Effective Glosses for Incidental Vocabulary Acquisition

Koichi NAGAYAMA  
*Kurume Junior High School Attached to Fukuoka University of Education*
Chizuru MORI  
*Fukuoka University of Education*

**Abstract**

The present study aims at investigating effective glosses for incidental vocabulary learning by adult L2 learners. Incidental vocabulary learning refers to subconscious learning from context and it can make use of marginal glosses on the text in order to enhance vocabulary acquisition while reading. There are two ways of supplying glosses. One is to make readers infer the meanings of the target words (Multiple-choice glosses) and the other is to simply give L2 synonyms (Single glosses). There has been a great deal of research on this issue, but the results are not consistent. In the present study, the subjects were divided into high and low proficiency groups. For gloss conditions, L2 Single gloss and two L2 Multiple-choice gloss types (similar and different) were prepared. Based on the results of the experiment, the most effective gloss type for each proficiency level was suggested and issues concerning the cognitive load of glosses were also discussed.

1. Introduction

The purpose of this study is to investigate effective glosses for incidental vocabulary learning. Incidental vocabulary learning means that one can learn L2 vocabulary unintentionally while reading or listening to a text. In incidental vocabulary learning through reading, both L1 and L2 researchers pointed out that it is important for the reader to infer the word meaning from its context. They have suggested that unknown words can best be learnt when presented in texts and when their meaning must be inferred from the context by learners themselves. However, several researchers (Konopak, 1988; Shatz & Baldwin, 1986) have demonstrated that contextual information is too misleading to allow correct inferences. One way to solve these problems is supplying the word meaning as a gloss. The glossing, however, seems to work against promoting retention for a long term. Therefore,
Hulstijn (1992) suggested Multiple-choice glossing as a remedy of this problem. Some researchers have studied which gloss type (Single vs. Multiple-choice) is more effective for incidental L2 vocabulary learning, but they have not shown consistent results confirming which gloss type was the most effective for incidental vocabulary acquisition.

2. Literature Review

2.1. Hulstijn (1992)

Hulstijn (1992) conducted five experiments to give an empirical contribution to 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. Among the five experiments, Experiments III and V were designed to investigate L2 learners' performance concerning Single gloss texts and Multiple-choice (MC) gloss texts. The subjects of the experiments were Turkish students learning Dutch as a second language.

2.1.1. Experiment III

In this experiment, he focused on comparing the Single gloss condition and the MC gloss condition while the subjects read the material with twelve marginal glosses. This consisted of twelve words: six of them were given the meaning as an L2 synonym and the other six were given the correct meaning in MC style. This MC consisted of one correct word with three distractors which were entirely different from the correct meaning. The MC procedure was as follows: At first, the subjects read a reading material, choosing six correct words in a MC gloss format. After the reading task, the subjects were asked to take a post test of the meaning of these twelve words in the original context. This post test was not known to the subjects in advance. The result was that retention in the MC gloss condition was significantly higher than in the Single gloss condition.

2.1.2. Experiment V

The subjects of this experiment did a reading task as before, and completed two unexpected meaning tests after the treatment (incidental learning task). Then the subjects were given a fifteen-minute break. After that, the experimenter returned the text to them and told that they would be, once again, tested on their knowledge of the words in the margin (intentional learning task). The posttests were given both in isolation and in context. The result was that the MC condition maintained better retention than the Single gloss condition, though this significant difference disappeared in the intentional task.

The two experiments showed that the MC condition made the readers recall the target words better in incidental vocabulary learning.

2.2. Watanabe (1997a)

Watanabe (1997a) investigated the effects of text modification and task on incidental learning of L2 vocabulary through reading. This text modification means the way of
supplying the explanation or the meaning of target words in an original text or a margin.

231 Japanese university students participated in the experiment. There were four groups according to the text type: the original text, the text with appositional vocabulary explanations, the text with marginal single glosses and the text with marginal MC glosses.

Appositional vocabulary explanations mean that the word meaning or the restatement is represented after the target word. The following is an example of the appositive.

Each year in the U.S. about 7,000 infants die in their cribs, babies' beds, for no apparent reason.

Although the appositives provide vocabulary explanations, the connection between the explanations and words being explained are not clearly shown. On the other hand, marginal single glosses provide explanations in an explicit manner. For instance:

Each year in the U.S. about 7,000 infants die in their cribs for no apparent reason.

MC marginal glosses were included to see whether more mental effort caused by the MC procedure would enhance the retention of the target word meanings. Interestingly, in terms of the distractors of MC, Watanabe (1997a) designed and adopted a quite different style from Hulstijn (1992). The MC consisted of two alternatives: a correct word meaning and a partially correct meaning of the target word which can be judged only in the context. For example:

Each year in the U.S. about 7000 infants die in their cribs for no apparent reason.

cribs=babies' beds

cribs=1. babies' beds

2. small beds

After reading each text, the subjects took vocabulary tests both in isolation and in context (immediate posttests), which were followed by the same test a week later (delayed posttests). The order of retention rate was the following in both tests: the lowest retention was the original, followed by the appositive, the MC gloss and the L2 Single gloss. There was a significant difference between the appositive condition and the marginal gloss conditions (L2 Single gloss and MC gloss) but there was no significant difference between these two marginal gloss conditions.

Clearly, these findings of Watanabe (1997a) were different from those of Hulstijn. Hulstijn (1992) reported that the inferred word meanings (i.e. the MC gloss group) caused better retention than the given word meanings (i.e. the Single gloss group), while in Watanabe (1997a), Single glosses seemed to be the most effective way for vocabulary learning. Watanabe (1997a) suggested two reasons for this discrepancy: 1. There were differences in the experimental design between his study and that of Hulstijn. 2. In terms of MC gloss distractors, Hulstijn prepared incorrect meanings, whereas Watanabe (1997a) provided his subjects with another correct meaning or partially correct meaning of each target word.

2.3. Watanabe (1997b)
In his study, Watanabe (1997b) focused on the effects of L1 Single glosses and L2 MC glosses on incidental vocabulary learning. He also investigated the relationship between the subjects' L2 proficiency and their score of the vocabulary tests. This was a new perspective which had not been dealt with in Hulstijn's experiments.

Twenty-four non-native speakers of English, that is, Japanese, Chinese and Korean students took part in the study as subjects. Fifteen target words were presented in the margin, five with MC glosses, the other five with translation in L1 and the remaining five with no glosses. The MC glosses used in this experiment consisted of a correct word and a distractor which could be judged more easily than Watanabe (1997a). For example:

Researchers uncover a new culprit in the mystery of sudden infant death syndrome.

culprit = 1. story 2. cause

In the low proficiency level, there was a significant difference between the No gloss group and the L1 Single gloss group which obtained the highest score in the later vocabulary test. On the other hand, in the high proficiency level, there was a significant difference between No gloss and L2 MC gloss which was the highest score. It was found that L2 MC glosses could cause higher retention in the high proficiency group, but it was not the case for the low proficiency group.

3. The Present Study

As mentioned earlier, there was a discrepancy between the earlier studies concerning the effects of Single glosses and MC glosses on incidental vocabulary learning. It is considered that there are two reasons for this: subjects' L2 proficiency and the construction policy of MC glosses. Hulstijn (1992) did not design his experiments to analyze the comparison by the subjects' L2 proficiency, whereas Watanabe (1997b) conducted his experiments taking the subjects' L2 proficiency into account. As the result, the MC gloss condition could cause better retention in the high proficiency level, but the results in the low proficiency level were different from Hulstijn (1992). As for the construction policy of MC glosses, Hulstijn (1992) used a word with a quite different meaning from the correct word as a distractor of each MC gloss. On the other hand, Watanabe (1997a) constructed the distractor of each MC gloss using a similar word to the correct word. In other words, Watanabe (1997a) provided the subjects with another correct meaning or partially correct meaning of each target word as a distractor. Generally speaking, it seems more difficult for subjects to choose the correct meaning from words with similar meanings; and thus, the types of MC glosses can be a factor to affect learners' incidental vocabulary acquisition.

In the present study, therefore, the differences of conditions as mentioned above were readjusted in terms of subjects' proficiency levels and gloss types. For gloss conditions, L2 Single gloss and two L2 MC gloss types (similar distractor and different distractor) were
prepared. The subjects were divided into two groups based on their L2 proficiency level (high or low) to investigate whether these three gloss types would cause different results in different proficiency levels.

3.1. Method

3.1.1. Subjects and Text types

In the present experiment, 122 Japanese university students served as subjects. These subjects were divided into two groups by their proficiency level based on their majors (English majors and non-English majors). The number of English majors was fifty-one students, and that of non-English majors was seventy-one students. Before carrying out the experiment, all subjects were tested for English proficiency with a simplified CELT that consisted of two sections: vocabulary and structure containing thirty-eight and thirty-seven questions respectively. The t test was used for the analysis and the result showed that there was a significant difference between these two groups (t=11.03, p<0.01). It was then confirmed that English majors could be regarded as the high proficiency group and non-English majors could be regarded as the low proficiency group in the present experiment.

The subjects in each proficiency level were divided into three gloss conditions randomly (L2 Single, L2 MC similar, and L2 MC different). The number of each condition group is as follows:

<table>
<thead>
<tr>
<th>Table 1. The Number of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Group</td>
</tr>
<tr>
<td>High proficiency</td>
</tr>
<tr>
<td>Low proficiency</td>
</tr>
</tbody>
</table>

The two MC gloss conditions were divided into the MC different gloss condition which was used in Hulstijn (1992) and the MC similar gloss condition based on Watanabe (1997a). The MC similar glosses of the present study consisted of the correct word and the distractor which was similar to the correct word meaning as in Watanabe (1997a). The target words were ten words, five in one text and five in the other text. Table 2 shows the examples of MC glosses in each proficiency level.

<table>
<thead>
<tr>
<th>Table 2. The Examples of MC Glosses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target word</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>MC similar</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>MC different</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

(* the distractor of the target word)
3.1.2. Materials and Readability

Two reading materials for each proficiency level were prepared. The reason why different materials were prepared for the high and low proficiency levels was that the main aim of this study was to compare the results of the three conditions in each proficiency level, not the direct comparison of the high and low proficiency levels. Therefore, it seemed more important to select reading materials appropriate for each level.

For the low proficiency group, two texts which contained 301 words and 245 words respectively were provided. On the other hand, for the high proficiency group, two texts which consisted of 368 words and 327 words were prepared. The readability of these two texts in each proficiency level was kept approximately the same by the use of the new Dale-Chall Readability formula (Chall & Dale, 1995). According to the formula, the readability of the texts in the low proficiency level was 9, while that of the texts in the high proficiency level was 13. (The percentage of items correctly answered in the reading comprehension tests in the experiment was 62% for the low proficiency group and 53% for the high proficiency group. It is reasonable to assume that these texts were not particularly too difficult for each group).

3.1.3. Procedure

Each subject of the three groups in each proficiency level received a test booklet. The booklet consisted of two texts which included ten underlined target words in total and comprehension tests. The subjects were instructed to read a text for ten minutes and to choose the correct answer of each MC if they were in the MC gloss conditions. After that, the subjects took the comprehension test without looking back the text. The same procedure was followed for the other text. After finishing this reading section, the subjects were told to take two unexpected vocabulary tests: the target words in isolation, and the target words in a part of the original context. They took these tests after the reading task (immediate tests), and the same test after a week (post tests).

In the two vocabulary tests, the subjects were allowed to write answers in either Japanese or English. The reason was that the purpose of this experiment was to investigate whether they remembered the word meanings. Therefore, it did not matter whether these word meanings were remembered in Japanese or in English. In terms of whether the target words were really unknown words to the subjects, the researchers asked the subjects to mark the words which they had already known before they received the experiment. Only the words which were confirmed unknown on the basis of this marking were analyzed in this experiment.

3.2. Results

In this analysis, the results of each condition were compared only within each proficiency level and these two proficiency levels were not compared to each other as mentioned earlier. The results are shown in Tables 3 and 4 and Figures 1,2,3 and 4.
Table 3. Means and Standard Deviations of the Three Conditions in the High Proficiency Level
(maximum=100%)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Immediate test</th>
<th>Post test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Isolation (SD)</td>
<td>Context (SD)</td>
</tr>
<tr>
<td>L2 MC similar</td>
<td>29.92 (16.67)</td>
<td>51.60 (16.15)</td>
</tr>
<tr>
<td>L2 MC different</td>
<td>19.93 (13.81)</td>
<td>42.74 (23.99)</td>
</tr>
</tbody>
</table>

Table 4. Means and Standard Deviations of the Three Conditions in the Low Proficiency Level
(maximum=100%)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Immediate test</th>
<th>Post test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Isolation (SD)</td>
<td>Context (SD)</td>
</tr>
<tr>
<td>L2 Single gloss</td>
<td>29.76 (16.00)</td>
<td>43.37 (20.55)</td>
</tr>
<tr>
<td>L2 MC similar</td>
<td>27.98 (16.52)</td>
<td>43.11 (16.78)</td>
</tr>
<tr>
<td>L2 MC different</td>
<td>38.32 (22.09)</td>
<td>46.72 (20.05)</td>
</tr>
</tbody>
</table>

Figures. Means of the Three Conditions on Each Test in Each Proficiency Level

1. Immediate test for the high proficiency level
2. Post test for the high proficiency level
3. Immediate test for the low proficiency level
4. Post test for the low proficiency level
Two-way ANOVA with repeated measures was used to examine the difference between the three conditions and the test types (Isolation test and Context test). In the high proficiency group, there was a significant difference among the three conditions in the immediate test \( (F(2,48)=7.59, p<0.01) \). There was also a significant difference between the test types \( (F(1,48)=112.27, p<0.01) \). However, there was not a significant interaction \( (F(2,48)=2.45, \text{n.s.}) \). Further analysis revealed that there was a significant difference both in Isolation test and in Context test \( (p<0.01) \). To locate the difference, a multiple comparison of Tukey method was applied, which showed that there was a significant difference between one condition and other three conditions both in Isolation test and in Context test. In terms of the difference between the test types, there was a significant difference between Isolation test and Context test in each of the three conditions.

In the post test, a significant difference among the three conditions disappeared and there was no significant difference \( (F(2,48)=2.56, \text{n.s.}) \). There was a significant difference between the test types \( (F(1, 48)=73.55, p<0.01) \) and also a significant interaction between the conditions and the test types \( (F(2, 48)=10.36, p<0.01) \). As the result of further analysis, there was no significant difference between any two of all three conditions in Isolation test. However, in Context test, there was a significant difference between the SG condition and the MC similar condition, between the MC similar condition and the MC different condition, and between the SG condition and the MC different condition \( (p<0.01) \). Concerning the difference between the test types, a significant difference was shown both in the MC similar condition and the MC different condition \( (p<0.01) \), while it disappeared in the SG condition.

The next analysis was held in the low proficiency group. In the immediate test, there was no significant difference among the three conditions \( (F(2,68)=1.04, \text{n.s.}) \). Between the test types, there was a significant difference \( (F(1,68)=55.53, p<0.01) \). The interaction between the gloss conditions and the test types was not significant \( (F(2,68)=1.50, \text{n.s.}) \). Regarding the test types, a significant difference was found between Isolation test and Context test in each of the three conditions \( (p<0.01) \).

In the post test, there was no significant difference among the three conditions \( (F(2,68)=0.40, \text{n.s.}) \). As in the immediate test, a significant difference was shown between the test types \( (F(1,68)=48.26, p<0.01) \), but there was no significant interaction between the gloss conditions and the test types \( (F(2,68)=2.21, \text{n.s.}) \). In terms of the test types, there was a significant difference between Isolation test and Context test in each of the three conditions, as in the immediate test \( (p<0.01) \).

4. Discussion

In terms of subjects' proficiency levels and gloss types, the following findings were obtained. In the high proficiency level, the MC similar group got the highest retention rate
both in the immediate and the post tests. The significant difference among the three conditions in Isolation test disappeared in the post test, while it remained in Context test. In the low proficiency level, the two MC gloss groups could not get a higher retention rate than the other group (the Single gloss group). As compared within the MC gloss types, the subjects of low proficiency level seemed to prefer the MC different condition to the MC similar condition.

The results of this experiment indicate that MC glosses have some effects in the high proficiency level. As Hulstijn (1992) pointed out, the subjects are likely to remember the meaning of an unknown word when they have inferred it by themselves, rather than when given the meaning. On the other hand, the low proficiency group could not remember the meaning of an unknown word even if they inferred it by themselves. As the reason, MC glosses seem to require more of the subjects’ serious effort on the target words than Single glosses. Watanabe (1997b) suggested that “MC glosses (inferred meanings) in reader’s L2 seemed to have induced cognitive effort and distinctive trace in memory for higher proficiency students but they were not as effective for lower proficiency students” (p.189). Therefore, the difference between the results in each proficiency level seems to be caused by whether this extra cognitive effort becomes an overload or not for the subjects of each proficiency level. For the high proficiency group, it appears that these efforts required to choose the correct answer in MC glosses did not become an overload. On the other hand, for the low proficiency group, the task might have been cognitively excessive.

The discrepancy of the results between the earlier studies may be explained by the experiment of the present study more clearly. Hulstijn (1992) showed that the MC gloss condition could cause better retention of the target words than the Single gloss condition. This tendency was confirmed by the MC gloss conditions in the high proficiency level of this experiment. According to Watanabe (1997a), however, any MC gloss conditions could not get a higher retention rate than the Single gloss condition. This result runs counter to the insistence by Hulstijn (1992), but is consistent with the MC similar gloss condition in the low proficiency level of the present experiment. Hulstijn (1992) and Watanabe (1997a) did not design their own experiments to explore the issue of the subjects’ L2 proficiency. It is inferred then that the difference between the two experiments may be explained by the L2 proficiency level of the subjects. That is, the reason why the MC gloss condition was preferred in Hulstijn (1992) was that the subjects might have had higher proficiency level of L2 than the subjects in Watanabe (1997a).

There is another point worth noting: the issue of how the level of adjacency (similarity) between the correct word and the distractor influences the retention of the target word meanings. As mentioned earlier, it should be noted that the MC similar condition caused a significantly higher retention rate in Context test of the post test in the high proficiency level. There seems to be two reasons for this. First, the subjects in the MC similar condition
seemed to pay more attention to each MC gloss since the distractor was very close to the correct word in meaning. The subjects might have referred to the immediate context more intensely to judge the correct word meaning. Second, the subjects in the MC similar condition seemed to be able to capture the target word meaning as a vague image referring to the immediate context better than the subjects in the MC different gloss condition. For example, the subjects of the MC similar condition in the high proficiency level answered “Samuzamushii (寒々しい)” in Japanese for the target word ‘bleak’, though the correct word meaning of this was “wild (荒涼とした)”. It is inferred that the answer was induced from the distractor of the MC similar condition, “cold”. The answer is incorrect in this context, but fairly close to the correct meaning. In this way, the subjects seemed to capture the target word meaning as a very close image of the target word referring to the immediate context.

The present study leaves room for further research. In the experiment, the subjects could not check whether they could choose the correct word of MC glosses until they finished all the tests. Therefore, additional studies will be needed to examine whether this result would be changed, if they could confirm the correct answers of MC glosses before they take a vocabulary test.

References


Appendix

Ten target words for High Group: distort, erect, appoint, confirm, intervene, magnification, rigorous, emphatic, derive, bleak

Ten target words for Low Group: huge, estimate, earn, profit, hire, rapid, alter, reduce, accommodate, innovation