Concreteness Effect on L2 Learners' Lexical Inferencing

Yuko MORIMOTO
Graduate School, University of Tsukuba

Abstract

A lot of studies have been investigating factors influencing the success rate of lexical inferencing, but concreteness factor has been neglected. The aim of this study is to examine effects of concreteness on lexical inferencing. Four types of test were made to infer the meaning of: (a) concrete unknown words in concrete sentences, (b) abstract unknown words in concrete sentences, (c) concrete unknown words in abstract sentences, and (d) abstract unknown words in abstract sentences. The result revealed that concreteness of words did not influence the success rate of lexical inferencing, whereas the concrete context raised the success rate significantly more than the abstract context. Moreover, the result of questionnaire indicated that participants felt lexical inferencing from concrete sentences was easier than that from abstract sentences. These results imply that images created from contexts helped EFL students infer meanings of words. Therefore, a concrete rather than an abstract context should be used when we teach how to infer unknown words, so that students become more confident.

1. Introduction

Lexical inferencing helps increase learners’ vocabulary knowledge. Jenkins, Matlock and Slocum (1989) claimed that two major ways to expand learners’ vocabulary were to teach meanings directly and to teach how to derive meanings of words. A number of studies have focused on lexical inferencing, although there is still room for investigating what types of word are easier to infer, and from what types of contexts learners can infer unknown words better. The current study shed light on these two issues from the perspective of concreteness.

2. Literature Review

2.1 Inferring Unknown Words

A large number of studies have investigated what factors help learners derive meanings of unknown words. In one of these studies, Fukkink (2005) presented two crucial factors: the context and the complexity of the word. As to the context, for instance, the results of Huckin and Bloch (1993) revealed that learners were more likely to derive correct meanings of unknown words when they used contextual clues within the same sentence as the unknown word; but if they used clues beyond the sentence, it became more difficult. Also, Mondria and
Wit-de Boer (1991) revealed that when a sentence contained a pregnant subject, verb, and function in lexical inferencing, the guessability of unknown words increased.

Concerning the complexity of the word, research can be divided into two types: studies about the form and the content of words. In terms of the form, Fraser (1999) found that learners were likely to recall known words with similar form to unknown words, and Qian (2004) pointed out that his participants used their knowledge of affixes in lexical inferencing. As to the content of words, Fukkink (2005) suggested that false attributes, which are not listed in dictionary definitions, were generated significantly more in deriving abstract words than in concrete words, which indicates that the quality of students’ definitions was better in concrete words. He claimed that abstract words were “conceptually challenging”; therefore his participants, who were primary-school students, had difficulty in inferring the meaning of abstract words. However, Fukkink carried out the research with L1 speakers; therefore how concreteness influences L2 lexical inferencing have not yet been clarified.

2.2 Concreteness

Concrete words are not only less conceptually challenging (Fukkink, 2005) but also have advantages in imageability. Paivio’s (1986) dual coding theory assumes that there are an imagery system and a verbal system in information processing, and “either system can be active without the other or both can be active in parallel” (p. 54). In processing concrete words or sentences, both systems are more likely to be activated because concreteness and imageability have a high correlation such as over $r = .90$ (de Groot & Keijzer, 2000); whereas in processing abstract words or sentences, the imagery system is less activated. Since concrete terms make both systems active and abstract terms are likely to activate only one system, much research has found advantages in concrete terms. For example, Jessen et al. (2004) reported that people processed concrete words faster than abstract words, and de Groot and Keijzer claimed that concrete sentences were better recalled than abstract sentences.

In the study of lexical inferencing, Mondria and Wit-de Boer (1991) used only concrete nouns as target words because they argued that “with this type of word the meaning can be conveyed relatively easily by means of a word-by-word translation” (p.255). However, it is ambiguous whether concrete words are easier to be inferred indeed, because Mondria and Wit-de Boer did not give a theoretical explanation for why concrete words are inferred more easily. Moreover, because neither the taxonomies of lexical inferencing by Haasstrup (1991) nor Bengeleil and Paribakht (2004) contain any factor of concreteness, the effect of concreteness on lexical inferencing is still unclear. In addition, the concreteness of word as well as that of context might affect the success rate of lexical inferencing, even though no studies have investigated it; therefore this study deals with both kinds of concreteness.

2.3 Purpose of the Study and Hypotheses
The purpose of the current study was to investigate how concreteness affected the success rate of lexical inferencing. In order to achieve this aim, two hypotheses (Hs) were set up based on the dual-coding theory, which indicates the advantage of concrete terms.

H1: Concrete unknown words are correctly inferred more than abstract unknown words.
H2: Unknown words in concrete sentences are correctly inferred more than those in abstract sentences.

3. Method

3.1 Participants

The participants in this study were 117 first-year university students whose majors were either humanities, agro-biological resources, or nursing and medical technology.

3.2 Materials

Two tests were prepared: a reading proficiency test, and a lexical inferencing test. A reading section from a TOEFL practice test (ETS, 2002), which was composed of five passages with a total of 50 items, was used to measure participants’ reading proficiency.

As for the lexical inferencing test, four types of test were made to see whether participants could derive meanings of: (a) concrete unknown words in concrete sentences (Type A); (b) abstract unknown words in concrete sentences (Type B); (c) concrete unknown words in abstract sentences (Type C); and (d) abstract unknown words in abstract sentences (Type D). The concreteness of words was assessed using the MRC (Medical Research Council) Psycholinguistic Database (Clark, 1997), in which concreteness is rated with a 700 point-scale (100 is the least concrete, and 700 is the most concrete). In this study, words rated over 450 were regarded as concrete and those rated below 350 were abstract. However, high frequency words were excluded since the participants probably knew the meanings of these words. At the same time, too culturally-dependent words (e.g., anele) were also excluded because the success rate of inferencing these words was assumed to be highly affected by participants’ cultural background knowledge.

Next, one sentence per target word was prepared. Care was taken that all words except for the target words should be of as high a frequency as the first 3000 frequency words in JACET 8000 (2003), because if the proportion of unknown words is too high, it is difficult to derive word meanings (Schmitt, 2000, p. 120). According to Kasahara (2005), at least the 3000 most frequent words counted by lemmas were known for the first-year university students. Therefore, it was assumed that participants in this study were also familiar with these words.

After preparing contexts, a total of 12 graduate students were asked to judge the
concreteness of these sentences independently, following Ellis and Beaton (1993). The rating instruction used in this study was the Japanese translation of the one used in Paivio, Yuille and Madigan (1968). They used the instruction for rating word-concreteness, whereas this study was for rating sentence-concreteness; hence, some parts of the original instruction were changed as necessary. After that, sentences with the target words replaced by blanks were presented to an English native speaker who was asked to fill in each blank with possible answers in order to confirm whether the guessability of the target sentences was the same. Consequently, 15 sentences per test type were selected to be used in this study. The average word-concreteness, sentence-concreteness, and the range of number of words in a sentence and the readability are shown in Table 1. Basically the unknown words being used across four types of test differed, but two concrete words were commonly used in Types A and C, and three abstract words were commonly used in Types B and D in order to analyze the effect of concreteness more precisely (see section 4.3). Following is the examples of each type of test: *It is very challenging to climb up such a steep crag* (Type A); *The passengers were surprised with an abrupt stop of the train* (Type B); *Marriage is not simply sharing the same name and abode* (Type C); and *Most people in the world abhor any kind of violence and war* (Type D).

<table>
<thead>
<tr>
<th>Type</th>
<th>Concreteness of contexts</th>
<th>Concreteness of unknown words</th>
<th>The range of number of words</th>
<th>The range of readability (Flesch-Kincaid)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Concrete (4.7-6.5)</td>
<td>Concrete (476-597)</td>
<td>10-23</td>
<td>4.8-9.9</td>
</tr>
<tr>
<td>B</td>
<td>Concrete (4.5-6.5)</td>
<td>Abstract (240-340)</td>
<td>11-19</td>
<td>4.8-11.0</td>
</tr>
<tr>
<td>C</td>
<td>Abstract (1.0-3.8)</td>
<td>Concrete (453-611)</td>
<td>10-17</td>
<td>3.8-9.9</td>
</tr>
<tr>
<td>D</td>
<td>Abstract (2.3-3.3)</td>
<td>Abstract (239-349)</td>
<td>9-18</td>
<td>6.7-10.5</td>
</tr>
</tbody>
</table>

*Note.* Concreteness of contexts was rated with seven-point scales (*one* was the most abstract and *seven* was the most concrete), and the concreteness of unknown words was rated with 700-point scales (100 was the most abstract and 700 was the most concrete).

3.3 Procedures
First, all participants took the reading proficiency test in 55 minutes, which ETS (2002) indicated as the time allotment. Then, participants randomly answered any two of the four test types of the lexical inference test, which took 30 minutes altogether. At the end of this test, participants rated how difficult each test type was on a five-point Likert scale.

3.4 Scoring
A correct answer was given one point in the reading proficiency test. The lexical inferencing test was scored from three perspectives: *Meaning, Word Class,* and *Context.* As for *Meaning,* if a participant’s answer was in the index of each target word in a dictionary
(Konishi & Minamide, 2002), the answer was given one point. In terms of Word Class, one point was given when the word class of a participant’s answer was the same as that of the target word. As to Context, when an answer fit into the context, one point was given even if it was not the meaning of the target word itself. For example, one of the target sentences is: *It is very challenging to climb up such a steep crag.* The target word is *crag*, which means “a high and very steep rough rock or mass of rocks,” even though some participants wrote *mountain* for the meaning of the *crag*. Their answers were not the correct meaning of *crag*, but the meaning of *mountain* fits into the context. Therefore, in this case, they obtained one point in Context. There were no partial points in any of the scorings. When a participant claimed he or she knew a target word, the item of the participant was excluded from analyses.

Two raters including the researcher scored the 88 answer sheets (22 sheets for each type of test), which were more than 35% for each test type. The inter-rater reliabilities, the Cronbach’s alphas, were .99 for Meaning, .99 for Word Class, and .95 for Context. Because high reliabilities were obtained, the researcher scored the rest of the answer sheets alone. In order to regulate the strictness of the raters, the final scores were the researcher’s.

4. Results and Discussions

The result of a one-way ANOVA for the reading proficiency test showed that there was no significant difference between the four groups (*F* (3, 229) = .95, *p* = .42). Therefore, these four groups were found to be equivalent in English reading ability.

4.1 Analysis by Success Rates of Inferencing

As revealed in the previous section (see section 3.4), there were three types of scoring for each participant’s answers in the lexical inferencing test: Meaning, Word Class, and Context. The descriptive statistics of each scoring are presented in Table 2. In Meaning and Context, Type B obtained the highest score, whereas in Word Class, Type A was the highest.

<table>
<thead>
<tr>
<th>Type</th>
<th>n</th>
<th>Meaning</th>
<th></th>
<th>Word Class</th>
<th></th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>58</td>
<td>5.72</td>
<td>1.86</td>
<td>13.78</td>
<td>1.64</td>
<td>11.10</td>
</tr>
<tr>
<td>B</td>
<td>60</td>
<td>6.52</td>
<td>2.00</td>
<td>11.20</td>
<td>2.05</td>
<td>11.90</td>
</tr>
<tr>
<td>C</td>
<td>57</td>
<td>4.26</td>
<td>1.61</td>
<td>12.46</td>
<td>2.00</td>
<td>10.44</td>
</tr>
<tr>
<td>D</td>
<td>58</td>
<td>3.62</td>
<td>1.79</td>
<td>13.17</td>
<td>1.87</td>
<td>10.67</td>
</tr>
</tbody>
</table>

*Note. A full mark of each scoring is 15.*

For each type of scoring, a two-way ANOVA (types of word × types of sentence) was

145
employed in order to see the effect of concreteness. Firstly, in terms of *Meaning*, the interaction was significant \( (F(1, 229) = 9.04, p < .00, \eta^2 = .004^{*1}) \). This result shows that the success rate of abstract unknown words were more likely to be affected by sentence types. Another finding was that the main effect of sentence types was significant \( (F(1, 229) = 83.35, p < .00, \eta^2 = .040) \), which indicated that participants retrieved correct meanings of unknown words from the concrete sentences more than from the abstract sentences. The main effect of word types was not significant \( (F(1, 229) = .33, n.s., \eta^2 = .000) \).

Secondly, the result of a two-way ANOVA in *Word Class* indicated that the interaction was again significant \( (F(1, 229) = 157.78, p < .00, \eta^2 = .004) \), which suggested that participants correctly inferred word classes of concrete words more in concrete sentences than in abstract sentences, but this difference disappeared in deriving abstract words. Apart from the interaction, the main effect of word types was significant \( (F(1, 229) = 13.96, p < .00, \eta^2 = .001) \), which means concrete words were more likely to be inferred in terms of word class. The main effect of sentence types was not significant \( (F(1, 229) = 1.72, n.s., \eta^2 = .000) \).

Thirdly, a two-way ANOVA was again conducted in *Context*. The interaction was not significant \( (F(1, 229) = 1.12, n.s., \eta^2 = .000) \). However, the main effect of sentence types was significant \( (F(1, 229) = 12.71, p < .00, \eta^2 = .002) \), which means that the participants derived meanings which fit into the context more in concrete sentences than in abstract sentences. The main effect of word types approached significance \( (F(1, 229) = 3.77, p < .10, \eta^2 = .001) \). However, we have to interpret this effect carefully because this result meant, contrary to the prediction, abstract words were inferred more correctly than concrete words.

### 4.2 Analysis by Different Kinds of Word Class

The number of word classes of target words in each test type was not controlled in this study, and it was possible that this discrepancy produced the difference in section 4.1. Moreover, further analyses were needed because the effect size \( (\eta^2) \) was small overall. The number of word class is described in Table 3, which shows there were more nouns in concrete sentences (Types A and C), and Type B had adjectives although other types had none.

<table>
<thead>
<tr>
<th>Table 3 The Numbers of Word Class Used in Each Test Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Type A</td>
</tr>
<tr>
<td>Type B</td>
</tr>
<tr>
<td>Type C</td>
</tr>
<tr>
<td>Type D</td>
</tr>
</tbody>
</table>

The result of the Mann-Whitney U test showed that there was no significant difference in success rates between any of the word classes within each test type. This means the type of
word class hardly affected participants’ performances in this study. Therefore, the result in section 4.1 was not caused by the effect of word class, but by the effect of concreteness.

4.3 Analysis by Common Words Across Test Types

In order to investigate the effect of concreteness of sentences in more detail, five common words were embedded in both contexts as written in section 3.2: two concrete words, *venom* and *derelict*, were used both in Types A and C, and three abstract words, *avowal*, *jeer*, and *teem*, were used in Types B and D. A pair example is shown below.

*The children in the park started running away at the sight of the big snakes with strong venom.* (A concrete sentence).

*The venom of this snake is nine times as powerful as that of ordinary snakes.* (An abstract sentence).

The result of t tests indicated that there were significant differences in four out of five words: *derelict* in Word Class (*p* < .01) and Context (*p* < .001); *avowal* in Meaning (*p* < .05); *jeer* in Meaning (*p* < .001); and *teem* in Meaning (*p* < .05) and Context (*p* < .05). This means that these words were correctly identified more in the concrete sentences than in the abstract sentences, and these four words had advantages at least either in Meaning or Context, both of which relate to meanings of the target words and contexts. Therefore, even if the target words were common across concrete and abstract sentences, participants still succeeded in lexical inferencing in concrete sentences more than in abstract sentences. This result reinforced the advantage of concrete sentences over abstract sentences in lexical inferencing, especially deriving exact meanings of unknown words or meanings that fit into the contexts.

4.4 Difficulty-Rating

As mentioned in section 3.3, each participant rated the difficulty of two types of test with a five-point Likert scale; *one* was the easiest and *five* was the most difficult. The descriptive statistics of difficulty are presented in Table 4.

<table>
<thead>
<tr>
<th>Type</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type A</td>
<td>57</td>
<td>4.14</td>
<td>.79</td>
<td>.10</td>
</tr>
<tr>
<td>Type B</td>
<td>60</td>
<td>3.75</td>
<td>.79</td>
<td>.10</td>
</tr>
<tr>
<td>Type C</td>
<td>57</td>
<td>4.26</td>
<td>.79</td>
<td>.10</td>
</tr>
<tr>
<td>Type D</td>
<td>57</td>
<td>4.51</td>
<td>.63</td>
<td>.08</td>
</tr>
</tbody>
</table>

*Note.* The number of participants differed from Table 1 since some participants missed rating difficulty.
The result of a one-way ANOVA showed there was a significant difference among the four types of test in difficulty \((F(3, 227) = 10.32, p < .00, \eta^2 = .120)\). Tukey HSD multiple comparisons were conducted and the result revealed that participants felt Type A was easier than Type D, and Type B was easier than Types A, C, and D. Considering that both Types A and B were made up with concrete sentences, learners perceived lexical inferencing from the concrete sentences to be easier than from the abstract sentences. However, there was no consistent tendency to indicate the advantage of either concrete or abstract unknown words.

4.5 Discussion

As a result of the analyses, no significant difference was found between the guessability of concrete and abstract words. Therefore, H1 (concrete unknown words are correctly inferred more than abstract unknown words) was rejected, which is inconsistent with Fukkink (2005). One of the biggest differences between his study and the current study was that Fukkink’s participants were elementary school students while participants in this study were university students. Adults are cognitively more developed; therefore they are mature enough to process concrete and abstract words similarly, whereas this is probably not the case for young children, since Fukkink suggested that abstract words were conceptually challenging. Another difference between the two studies is that Fukkink conducted his study with L1 but this study used L2. Therefore, concreteness of L1 unknown words might influence lexical inferencing as Fukkink discovered, while L2 target words do not, as in the result of this study. However, further studies are necessary to affirm which explanation is more appropriate.

On the other hand, the result showed that unknown words in concrete sentences were more likely to be inferred correctly than in abstract sentences. This result was obtained in terms of Meaning and Context, analyses by common words across test types, and even in difficulty-ratings by participants. Therefore, H2 (unknown words in concrete sentences are correctly inferred more than those in abstract sentences) was supported except for Word Class. Considering that both concrete sentences and abstract sentences had the same guessability (see section 3.2), this difference came not from a discrepancy in quality and quantity of linguistic clues, but from concreteness of sentences. Participants were more likely to visualize, or at least utilize the image being represented in concrete sentences, which helped them derive unknown words from context. However, the reason for non-significant effects on Word Class between concrete and abstract sentences was that Meaning and Context were directly related to the content of contexts, while Word Class was not. In other words, participants had to understand the meaning of the sentence to obtain correct answers in Meaning and Context, whereas in terms of Word Class it was possible to answer with only syntactic knowledge. Therefore, it can be said that the effect of sentence-concreteness is only valid when learners have to utilize to content of the sentence in lexical inferencing.
5. Limitations and Implications

There are three limitations in this study. First, this study only dealt with contexts consisting of one sentence. Although using a short context has the merit of restricting the factors affecting lexical inferencing, it does not represent authenticity of reading. Apparently further research is needed to examine the effect of concreteness with longer contexts. Second, the rating of sentence-concreteness was done by participants' intuition, not using any standards. Although past studies which dealt with concrete and abstract sentences (e.g., Ellis & Beaton, 1993) also did not use an objective standard, fixed criteria will be needed to pursue further studies. The third inadequacy is that although Barsalou (1989) suggested that people had different impressions of concepts, individual impressions about the concreteness of sentences were not included in this study. It is possible that the effect of concreteness would become clearer if future studies checked the concreteness of individuals' impressions.

A pedagogical implication is that it is better to use concrete rather than abstract contexts when we teach how to derive unknown words. This study revealed that lexical inferencing was performed better from concrete sentences than from abstract sentences; therefore, inferencing from concrete sentences increases the success rate, which will make learners more confident. However, it is presumably also important to allow learners to gradually become accustomed to abstract contexts because authentic texts are not always created from concrete contexts. In order to teach lexical inferencing effectively, it is necessary for educators to use concrete and abstract contexts properly.

Acknowledgement

I am grateful for three anonymous reviewers for their valuable comments. I also wish to thank Professor Yuji Ushiro and Professor Akiyo Hirai in University of Tsukuba for their innumerable advice to improve this paper.

Note

$\eta^2$ (eta squared: effect size) is calculated by $SS_{\text{effect}}$ divided by $SS_{\text{total}}$ where $SS_{\text{effect}}$ is the sum of squares for the effect and $SS_{\text{total}}$ is the total sum of squares (Kline, 2004, p. 180). $0.010 \leq \eta^2 \leq 0.059$; $0.059 \leq \eta^2 \leq 0.138$; $0.138 \leq \eta^2 \leq 0.418$ (large; Cohen, 1988, pp. 284-287). An effect size of 0.009 and below is regarded as having no effect.

References

In S. Vosniadou & A. Ortony (Eds.), Concepts and conceptual development: Ecological and intellectual factors in categorization (pp. 101-140). Cambridge University Press.


