low in semantic relatedness, $F(1, 100) = 92.14$, $p < .001$, $\eta_G^2 = .48$; $F(1, 100) = 168.38$, $p < .001$, $\eta_G^2 = .63$, but as expected, the effects were larger for those that were low in semantic relatedness than high in semantic relatedness ($\eta_G^2 = .63$ vs. .48; $M_{diff} =$ 1.95 [1.65, 2.25] vs. 1.15 [0.91, 1.39]). Similarly, semantic relatedness effects were significant for sentences both high and low in causal relatedness, $F(1, 47) = 34.07$, $p < .001$, $\eta_G^2 = .25$; $F(1, 53) = 181.49$, $p < .001$, $\eta_G^2 = .61$, but were more prominent for those low in causal relatedness than high in causal relatedness ($\eta_G^2 = .61$ vs. .25; $M_{diff} = 1.55$ [1.31, 1.79] vs. 0.75 [0.45, 1.05]). These data replicate the interaction effect found in Experiment 1, regardless of the different experimental design.

Table 6. *Summary table for three-way ANOVA of the effects of L2 reading proficiency, causal relatedness, and semantic relatedness on coherence ratings for Experiment 2*

Source	*SS*	*df*	*MS*	*F*	*p*	η_G^2
			Between-participants			
Proficiency (P)	0.17	1	0.17	0.35	.557	.00
Causal Relatedness (C)	117.88	1	117.88	240.60	< .001	.57
P × C	3.52	1	3.52	7.18	.009	.04
Error (P ×C)	48.99	100	0.49			
			Within-participants			
Semantic Relatedness (S)	72.09	1	72.09	172.41	< .001	.44
S × P	2.86	1	2.86	6.84	.010	.03
S × C	6.59	1	6.59	15.77	< .001	.07
S × C × P	1.51	1	1.51	3.62	.059	.02
Error (S × C × P)	41.81	100	0.42			

Moreover, there were other two significant interaction effects: One was between causal relatedness and L2 reading proficiency, and the other was between semantic relatedness and L2 reading proficiency. Post-hoc analysis for the interaction of causal relatedness and L2 reading proficiency indicated that causal relatedness significantly affected coherence judgments made by both higher (CR-High: $M = 4.32$ [4.15, 4.50], $SD = 0.68$; CR-Low: $M = 2.54$ [2.28, 2.79], $SD = 0.98$) and lower (CR-High: $M = 4.00$ [3.69, 4.31], $SD = 1.01$; CR-Low: $M = 2.74$ [2.47, 3.01], $SD = 1.03$) proficiency groups, $F(1, 55) = 191.56$, $p < .001$, $\eta_G^2 = .67$; $F(1, 45) = 36.44$, $p < .001$, $\eta_G^2 = .45$, but the effects were greater for the higher proficiency group ($\eta_G^2 = .67$ vs. .45; $M_{diff} = 1.78$ [1.60, 1.96] vs. 1.26 [1.05, 1.47]). The simple main effect of proficiency was only significant when causal relatedness was high, $F(1, 47) = 4.51$, $p = .039$, $\eta_G^2 = .05$, with higher ratings for the higher proficiency group than the lower proficiency group ($M_{diff} = 0.32$ [0.12, 0.53]). These results indicate that the increase of ratings caused by the high causal relatedness was greater for the higher proficiency group than the lower proficiency group, suggesting that causal relatedness more strongly affected higher proficiency readers' judgments than lower proficiency readers' judgments.

The opposite pattern to these results was found for the interaction of semantic relatedness and L2 reading proficiency. Post-hoc analysis found that semantic relatedness significantly influenced the judgments made by both higher (SR-High: $M = 3.91$ [3.68, 4.15], $SD = 0.89$; SR-Low: $M = 2.97$ [2.62, 3.32], $SD = 1.34$) and lower (SR-High: $M = 4.00$ [3.78, 4.23], $SD = 0.78$; SR-Low: $M = 2.55$ [2.26, 2.84], $SD = 1.08$) proficiency groups, $F(1, 55) = 72.38$, $p < .001$, $\eta_G^2 = .36$; $F(1, 45) = 95.16$, $p < .001$, $\eta_G^2 = .51$, but the effects were relatively greater for the lower proficiency group ($\eta_G^2 = .51$ vs. .36; $M_{diff} = 1.45$ [1.06, 1.83] vs. 0.95 [0.53, 1.37]). The simple main effect of proficiency was only significant when semantic relatedness was low, $F(1, 100) = 4.14$, $p = .045$, $\eta_G^2 = .04$, with relatively higher ratings for the higher proficiency group than the lower proficiency group ($M_{diff} = 0.42$ [-0.06, 0.89]). These results demonstrate that the decrease of ratings caused by low semantic relatedness was greater for the lower proficiency group than the higher proficiency group, indicating that semantic relatedness more greatly influenced coherence judgments made by lower proficiency readers than those made by higher proficiency readers.

The three-way interaction between causal relatedness, semantic relatedness, and proficiency approached significance. Descriptive statistics implied that the interaction effect between

proficiency and causal relatedness was more prominent in the SR-Low condition, while the interaction effect of proficiency and semantic relatedness was more remarkable in the CR-High condition. Nevertheless, more research is needed to draw a definitive conclusion on whether this three-way interaction exists.

In sum, Experiment 2 partially replicated the results of Experiment 1. It confirmed the interaction effect of causal and semantic relatedness on coherence judgments. As in Experiment 1, causal relatedness had greater effects on judgments of less semantically related sentences, whereas semantic relatedness had greater effects on judgments of less causally related sentences. Therefore, it is clear that this interaction effect exists on coherence judgments made by L2 readers.

Furthermore, Experiment 2 found that L2 reading proficiency intervened in the effects of causal and semantic relatedness individually; that is, causal relatedness had greater impact on judgments made by higher proficiency readers, whereas semantic relatedness had stronger effects on those made by lower proficiency readers. These interactions of each type of relatedness and proficiency were found clearly only in Experiment 2, where participants read sentences that were consistently either low or high in causal relatedness. Given this, the effects of causal relatedness might be significant enough to overwhelm the effects of semantic relatedness and proficiency, as mentioned in the introduction of this experiment.

General Discussion and Conclusions

The aim of this study was to explore standards of coherence that L2 readers (Japanese readers of English) employ while reading. A pair of experiments examined the effects of two types of connectivity between sentences (i.e., causal and semantic relatedness) on text coherence judgments made by L2 readers with different reading proficiency levels. The obtained results provided answers for the two RQs addressed in this study.

Regarding RQ1 (the effects of causal and semantic relatedness), L2 readers' perceptions of text coherence were affected by both causal and semantic relatedness between sentences in the same way as L1 readers (Todaro et al., 2010). Sentence pairs were perceived as more coherent when causal or semantic relatedness was high. Further, causal and semantic relatedness had an interaction effect such that the effects of causal relatedness were stronger when sentences were low in semantic relatedness, whereas the effects of semantic relatedness were more prominent when sentences were low in causal relatedness. Because these results were obtained in both Experiments 1 and 2, it is strongly suggested that L2 readers base their standards of coherence on both causal and semantic relatedness between sentences. The robust impact of causal relatedness strongly supports the view that L2 readers are sensitive to causal coherence of two-sentence text as was suggested by previous studies using recall and self-paced reading tasks (Horiba, 1996b; Shimizu, 2009). The specific contribution of the present study is the finding that semantic relatedness is also a factor that affects L2 readers' perceptions of text coherence, and that it has an interaction effect with causal relatedness. However, the effects of semantic relatedness may be overwhelmed by those of causal relatedness, to some extent.

As for RQ2 (the effects of L2 reading proficiency), L2 readers with different proficiency levels were likely to adopt different standards of coherence in some conditions. Experiment 1 implied the differential effects of semantic relatedness across L2 reading proficiency. The following experiment, Experiment 2, further revealed that when reading sentences consistent in causal relatedness, lower proficiency readers weighed semantic relatedness in their coherence judgments to a greater extent than higher proficiency readers. This suggests that lower proficiency L2 readers base their standards of coherence more on semantic relatedness compared to higher proficiency L2 readers. On the other hand, under the same circumstances, higher proficiency readers weighed causal relatedness in their judgments more heavily than do lower proficiency readers. Such different patterns in the results of higher and lower proficiency readers might be explained in a similar way as in L1 research: Less proficient readers have a standard of coherence that emphasizes semantic relatedness because such readers tend to build a surface- or textbase-level representation, where only the explicit meaning of the text is represented. In contrast, proficient readers have a standard of coherence that emphasizes causal relatedness because such readers often build a text representation at the situation-model level, where causality is inferred and preserved by the readers (Todaro et al., 2010; Wittwer & Ihme, 2014). Nevertheless, it should be noted again that these interactions of each type of relatedness and proficiency are likely to be prominent in certain conditions (e.g., when reading sentences consistent in causal relatedness).

In conclusion, the present experiments demonstrated that L2 readers' perceptions of text coherence generally vary according to causal and semantic relatedness between sentences that are intentionally manipulated. Thus, L2 readers base their standards of coherence on causal and semantic relatedness between sentences. However, the results also revealed that not all L2 readers perceive the same texts to be similarly coherent in some conditions. It is possible that there exist individual differences in the perception of text coherence, such that less proficient L2 readers emphasize semantic relatedness in their perception, while more proficient readers emphasize causal relatedness in their perception. Hence, readers with different L2 reading skill levels might adopt different standards of coherence.

Implications

The resulting insight from this study has theoretical, methodological, and educational implications. From a theoretical perspective, this study suggests that the basic notion of standards of coherence, which has been proposed in L1 reading research (van den Broek et al., 2011; van den Broek et al., 2005; van den Broek et al., 1995), can be applicable to L2 reading. The present results demonstrated that L2 readers do employ standards of causal and semantic coherence while reading texts, and that these standards may vary between individuals. These types of standards and their variation patterns are very similar to those suggested in studies with L1 readers. Thus, standards of coherence might not be directly related to either L1 or L2 reading, but simply reading proficiency in general. Because standards of coherence are a central concept of the landscape model, the model may be able to describe the processes of establishing coherence in L2 as well as L1 text comprehension. It will be interesting to examine other aspects of the landscape model in the L2 context in the future.

From a methodological perspective, this study computed semantic relatedness using LSA. This technique was used in this study to assess the semantic similarities between sentences, and the results clearly indicated that LSA similarities affect the perception of text coherence by L2 readers of English. This supports the view that LSA similarities between sentences can be used as a metric for the readability of L2 texts (Crossley et al., 2012; Crossley et al., 2007; Crossley & McNamara, 2008; Nahatame, 2012); more specifically, LSA similarities may be indicative of the perceived coherence of the L2 text. Although most previous L2 studies have focused on the causal coherence of the text and learners' comprehension, it is recommended that future studies also focus on semantic coherence captured by LSA.

With regard to educational implications, L2 teachers should recognize the possibility that not all L2 learners perceive the same text to be similarly coherent. As the present findings suggest, L2 readers with lower reading proficiency, compared to those with higher reading proficiency, are likely to perceive coherence based on a relative surface level of semantic relatedness. Thus, although the semantic relatedness between sentences is one of the elements comprising coherence, L2 readers with lower reading proficiency may benefit from educational intervention focusing their attention on causal relatedness. Such intervention may help them achieve a deeper level of text comprehension.

Limitations and Avenues for Future Research

This study was only the first step in directly examining standards of coherence in L2 reading. Therefore, it can be developed and extended in several ways. Some of the possible directions for future research are noted below along with the limitations of this study.

First, although this study revealed some properties of standards of coherence employed by L2 readers, a similar examination should be conducted with L2 readers who have different profiles from those in this study. Specifically, given that this study was conducted with Japanese university students with beginning- to intermediate-levels English proficiency, it is worth conducting a similar study with L2 readers of English who are more proficient (e.g., advanced-level university students). Such a study will help us gain a deeper understanding of how L2 (reading) proficiency is related to standards of coherence.

Second, the present study focused on causal and semantic relatedness between sentences because prior studies suggest that these types of relatedness contribute to cohesiveness of a text and coherence of mental representations. Nevertheless, note again that there are other types of coherence, such as referential, spatial, and temporal coherence. Van den Broek et al. (2011) suggest that standards of referential and causal coherence are likely to be stringently adopted, while standards of spatial coherence tend to be less stringently adopted. Thus, it would be interesting to examine other types of coherence aside from causal and semantic coherence, and compare how stringently these types of standards of coherence are adopted in L2 reading.

Third, the present study examined the results of coherence judgments but not the judgment processes. Todaro et al. (2010) combined their data for the final judgment ratings with judgment times and sentence reading times to provide a better picture of causal and semantic relatedness

effects on comprehension processes. Additionally, they also examined whether the use of semantic relatedness on coherence judgments depends on controlled or automatic processes by providing instructions that emphasized causal relatedness or by limiting the text presentation time and coherence judgment times. In future studies, it would be informative to examine such processes and the automaticity of coherence judgments by L2 readers, as in Todaro et al.

Fourth, as the first step in exploring standards of coherence employed during L2 reading, this study used experimental passages consisting of two sentences that were easily comprehensible. Future studies may explore standards of coherence employed during reading of a whole, longer L2 passage. Such studies will allow us to assess whether the present findings are generalizable to longer texts.

Finally, it is important to investigate whether and how causal and semantic relatedness affect the actual processing and memory of L2 texts. Consider that participants in the present study were engaged in a task where they were asked to make conscious judgments on text coherence. Thus, it is unclear how the effects of causal and semantic relatedness (especially the latter) may act on an L2 reader's comprehension under more natural reading situations where such judgments are not required. Addressing this issue will extend the literature on cognitive processes involved in L2 reading.

Acknowledgments

This research was supported by Grants-in-Aid for Research Activity Start-up (15H06571) of the Japan Society for the Promotion of Science. I am grateful to Dr. Michael B. W. Wolfe for providing the experimental materials for this study. I would also like to express my sincere appreciation to the anonymous reviewers for their valuable and insightful comments on an earlier version of this manuscript.

Notes

1. Keenan et al. (1984) and Myers et al. (1987) indicated that sentences of intermediate levels of relatedness were better recalled than more and less related sentences. They suggested that readers needed to process these sentences with greater effort to infer the relationship between the sentences, resulting in better memory.

2. For more information about Eiken, see http://www.eiken.or.jp/eiken/en/

3. LSA was performed using the Sentence Comparison program from the University of Colorado website (http://lsa.colorado.edu). All LSA cosines were computed based on the General Reading up to First Year College corpus with 300 dimensions. The protagonist's name in the prime sentence and the pronominal reference in the target sentence were omitted from the analysis.

References

Committee of Revising the JACET Basic Words (Ed.). (2003). *JACET list of 8,000 basic words*. Tokyo, Japan: Author.

Crossley, S. A., Allen, D., & McNamara, D. S. (2012). Text simplification and comprehensible input: A case for an intuitive approach. *Language Teaching Research*, *16*, 89–108. doi: 10.1177/1362168811423456

Crossley, S. A., Louwerse, M., McCarthy, P. M., & McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *The Modern Language Journal*, *91*, 15–30. doi: 10.1111/j.1540-4781.2007.00507.x

Crossley, S. A., & McNamara, D. S. (2008). Assessing second language reading texts at the intermediate level: An approximate replication of Crossley, Louwerse, McCarthy, and McNamara (2007). *Language Teaching*, *41*, 229–409. doi: 10.1017/S0261444808005077

Ferstl, E. C., & von Cramon, D. Y. (2001). The role of coherence and cohesion in text comprehension: An event-related fMRI study. *Cognition Brain Research*, *11*, 325–340. doi: 10.1016/S0926-6410(01)00007-6

Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes*, *25*, 285–307. doi: 10.1080/01638539809545029

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, *3*, 371–398. doi: 10.1111/j.1756-8765.2010.01081.x

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*, 193–202. doi: 10.3758/BF03195564

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

Horiba, Y. (1996a). Comprehension processes in L2 reading: Language competence, textual coherence, and inferences. *Studies in Second Language Acquisition*, *18*, 433–473. doi:10.1017/S0272263100015370

Horiba, Y. (1996b). The role of elaborations in L2 comprehension: The effect of encoding task on recall. *The Modern Language Journal*, *80*, 151–164. doi: 10.1111/j.1540-4781.1996.tb01155.x

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

Horiba, Y. (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.

Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effect of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, *23*, 115–126. doi: 10.1016/S0022-5371(84)90082-3

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394. doi:10.1037/0033-295X.85.5.363

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis.

Discourse Processes, *25*, 259–284. doi: 10.1080/01638539809545028
Lee, S. K. (2009). Topic congruence and topic interest: How do they affect second language reading comprehension? *Reading in a Foreign Language*, *21*, 159–178.
Li, M., & D'Angelo, N. (2015). Higher-level process in second language reading comprehension. In X. Chen, V. Dronjic, R. H. Park (Eds.), *Reading in a second language: Cognitive and psycholinguistic issues* (pp. 158–194). London, UK: Routledge.
McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. New York: Cambridge University Press.
McNamara, D. S., & Magliano, J. P. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 51, pp. 297–384). New York: Elsevier Science.
Morishima, Y. (2013). Allocation of limited cognitive resources during text comprehension in a second language. *Discourse Processes*, *50*, 577–597. doi: 10.1080/0163853 X.2013.846964.
Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, *26*, 453–465. doi: 10.1016/0749-596X(87)90101-X
Nahatame, S. (2012). Latent semantic analysis (LSA) ni yoru kusho hoju gata dokkai test no kaimei: Bun level no imiteki kanrendo wo kanten to shite. [Investigation of a gap-filling test using latent semantic analysis (LSA): Focusing on semantic relatedness between sentences]. *STEP (the Society for Testing English Proficiency) Bulletin*, *24*, 42–58.
Nahatame, S. (2014). Strategic processing and predictive inference generation in L2 reading. *Reading in a Foreign Language*, *26*, 54–77.
Shimizu, H. (2009). The effects of causal relatedness on EFL learners' reading comprehension and inference generation. *ARELE (Annual Review of English Language Education in Japan)*, *20*, 31–40.
Todaro, S., Millis, K., & Dandotkar, S. (2010). The impact of semantic and causal relatedness and reading skill on standards of coherence. *Discourse Processes*, *47*, 421–446. doi: 0.1080/01638530903253825
Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*, 612–630. doi: 10.1016/0749-596X(85)90049-X
Traxler, M. J. (2011). *Introduction to psycholinguistics: Understanding language science*. Hoboken, NJ: Wiley-Blackwell.
Ushiro, Y. (2010). *Flexibility of updating situation models: Schema modification processes of Japanese EFL readers*. Tokyo, Japan: Kairyudou.
van den Broek, P. (1988). The effects of causal relations and hierarchical position on the importance of story statements. *Journal of Memory and Language*, *27*, 1–22. doi: 10.1016/0749-596X(88)90045-9
van den Broek, P., Bohn-Gettler, C., Kendeou, P., Carlson, S., & White, M. J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 123–140). Greenwich, CT: Information Age Publishing.
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. doi: 10.1080/0163853X.2005.9651685
van den Broek, P., Risden, K., & Husebye-Hartmann, E. (1995). The role of readers' standards

for coherence in the generation of inferences during reading. In R. F. Lorch, Jr., & E. J. O'Brien (Eds.), *Sources of coherence in text comprehension* (pp. 353–373). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Warren, P. (2012). *Introducing psycholinguistics*. New York: Cambridge University Press.

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. doi: 10.1080/0163853X.2013.855577

Wolfe, M. B. W., Magliano, J. P., & Larsen, B. (2005). Causal and semantic relatedness in discourse understanding and representation. *Discourse Processes*, *39*, 165–187. doi: 10.1080/0163853X.2005.9651678

Yoshida, M. (2012). The interplay of processing task, text type, and proficiency in L2 reading. *Reading in a Foreign Language*, *24*, 1–29.

Appendix

Experimental passages used in this study

1.	Mary could not find anything to read in the library.	(CR-High / SR-High)
	Mary wanted to look for recipes for her dinner party.	(CR-High / SR-Low)
	Mary went to the library to look for something to read.	(CR-Low / SR-High)
	Mary was having a dinner party for her office.	(CR-Low / SR-Low)
	She went to the bookstore to get new books.	(Target sentence)
2.	Frank lost control of his new sports car.	(CR-High / SR-High)
	Frank was looking at the screen of his new cell phone.	(CR-High / SR-Low)
	Frank bought a new sports car.	(CR-Low / SR-High)
	Frank bought a new cell phone.	(CR-Low / SR-Low)
	He got into an accident with a truck.	(Target sentence)
3.	Mary ate some food that had a spoiled taste.	(CR-High / SR-High)
	Mary was nervous about giving a speech in class.	(CR-High / SR-Low)
	Mary wanted to try the food at the new restaurant.	(CR-Low / SR-High)
	Mary went to school today.	(CR-Low / SR-Low)
	She felt sick to her stomach.	(Target sentence)
4.	Frank found typing errors in his memo.	(CR-High / SR-High)
	Frank's coffee was very cold this morning.	(CR-High / SR-Low)
	Frank was dictating an important memo.	(CR-Low / SR-High)
	Frank asked for a cup of hot coffee.	(CR-Low / SR-Low)
	He yelled at his secretary.	(Target sentence)
5.	Mary found that her baby was very sick.	(CR-High / SR-High)
	Mary's husband fell to the floor.	(CR-High / SR-Low)
	Mary recently became a nurse.	(CR-Low / SR-High)
	Mary's husband was working in his room.	(CR-Low / SR-Low)
	She called her family doctor at once.	(Target sentence)
6.	Frank kicked the living room wall, angrily.	(CR-High / SR-High)

	Frank felt a large earthquake begin.	(CR-High / SR-Low)
	Frank and Mary talked about the carpet in the living room.	(CR-Low / SR-High)
	Frank noticed that the neighborhood was very quiet.	(CR-Low / SR-Low)
	A mirror fell down and shattered on the floor.	(Target sentence)
7.	Mary was very tired at school because she stayed up all night.	(CR-High / SR-High)
	Mary did not go home until 4:00 a.m.	(CR-High / SR-Low)
	Mary had to finish her homework assignment by the following week.	(CR-Low / SR-High)
	Mary met a handsome, tall guy at the party.	(CR-Low / SR-Low)
	She fell asleep during class.	(Target sentence)
8.	Frank thought Mary would be a great wife for him.	(CR-High / SR-High)
	Frank knew that Mary had a lot of money.	(CR-High / SR-Low)
	Frank met Mary at a divorce support group.	(CR-Low / SR-High)
	Frank met Mary at the bank.	(CR-Low / SR-Low)
	He decided to propose marriage to Mary.	(Target sentence)
9.	Mary's fingers slipped while she was slicing the steak.	(CR-High / SR-High)
	Mary did not know the top of the fence was spiked.	(CR-High / SR-Low)
	Mary sliced the steak with a sharp knife.	(CR-Low / SR-High)
	Mary erected a tall fence in the back yard.	(CR-Low / SR-Low)
	She cut her hand badly.	(Target sentence)
10.	Frank shot with the gun at the police cars.	(CR-High / SR-High)
	Frank drank too much before driving home.	(CR-High / SR-Low)
	Frank was very upset when he saw the police cars.	(CR-Low / SR-High)
	Frank went to a party with some friends.	(CR-Low / SR-Low)
	He was arrested and put in jail.	(Target sentence)
11.	Mary received a package meant for the previous resident of her home.	(CR-High / SR-High)
	Mary received the wrong CD from the music store.	(CR-High / SR-Low)
	Mary checked to see if the mail had been delivered.	(CR-Low / SR-High)
	Mary had ordered a CD from the music store.	(CR-Low / SR-Low)
	She put the envelope back into the mailbox.	(Target sentence)
12.	Frank remembered that it was his wedding anniversary.	(CR-High / SR-High)
	Frank was two hours late on his way home.	(CR-High / SR-Low)
	Frank was weeding the garden in the back yard.	(CR-Low / SR-High)
	Frank was preparing for a presentation at work.	(CR-Low / SR-Low)
	He decided to buy his wife some flowers.	(Target sentence)
13.	Mary didn't take anything to wear in the pool.	(CR-High / SR-High)
	Mary lost her luggage at the international airport.	(CR-High / SR-Low)
	Mary was invited to swim with her friends.	(CR-Low / SR-High)
	Mary went to Alaska during her spring vacation.	(CR-Low / SR-Low)
	She borrowed a swimming costume at the gym.	(Target sentence)
14.	The cashier could not give change for Frank's hundred-dollar bill.	(CR-High / SR-High)
	Frank's girlfriend insisted on treating him to a movie.	(CR-High / SR-Low)
	Frank went to pay his bill.	(CR-Low / SR-High)

	Frank arrived at the ticket counter with his girlfriend.	(CR-Low / SR-Low)
	He put his money back into his wallet.	(Target sentence)
15.	Mary was told that people had found a shark in the sea.	(CR-High / SR-High)
	Mary ate too much at lunch with her friends.	(CR-High / SR-Low)
	Mary arrived at the beach in her swimming costume.	(CR-Low / SR-High)
	Mary sat on a chair at the cafeteria.	(CR-Low / SR-Low)
	She decided to wait before going for a swim.	(Target sentence)
16.	Frank had to pass his last exam to graduate.	(CR-High / SR-High)
	Frank wanted to get a black belt in karate.	(CR-High / SR-Low)
	Frank majored in physics at his university.	(CR-Low / SR-High)
	Frank had a black belt in karate.	(CR-Low / SR-Low)
	He began to study for his final test.	(Target sentence)
17.	Mary wanted to have a special dinner.	(CR-High / SR-High)
	Mary passed her medical license exam.	(CR-High / SR-Low)
	Mary loved to cook elaborate meals.	(CR-Low / SR-High)
	Mary was studying to be a doctor.	(CR-Low / SR-Low)
	She went to a fancy restaurant.	(Target sentence)
18.	Frank collapsed while he was seeing patients in his office.	(CR-High / SR-High)
	Frank was hit on the head with a hard baseball.	(CR-High / SR-Low)
	Frank was a doctor in a big city.	(CR-Low / SR-High)
	Frank had a great seat at the baseball game.	(CR-Low / SR-Low)
	He was carried to a hospital in an unconscious state.	(Target sentence)
19.	Mary turned on her table fan to cool the room.	(CR-High / SR-High)
	Mary let out a very loud sneeze.	(CR-High / SR-Low)
	It was hot in the room where Mary was working on her homework.	(CR-Low / SR-High)
	Mary had allergy attacks during the summer.	(CR-Low / SR-Low)
	The papers on her desk blew onto the floor.	(Target sentence)
20.	Frank's eyes were sensitive to strong light.	(CR-High / SR-High)
	Frank wanted to create a romantic mood.	(CR-High / SR-Low)
	Frank went into the well-lit bedroom.	(CR-Low / SR-High)
	Frank remembered that it was his wedding anniversary today.	(CR-Low / SR-Low)
	He turned the lamp down low.	(Target sentence)

About the Author

Shingo Nahatame, Ph.D., is a lecturer in the Department of Education at Kyoei University, Japan. His research interests include the cognitive processes of second language reading, especially discourse processing and understanding. E-mail: nahatame@kyoei.ac.jp. Website: https://sites.google.com/site/snahatame/