Effects of Text Readability on Incidental Vocabulary Learning
Through Glosses

Kiwamu KASAHARA
Graduate School, University of Tsukuba

Abstract

The purpose of this paper is to investigate the effects of text readability on incidental vocabulary learning through glosses. Sixty-six university students participated in this experiment. Each participant read four English passages with different readability. Each passage had four target words with L1 (Japanese) marginal glosses. Afterwards, an unexpected vocabulary test on those target words was conducted. The main results were: (a) text readability affects incidental vocabulary learning through glosses and, (b) incidental vocabulary learning is unlikely to occur when the text readability is so high that the text demands too much cognitive load from learners. Some pedagogical implications are drawn from these findings.

1. Introduction

In EFL learning, learners enlarge their vocabulary unintentionally through various types of activity as in the case of L1. According to Hatch and Brown (1995), intentional vocabulary learning is “being designed, planned for, or intended by teacher or student,” while incidental learning is “the type of learning that is a by-product of doing or learning something else.” The latter plays a crucial role in increasing learners’ vocabulary. Of the various types of language activity, reading is thought to provide particularly abundant opportunities for incidental vocabulary learning (Pitts, White & Krashen, 1989; Day, Omura & Hiramatsu, 1991).

Incidental vocabulary learning through extensive reading is, however, often interrupted by the usual learner’s strategy of simply ignoring unknown words. Therefore, it is extremely important to let EFL learners “pick up” the meaning of unknown words when they encounter them. There are several ways to make this happen, such as guessing the meaning of unknown words from the context, consulting a dictionary, and referring to a gloss (Mitarai & Aizawa,
Among these approaches, several studies have shown the superiority of referring to a gloss, pointing out the defects of other approaches (Hulstijn, Hollander & Greidautus, 1996; Aizawa, 1998). Inferring the meaning of unknown words is often a difficult task when the contextual information is limited, and moreover, it can cause wrong guesses. Looking up in a dictionary is avoided in many cases because the readers do not want to discontinue reading.

In addition to these studies, there have been several on what kinds of gloss are effective for incidental vocabulary learning. Some examine their effective positions, some investigate which language (L1 or L2) is better for glosses, while others survey effective types of gloss (single or multiple-choice).

However, there does not seem to be any research on the relation between text readability and incidental vocabulary learning through glosses. The former might play an important role in incidental vocabulary learning because understanding a text is a necessary condition for incidental learning to happen. Therefore, this study examines this relationship, and tries to clarify what level of text is suitable for incidental vocabulary learning through glosses.

2. Literature review

As to the position of glosses, Holly and King (1971) investigated the effectiveness on incidental vocabulary learning of the three gloss positions, i.e. at the side of the page, at the bottom of the page, and in an attached list. The result shows no significant difference between them. There seems to be a general tendency among learners to prefer glosses in the side margin (Jacobs, Dufon & Hong, 1994). This is because they can refer to the glosses without breaking the stream of reading activity.

In terms of the language of glosses, several investigations have been conducted. Jacobs et al. (1994) and Myong (1995, cited in Nation, 2001) shows no significant difference between L1 and L2 glosses. Mitarai and Aizawa (1999) find the superiority of L1 glosses because “L1 meaning network is easier to construct than that of L2” (p. 79). Nation (2001) has concluded that the language of glosses makes little difference as long as the meaning is clear to the learners.

Among the studies on the type of gloss, several have considered whether single or multiple-choice (MC) glosses are more effective. In the latter, readers are given two or more synonyms or explanations for one unknown word, and choose one correct meaning, judging from the context. This type of gloss is devised on the basis of a ‘mental effort’ hypothesis, “which predicts that the retention of an inferred word meaning will be higher than the retention of a given word meaning” (Hulstijn, 1992). Hulstijn (1992) proves the superiority of MC glosses in three of his five experiments, while Watanabe (1997a) shows no significant difference between them. Some researchers (Nagayama & Mori, 2003; Watanabe, 1997b) have
shown that this discrepancy occurs because of the difference in the subjects’ English proficiency. According to their studies, learners with high English proficiency can take advantage of MC glosses but there is still a possibility that they may make a wrong guess.

3. The Present Study

The aim of this study is to investigate the effects of text readability on incidental vocabulary learning through glosses. This is based on the assumption that readability greatly affects understanding of the text, which is necessary to make incidental vocabulary learning happen. It is surprising that little is known about the relation between text readability and incidental vocabulary learning. Some research should be done to clarify this relation.

Since we are not concerned here with what type of gloss is effective, the meanings of the glosses adopted in this study should be as clear as possible to the participants. Taking the results of the previous literature above into consideration, L1 (Japanese) single glosses in the side margin are best. For instance:

*On Friday morning, I ordered two cases of scallops from your company.*

3.1 Research Questions

The following research questions were addressed in this study.

(1) Does text readability affect incidental vocabulary learning through glosses?
(2) Does incidental vocabulary learning through glosses occur only to a negligible degree when the text readability showing U.S. school-grade is so high that the text is beyond learners’ understanding?

3.2 Participants

Two regular English classes of first-year students at a national university took part in this experiment. One class consisted of 31 students who were studying Engineering systems; the other comprised 35 students majoring in Humanities. In order to check the participants’ reading ability, a reading comprehension test consisting of questions from a TOEFL reading section was conducted before carrying out the experiment. According to the results, the participants were divided into two groups. It was decided that 33 of them belonged to the high proficiency group while the other 33 belonged to the low proficiency group. The means and standard deviations in each group are shown in Table 1. There was a significant difference between the two groups in the result of the t test conducted on their scores of the reading comprehension test \((t = 12.01, p < 0.01)\).
### 3.3 Materials

Four reading materials with different readability were prepared. The readability of each text was measured by the Flesh-Kincaid Grade Level. The topics of these texts were all concerned with business matters in order to minimize different effects on learning by different topics. Texts A and B were taken from a reference book on TOEIC, Text C from a book on business letters, and Text D from a reference book on TOEFL. Priority in choosing materials was given to readability over length. Consequently, Texts C and D were longer than Texts A and B by about 40 words because of their high readability. Four target words, which the author thought would be unknown to ordinary freshmen at college, were selected from each text and an L1 single gloss as mentioned above was put in the right margin beside each target word. The number of total words, F-K Grade Level and target words are shown in Table 2.

### Table 2

<table>
<thead>
<tr>
<th>Text</th>
<th>Total words</th>
<th>F-K Grade Level</th>
<th>Target words</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>151</td>
<td>4.1</td>
<td>scallop, cod, halibut, mishap</td>
</tr>
<tr>
<td>B</td>
<td>148</td>
<td>6.1</td>
<td>Battered, surge, downgrade, brokerage</td>
</tr>
<tr>
<td>C</td>
<td>189</td>
<td>8.7</td>
<td>sloppy, implement, reship, gratis</td>
</tr>
<tr>
<td>D</td>
<td>188</td>
<td>12.0</td>
<td>unprecedented, staggering, perversion, unduly</td>
</tr>
</tbody>
</table>

### 3.4 Pilot Study

A pilot study was carried out to determine the length of time the participants would need to read each passage and to predict whether or not they would have any knowledge of the target words. Three sophomores at the same university read the four texts and took the reading comprehension and vocabulary tests. As a result, it was found that Tests C and D required more time to read than Texts A and B. Therefore, the idea of randomizing the order of texts to avoid the order effect was rejected, and it was decided that the participants would read the texts in the same order, A, B, C and D. It was also found that they had little knowledge of the target words.

### 3.5 Procedure
This study was conducted in two English classes during a regular 75-minute class session in the winter of 2004. Each student received a test booklet consisting of the four texts mentioned above. Each text had three comprehension questions. The students had to read the text and choose a correct answer from four given choices to each question. All the participants had the same amount of time to complete the task for each text. They were not allowed to go on to the next text until so instructed. A period of 21 minutes was allotted to this reading section.

After the explanations and correct answers were shown to the students for 15 minutes, the booklets were collected. Then, they were told to take an unexpected vocabulary test, which included 16 target words with glosses in the texts, and four dummy words that had not been in the texts. The subjects had to write the meaning of those words in Japanese in 10 minutes. If they already knew some of the words tested they had to put checkmarks in the parentheses next to those words. The words with checkmarks were excluded from the analysis.

3.6 Scoring

For the comprehension test, one point was given for each correct answer. Therefore, the highest score for each text was three. For the vocabulary test, two points were given for a fully correct answer and one point for a partly correct answer. For example, in the case of the word “cod”, the correct answer is “Tara” (鯖), while “some kind of fish” (魚の一一種) is a partly correct answer. Hence, the maximum for each text was eight. If a participant put a checkmark next to a word and the answer was correct, the word was excluded from the analysis.

3.7 Data Analysis

Since neither normal distribution of the data nor homogeneity of variance could be assumed, the nonparametric Friedman test was conducted to see if there was any difference in the four texts in terms of both the reading comprehension and vocabulary test within each group. A Wilcoxon signed-ranks test with a Bonferroni correction was employed for the multiple comparisons. In order to see the difference between the two groups in each test, a Wilcoxon rank sum test was applied.

3.8 Results

The means and standard deviations of the reading comprehension tests and the vocabulary tests are given in Tables 3 and 4 respectively. Table 5 shows the results of the Wilcoxon signed-ranks tests to show the differences between the texts in the two tests within each group. Table 6 reveals the differences between the groups in both the tests according to the result of the Wilcoxon rank sum test.
Table 3

Means and Standard Deviations of the Reading Comprehension Test Scores in Each Group (maximum = 3)

<table>
<thead>
<tr>
<th>Group</th>
<th>Text A</th>
<th>Text B</th>
<th>Text C</th>
<th>Text D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>High</td>
<td>2.91</td>
<td>0.29</td>
<td>1.88</td>
<td>0.74</td>
</tr>
<tr>
<td>Low</td>
<td>2.79</td>
<td>0.48</td>
<td>1.91</td>
<td>0.77</td>
</tr>
<tr>
<td>Total</td>
<td>2.85</td>
<td>0.40</td>
<td>1.90</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Table 4

Means and Standard Deviations of the Vocabulary Test Scores in Each Group (maximum = 8)

<table>
<thead>
<tr>
<th>Group</th>
<th>Text A</th>
<th>Text B</th>
<th>Text C</th>
<th>Text D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>High</td>
<td>2.73</td>
<td>2.05</td>
<td>2.36</td>
<td>2.04</td>
</tr>
<tr>
<td>Low</td>
<td>1.72</td>
<td>1.72</td>
<td>1.33</td>
<td>1.29</td>
</tr>
<tr>
<td>Total</td>
<td>2.23</td>
<td>1.94</td>
<td>1.85</td>
<td>1.77</td>
</tr>
</tbody>
</table>

Figures. Means of the two groups and the total in each test

Table 5

Z Values in Wilcoxon Signed-Ranks Test: Pairwise Comparisons of Reading Comprehension and Vocabulary Scores within Each Group

<table>
<thead>
<tr>
<th></th>
<th>Reading comprehension test</th>
<th>Vocabulary test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>B - A</td>
<td>-4.428**</td>
<td>-3.883**</td>
</tr>
<tr>
<td>C - A</td>
<td>-4.110**</td>
<td>-3.989**</td>
</tr>
<tr>
<td>C - B</td>
<td>-1.118</td>
<td>-1.281</td>
</tr>
<tr>
<td>D - B</td>
<td>-1.118</td>
<td>-3.682**</td>
</tr>
<tr>
<td>D - C</td>
<td>-0.202</td>
<td>-3.048**</td>
</tr>
</tbody>
</table>

Note: *p < .05; **p < .01; A = Text A; B = Text B; C = Text C; D = Text D.
3.8.1. Reading Comprehension Test

In the high proficiency group, the Friedman test revealed that there was a significant difference between the texts ($\chi^2(3, N=33) = 36.2, p < .01$). The Wilcoxon signed-ranks post hoc test showed that there was a significant difference between Text A and the other texts, B, C and D. This means that the students in the high proficiency group understood Text A significantly better than the other texts. In the low proficient group, the Friedman test also showed that there was a significant difference between the texts ($\chi^2(3, N=33) = 49.9, p < .01$). In addition to the similar significant difference between Text A and the other texts, the Wilcoxon singed-rank post hoc test revealed that the scores of Texts B and C were significantly better than that of Text D. This means that the students in the low proficiency group understood Text A significantly better too, but the comprehension of Text D was significantly less than for Texts B and C. In terms of the comparison between the groups, the Wilcoxon rank sum test showed that the scores of the high proficiency group were significantly better than those of the low proficiency group only in the case of Text D.

3.8.2. Vocabulary Test

As to the vocabulary test, the results of the two groups were exactly the same. According to the Friedman test, there was a significant difference between the texts in each group ($\chi^2(3, N=33) = 49.9, p < .01$ in the high proficiency group; $\chi^2(3, N=33) = 49.9, p < .01$ in the low proficiency group). The Wilcoxon post hoc test showed that the scores of vocabulary tests on Text D were significantly worse than those of the other texts in both groups. These facts seem to be a reflection of the extremely low scores in the vocabulary test on Text D in both groups. The Wilcoxon rank sum test revealed that the high proficiency group was significantly superior to the low proficiency group in all the tests except Text D.

### Table 6

<table>
<thead>
<tr>
<th>Text</th>
<th>Reading comprehension test</th>
<th>Vocabulary test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text A</td>
<td>-1.099</td>
<td>-2.100*</td>
</tr>
<tr>
<td>Text B</td>
<td>-.285</td>
<td>-2.138*</td>
</tr>
<tr>
<td>Text C</td>
<td>-.948</td>
<td>-2.498*</td>
</tr>
<tr>
<td>Text D</td>
<td>-3.958**</td>
<td>-.167</td>
</tr>
</tbody>
</table>

Note: *p < .05; **p < .01
4. Discussion

In both groups, the results of the reading comprehension tests were not consistent with those of the vocabulary tests. For instance, though the scores of the reading comprehension tests on Text A were significantly higher in the two groups than those on the other texts, there was no significant difference between the three texts except for Text D in the vocabulary test. Besides, in the case of the high proficiency group, the significantly lower scores of Text D did not reflect the fact that there was no significant difference in the reading comprehension test between Texts B, C and D. The reason for the discrepancy lay in the fact that the target words in the vocabulary test were not necessarily relevant to the answers to the reading comprehension questions. In fact, the correlations between the two tests were low (.16 in the high proficiency group and .30 in the low proficiency group).

Now, let us examine the research questions in terms of the results of the two kinds of test and data analysis. The answer to the first research question of whether test readability affects incidental vocabulary learning through glosses is affirmative. This conclusion comes from the facts that the higher readability a text has, the lower score it shows in the vocabulary test, and that there was a significant difference between the vocabulary test on Text D and those on the other texts in both the high and low proficiency group. In spite of the difference in proficiency, the difficulty of the text shown in readability has some influence on the likelihood of incidental vocabulary learning through glosses.

The second research question concerns whether incidental vocabulary learning occurs to only a negligible degree when the text readability is so high that the text is beyond learners' understanding. The answer to this question seems to be partly affirmative. In the low proficiency group, the mean of the reading comprehension test on Text D was significantly lower than those on the other texts. The results of the vocabulary tests are almost parallel: the mean on Text D is significantly lower than those on the other texts. Considering the fact that 27 students out of 34 got zero points in the vocabulary test on Text D, hardly any vocabulary gain from this text is seen because its readability was so high that students could not understand the content clearly. Therefore, the results in the low proficiency group are thought to be affirmative to the second question.

However, the results in the high proficiency group seem to be contradictory to the above results. Though there was no significant difference in the reading comprehension tests between Texts B, C and D, the vocabulary gain from Text D was significantly less than from the other texts. The mean score was almost as low as that in the low proficiency group, and 27 out of the 33 participants got zero points. These facts suggest that they understood Text D as well as Texts B and C, but little vocabulary gain was seen from Text D, unlike Text B and C. The main reason for this is the low correlation between the two tests as we mentioned above. The target words did not always play a crucial role, especially not on Text D, when the participants
answered the comprehension questions.

Another plausible reason may lie in the assumption that the target words on Text D might have been harder for the participants to learn than those in the other texts. The target words that showed a high percentage of correct answers were: (a) short concrete nouns (e.g. cod, scallop); (b) words whose meanings were relatively easy for the participants to infer from their morphemes (e.g. reship, downgrade); and (c) words that were relevant to the comprehension questions (e.g. cod, reship). The target words in Text D (unprecedented, staggering, perversion and unduly) did not have such characteristics or conditions.

To summarize, it could be said that there is some kind of threshold level in readability where incidental vocabulary learning hardly occurs, but it is still not clear which variables, readability or the difficulty of the target words have more influence on incidental vocabulary learning through glosses.

5. Conclusion

In this study, we have found that text readability affects incidental vocabulary learning through glosses. In addition, learning is unlikely to occur when text readability is so high that the text becomes too much of a burden for readers.

When students are expected to increase their vocabulary through extensive reading materials with glosses, the materials should be easy enough for them to understand the content without too much effort. In this case, text readability can be an important indicator in deciding which material should be chosen.

At the same time, it should be noted that incidental vocabulary gain through this type of reading is usually not large, as shown in this and other studies (Hulstijn, Hollander, & Greidaunus, 1996). In order to increase vocabulary through extensive reading, students should be given several chances to see the target words and word-and-form consolidation when necessary.

This study leaves several points for further research to consider. First, the target words in this experiment were chosen regardless of how difficult they were for students to learn. Different parts of speech might cause different levels of difficulty. In this study, the participants recognized short concrete nouns relatively better than other words. The morphemes of the target words could also have helped them to remember their meanings. Besides, words in the same category may be easier to learn than those that are totally different to each other. For instance, the target words in Text A were all related to seafood, which contributed to students’ success in finding the meanings of the words like “cod” or “scallop.” Second, this study was conducted without considering the relevance of the target words to the comprehension questions. The target words that played an important role in answering those questions might have attracted more attention from the participants than those that did not.
Therefore, additional studies taking these points into consideration are needed in the future. That is, target words that have the same difficulty and are under the same conditions should be chosen in the same type of experiment.

References


