

# An Experimental Research on the Effects of Types of Glossing on Incidental Vocabulary Acquisition through Reading\*

Shan Liu  
Zhejiang Gongshang University, Hangzhou, China

**Abstract**—A large amount of research has been conducted to delve into the means of improving the rate of incidental vocabulary acquisition through reading, which includes providing glosses or annotations, increasing the reoccurrence of the target words and taking the advantage of dictionaries. But little has been done on the effects of different types of glossing on incidental vocabulary acquisition through reading. The present study has been aimed at finding out whether the three different types of glossing, i.e. glossing in Chinese, glossing in English, and glossing in both Chinese and English, have different effects on enhancing the incidental vocabulary acquisition rate through reading. From the data analysis of the present study, it has been found that in terms of enhancing the depth of vocabulary knowledge in incidental vocabulary learning through reading, glossing in both Chinese and English is the most effective glossing type, whether it be in the immediate retention testing section or the delayed retention testing section. In terms of enhancing the breadth of vocabulary knowledge, glossing in Chinese is the most effective one in the immediate retention of the breadth of word knowledge, and glossing in both Chinese and English is the most effective in the delayed retention of the breadth of word knowledge. On the basis of these findings, implications and suggestions for the arena of foreign language teaching and other related fields have been proposed.

**Index Terms**—type of glossing, incidental vocabulary acquisition, reading; breadth of vocabulary knowledge, depth of vocabulary knowledge

## I. INTRODUCTION

After having been neglected for decades in the arena of foreign or second language teaching and learning, the teaching of vocabulary has regained its rightful and significant place in the field of language teaching. Much has been done on how to enhance the acquisition of vocabulary (Chen and Truscott, 2010; Gui, 2015; Saz et al., 2015; Tanaka, 2017; Zeeland and Schmitt, 2013).

Large numbers of scholars focus on ways to enhance incidental vocabulary acquisition through reading, including adding text-based vocabulary exercises, introducing dictionary use, providing glosses, annotations, whether in visual or aural forms (Chen and Truscott, 2010; Chun and Plass, 1996; Coady, 2001; Gardner, D., 2004; Grabe and Stoller, 2001; Gui, 2015; Saz et al., 2015; Wesche and Paribakht, 2000).

## II. RESEARCH QUESTIONS OF THE PRESENT STUDY

The present study has been aimed at delving into the effects of different types of glossing, that is, glossing in Chinese, glossing in English, and glossing in both Chinese and English, on incidental vocabulary acquisition through reading.

The specific research questions are listed as follows:

Q1: Does a better comprehension of the reading text contribute to a higher retention rate of word knowledge? In other words, is it true that the better one understands the text, the better one could acquire new vocabulary?

Q2: Which type of glossing, among the three types of glossing, i.e. glossing in Chinese, glossing in English and glossing in both Chinese and English, is the most effective in enhancing incidental vocabulary acquisition through reading?

## III. RESEARCH DESIGN

### A. Subjects

Three classes of non-English-major freshmen from a college in Hangzhou, China took part in the study. One class took part in the pilot study, and the other two classes took part in the formal tests. All the students in the three classes

---

\* This paper has been sponsored by the Research Project of Zhejiang Federation of Humanities and Social Sciences Circles (No.16NDJC197YB), the Research Project of the Department of Education of Zhejiang Province (No. Y201534465), Zhejiang Province's Higher Education Reform Research Project on Classroom Teaching (No. kg2013550) and the Reform Research Project on Classroom Teaching of Commerce of Hangzhou, Zhejiang Gongshang University.

are of the same proficiency, which is guaranteed by the fact that in this college students have to take part in an English proficiency test and then, according to the scores in this test, students of the same proficiency in English will be grouped into certain classes of the same level. The class taking part in the pilot study consists of 35 students, including 17 male students and 18 female students. The other two classes consist of 68 students, 33 male students and 35 female students. However, in the formal testing, among all the students, only 62 took part in all the tests and finished all the papers. What is more, two among the 62 students misunderstood one of the tests, thus being eliminated from the data analysis. In addition, eight students in the two classes in total were randomly selected to take part in the semi-structured interviews during breaks immediately after the two testing sections respectively.

### B. Target Words

There are 18 target words in total in this experiment, and all of them are picked up according to the College English Curriculum Requirements (For Trial Implementation) issued by the department of higher education of the Ministry of Education and the national English Curriculum Standards for Senior Middle Schools released by the Ministry of Education. In other words, they are beyond the requirement of senior high school and are required in college English study.

What needs to be pointed out is that all the forms of the original target words in the study are replaced by nonsense words, or pseudo words, which have been directly taken from Webb (2007) or Nation (1990) or coined based on the nonsense words used by Webb or Nation. According to Webb (2005, 2007), compared with using original target words, there are several advantages in replacing the L2 forms of the target words. First of all, this is to make sure that subjects have no previous information of any kind concerning the target words, which is significant because whatever gains are observed can be ascribed solely to the treatment. Accordingly, the researcher may get a more accurate assessment of what they are trying to measure when they use nonsense words (*ibid*). In studies in which authentic L2 target words are used, the results might be questioned in that it cannot be certain that the learners did not have any prior knowledge of the target words (Webb, 2007). In most of these studies, although researchers use a pre-test/post-test design, there is still the possibility that the pre-test may not be sensitive enough to show that the learners had partial knowledge of the target words or that the participants have partial knowledge of the form, associations, syntax, and grammatical functions but just because of their uncertainty about the meaning they respond that the word is unknown (*ibid*). There is also the possibility that the pre-tests made the participants aware that the focus of their task was vocabulary learning (*ibid*). On the other hand, researchers using authentic words may find it comparative difficult to find the target words in context in which all the running words are known, since real target words are more likely to be low frequency words, which in turn tend to co-occur with other low frequency words (*ibid*).

In addition to the advantages of using nonsense words claimed by Webb, employing nonsense words was also owing to the results that we have found out in the pilot study. From the pilot study, we found that more than half of the original target words have been mastered by more than 60% of the students. That is to say, most of them knew most of the original target words. This fact has also been testified by the interview following the tests for some students. As a consequence, nonsense words were introduced to the present study.

In addition, the effect of the part of speech of the target words on incidental vocabulary acquisition has been controlled in the present study. The part of speech, or grammatical category, of words has been proved by scholars to have a great impact on vocabulary acquisition (Horst et al cited in Webb, 2007; Laufer, 2002; Coady, 2001). According to Laufer (2002), certain grammatical categories are more difficult to learn than others. Nouns seem to be the easiest; adverb, the most difficult; verbs and adjectives are somewhere in between. As a result, only nouns and verbs are taken as target words in this experiment. More specifically, there are 12 nouns and six verbs, among which four nouns and two verbs are glossed in Chinese, another four nouns and two verbs in English, the remaining four nouns and two verbs in both Chinese and English. In addition, the choice of words to be glossed in the three types is random.

### C. Instruments

A reading comprehension test, a test of the breadth of vocabulary knowledge, and a test of the depth of vocabulary knowledge are used as instruments in this study. In addition, semi-structured interviews are also employed in this experiment.

#### 1. Reading comprehension test (CT)

The reading material used in this study (see Appendix VI) is a passage that has been taken from Unit One, Book One in College English, chief-edited by Dong Yafen and published by Shanghai Foreign Language Education Press. This set of textbooks is considered as an outstanding one and it has been recommended by the Ministry of Education. The passage remains the same with the original one except that the original target words were replaced with nonsense words and a slight change was made in the last paragraph, (i.e. we changed the form of a noun 'accumulation' into the gerund of its verb 'accumulating' to make the ratio of nouns to verbs as 2: 1).

There are 774 words in total in the reading passage, which contains 17 new words apart from the 18 target words. That means the coverage of words is about 97%, which is just between 95% and 98%. According to much research done, for basic necessity, 95% is needed, and the coverage of 98% is needed for reading for pleasure (Hirsh and Nation, 1992 cited in Laufer, 2001; Coxhead, 1998 cited in Alderson, 2007). So the present study, the coverage of words in the passage is just within this range.

The 17 new words are annotated in Chinese immediately after the word in the parentheses and the 18 target words are glossed respectively in Chinese, in English or in both Chinese and English in the margin. The glosses of the target words are taken directly from the original text book, including its part of speech and its definition in Chinese, in English or in both.

Following the passage are ten comprehension questions in the form of multiple choice. All the questions are designed on the basis of information in the passage and they cover all the paragraphs in the passage.

There are two purposes in designing the comprehension questions. Firstly, it is to answer the research question whether a better comprehension of the text leads to a better retention of vocabulary knowledge. This could also investigate how the comprehension of the text by students is related to their retention of word knowledge in the process of incidental vocabulary acquisition. In the second place, it is to ensure that subjects will not intentionally learn the target words, for their attention is drawn by the task of comprehending the text and completing all the questions following the text.

One point is awarded to one correctly made choice, and no reduction of points for wrong answers. Therefore, the highest score possible for this test is ten.

#### 2. Test of the breadth of vocabulary knowledge (BT)

The test of the breadth of word knowledge (see Appendix V) employs the form of the Vocabulary Levels Test by Nation (1990). Three items of definitions of the same part of speech are put in one group on the left side and six corresponding choices are on the right side. Subjects need to choose three from the six choices to match the three definitions given on the left part. However, there is a prominent difference from that of Nation in the present test in that all the three items in one group on the left part may be in Chinese or in English, for the purpose of being in line with the different glossing types of the target words in the original reading comprehension test. Since among the 18 target word there are four nouns and two verbs which are glossed in both Chinese and English, the definition of two nouns and one verb among them is given in English and the definition of the remaining two nouns and a verb are in Chinese.

For example:

- |                                |             |
|--------------------------------|-------------|
| 1. ___ knowledge of sb. or sth | a. service  |
| 2. ___ chance to do sth.       | b. denent   |
| 3. ___ careful and hard effort | c. aleand   |
|                                | d. cheltian |
|                                | e. tobacco  |
|                                | f. vinegar  |
| 4. ___ 抱怨                      | a. project  |
| 5. ___ 分配 (任务)                 | b. taspe    |
| 6. ___ 把.....塞进                | c. fix      |
|                                | d. demart   |
|                                | e. segat    |
|                                | f. explain  |

All the definitions on the left are taken directly or shortened a little for the purpose of getting a shorter definition from the original glosses in the reading passage. One point is awarded to each correctly answered item and therefore the full score for this test is 18 points in total.

#### 3. Test of the depth of vocabulary knowledge (DT)

The test of the depth of vocabulary knowledge (DT, see Appendix VI) has been designed on the basis of Qian (2002)'s DVK format. All of the 18 target words, including 12 nouns, 6 verbs, have been taken in the test of the depth of vocabulary knowledge.

An example is given in the following:

faddam \_\_\_\_\_

|                 |           |          |            |
|-----------------|-----------|----------|------------|
| A. civilization | B. plan   | E. full  | F. rich    |
| C. approach     | D. policy | G. great | H. ancient |

As could be seen from the above example, there are four choices in the left box for synonymy. It is a little different from what Qian's format in that he has included another aspect of vocabulary knowledge: polysemy. In the present test, only synonym has been taken into consideration and the aspect of polysemy has been dropped because all the target words only appear once in the passage and only one meaning is available in the testing process for subjects. And the four choices in the right box are for collocations, which is the same with Qian's format. All the correct choices are designed based on *Longman Dictionary of Contemporary English*, *Collins COBUILD Dictionary* and *The LTP Dictionary of Selected Collocations*.

The correct answers are always four in number for each item, and the four choices may be one from the left side and three from the right, or two from both sides respectively, or one from the left side and three from the right. There are 18 items in this test, and one point is awarded to each correct choice among the four answers for each item. As a result, the full score for each item is four points.

#### 4. Semi-structured interview

The first semi-structured interview in the first section of testing (see Appendix VII) was designed mainly to see whether the subjects read glosses or not, or which kind of glossing they paid more attention to and which kind of glossing they find more beneficial. The second semi-structured interview (see Appendix VIII) was intended to see how students felt about the word tests and word remembering.

5. Pilot study

A pilot study was carried out before the formal experiment. One class of non-English-major freshmen of the same proficiency in the English language with the subjects in the experiment took part in the pilot study. The testing papers used in the pilot study all employed the original true words as target words, including the comprehension test (see Appendix I), the test of the breadth of vocabulary knowledge (see Appendix II), and the test of the depth of vocabulary knowledge (see Appendix III). After the three tests, six students are randomly interviewed. Revision was later made according to the results collected from the pilot study.

The most interesting point found in the pilot study is that more than half of the original target words have been mastered by most of the students. That is to say, most of them have mastered much more words than required by the national English Curriculum Standards for Senior Middle Schools released by the Ministry of Education. This was also testified by the following interview with the students. As a result, the author made important revision in the experimental material and nonsense words were introduced to this study.

D. Procedure

Two classes of non-English-major freshmen from a college in Hangzhou, China took part in the present research. All the participants are of the same proficiency in English, which has been ensured by the classification of English classes according to their English scores in the large-scale National College Entrance Examination in China. All the tests were executed during normal periods of classes. In the first stage, subjects were required to finish a reading comprehension test (CT), which contains a reading passage followed by ten comprehension questions. There are 18 target words in the passage, among which six are glossed in Chinese, six in English and the remaining six in both Chinese and English. After all the subjects had finished this comprehension test, papers were collected. Then the test of the depth of knowledge of the target words (DT) was administered preceding to the administration of the test of the breadth of knowledge of the target words (BT) for the purpose of reducing practice effect. All the three tests cost two periods of normal classes.

One week after the first section of testing (or the immediate testing), the second section of testing (or the delayed testing), including the test of the breadth of vocabulary knowledge (BT) and the test of the depth of vocabulary knowledge (DT) was carried out in a normal period of class to investigate the delayed retention of the target words, both in terms of the breadth knowledge and in terms of the depth knowledge. The same testing papers were used in the second testing section and similarly, the test of the breadth of vocabulary knowledge was preceded by the test of the depth of vocabulary knowledge. The whole section cost half an hour in total.

During the class breaks after the immediate and delayed tests respectively, students were randomly selected to be interviewed by way of semi-structured interview. The semi-structured interviews was designed mainly to see whether the subjects read glosses or not, or which kind of glossing they paid more attention to and which kind of glossing they found more beneficial.

IV. RESULTS AND DISCUSSION

A. Reliability of the Five Tests Used in This Study

The following table (Table 4-1) illustrates the reliabilities of the five tests including the comprehension test (CT), two tests of the breadth of vocabulary knowledge (BT) and two tests of the depth of vocabulary knowledge (DT) employed in this experiment.

TABLE 4-1  
RELIABILITY OF THE FIVE VOCABULARY TESTS

| Tests | Cronbach's Alpha | N of Items |
|-------|------------------|------------|
| BT1   | .804             | 18         |
| BT2   | .714             | 18         |
| DT1   | .841             | 18         |
| DT2   | .941             | 18         |
| CT    | .926             | 10         |

BT1=the First Breadth Test  
 BT2=the Second Breadth Test  
 DT1=the First Depth Test  
 DT2=the Second Depth Test

As can be seen from the above table, in all the tests, no item in all the five tests has been excluded. In addition, the Cronbach's Alpha index of the second test of depth of vocabulary (DT2) reaches as high as .941 and that of the comprehension test (CT) reaches a level of .926. Even the lowest of the five (BT2) is as high as .714. The Cronbach's Alpha index of the first test of the breadth of vocabulary knowledge is a little higher than the second, and the first test of

the depth of vocabulary knowledge ranks third with a Cronbach's Alpha index of .841.

That is to say, all of the five tests boast a sound degree of reliability, thus testifying the reliability of the five tests. The Cronbach's Alpha index of the second test of the depth of vocabulary knowledge is higher than .91, which is the Cronbach's Alpha index gained the investigation by Qian and Schedl (2004), and the other .841 is lower than this figure. Therefore, all of them could be employed in the future studies.

#### B. Correlation between Comprehension and Incidental Vocabulary Acquisition

This research question of "Does a better understanding of the text lead to a higher rate of incidental vocabulary acquisition?" can be addressed from two perspectives, i.e. from the perspective of the breadth of vocabulary knowledge and from the perspective of the depth of vocabulary knowledge.

##### 1. Correlation between comprehension and the breadth of vocabulary knowledge

Does a better understanding of the text lead to a higher rate of incidental vocabulary acquisition in terms of the breadth of vocabulary knowledge? Before answering this question, the following table of the descriptive statistics (Table 4-2) needs to be examined.

For the purpose of a clearer demonstration and comparison, the scores used here in these two tests are the average of each item in the test paper, instead of the total score of the whole paper. That is to say, the full average score is one point for both the comprehension test and the first breadth test of vocabulary knowledge.

TABLE 4-2  
DESCRIPTIVE STATISTICS OF CT AND BT1

|     | Mean | S.D. | N  |
|-----|------|------|----|
| BT1 | .441 | .233 | 60 |
| CT  | .793 | .145 | 60 |

CT=Comprehension Test  
BT1=the First Breadth Test

It could be seen that the mean of the first breadth test of vocabulary knowledge is 0.441, lower than that of the comprehension test, whose mean is 0.793. This may contribute to the fact that most of them could understand the text well, which is also in accordance with what has been obtained from the interview. In the interview after the immediate retention tests, five among the eight students interviewed told the author that they could understand the text well in spite of the fact that there were many words which they did not know.

And as to the question itself, Table 4-3 gives a definite answer in the following table:

TABLE 4-3  
CORRELATIONS BETWEEN CT AND BT1

|     |                     | CT      | BT1     |
|-----|---------------------|---------|---------|
| CT  | Pearson Correlation | 1       | .278(*) |
|     | Sig. (2-tailed)     |         | .032    |
|     | N                   | 60      | 60      |
| BT1 | Pearson Correlation | .278(*) | 1       |
|     | Sig. (2-tailed)     | .032    |         |
|     | N                   | 60      | 60      |

\* Correlation is significant at the 0.05 level (2-tailed).

CT=Comprehension Test  
BT1=the First Breadth Test

It is obvious that the comprehension test is significantly correlated with the first breadth test of vocabulary knowledge ( $p=.032<.05$ ). In other words, generally speaking, students who have obtained a higher score in the comprehension test tend to gain a higher score in the first breadth test of vocabulary knowledge. That is to mean, better understanding of the text, which is operationalized as the score in the comprehension test, will lead to a higher rate of incidental vocabulary acquisition in terms of the breadth of word knowledge. As a result, the previous question raised in the beginning of this part has been answered.

This conclusion could be understood easily. In the course of reading the text and finishing all the comprehension questions, students had to pay at least some attention to the meaning of the nonsense target words in the text in order to understand the whole text. Though they did not consciously remember these words by heart, for they had not been informed of the following tests of these words in the beginning, the aspect of the meaning of some target words in the context of the reading text was indeed incidentally acquired. And the better one understood the meaning of the text, the more he was familiar with the details of the text, including knowledge about the target words, thus rendering a higher rate of incidental vocabulary acquisition in terms of the breadth of vocabulary knowledge.

##### 2. Correlation between comprehension and the depth of vocabulary knowledge

Does a better understanding of the text lead to a higher rate of incidental vocabulary acquisition in terms of the depth of vocabulary knowledge? Similarly, the descriptive statistics in the following table (Table 4-4) are to be looked into. For the purpose of a clearer demonstration and comparison, the scores used here in these two tests are the average of each item in the test paper, instead of the total score of the whole paper. That is to say, the full score for each item is one

point for the comprehension test and four points for the first depth test of vocabulary knowledge.

TABLE 4-4  
DESCRIPTIVE STATISTICS OF CT AND DT1

|     | Mean  | Std. Deviation | N  |
|-----|-------|----------------|----|
| DT1 | 1.836 | .553           | 60 |
| CT  | .793  | .145           | 60 |

CT=Comprehension Test  
DT1=the First Depth Test

Since the mean and the standard deviation have been analyzed in the pervious part, only the first depth test of vocabulary knowledge is to be explored. As could be seen from the table above, the mean of the first depth test is about 1.836 on a scale of four points, which demonstrates that less than half of the depth knowledge has been acquired. What is more, its standard deviation is about 0.553.

TABLE 4-5  
CORRELATIONS BETWEEN CT AND DT1

|     |                     | DT1   | CT    |
|-----|---------------------|-------|-------|
| DT1 | Pearson Correlation | 1     | -.047 |
|     | Sig. (2-tailed)     |       | .723  |
|     | N                   | 60    | 60    |
| CT  | Pearson Correlation | -.047 | 1     |
|     | Sig. (2-tailed)     | .723  |       |
|     | Covariance          | -.004 | .021  |
|     | N                   | 60    | 60    |

CT=Comprehension Test  
DT1=the First Depth Test

The above table (Table 4-5) presents to the reader clearly whether the comprehension test and the first depth test of vocabulary knowledge are correlated with each other.

From the above table, we can see that these two factors are not correlated with each other, for the value of p is as high as 0.723. Accordingly, we could come to the conclusion that the comprehension test and the first depth test of vocabulary knowledge are not correlated with each other. This indicates that a better understanding of the text may not necessarily lead to a better retention of the depth of word knowledge in the process of incidental vocabulary acquisition.

This conclusion could be easily understood that one encounter with a new word in a reading text is far from enough to acquire all aspects of word knowledge, such as synonyms and collocations. It is a long process to develop one's depth of word knowledge, which has been proved by many.

On the other hand, while reading the text and answering the comprehension question, most subjects have not paid much attention to the target words themselves. In most cases, only the superficial meaning of the target words has been noticed and other aspects have been ignored.

Therefore, the conclusion could be drawn that a better understanding of the text does contribute to a higher rate of incidental vocabulary acquisition from the perspective of the breadth of word knowledge. However, a better understanding of the text does not guarantee a higher rate of incidental vocabulary acquisition in terms of the depth of word knowledge.

*C. Effects of Different Types of Glossing on Incidental Vocabulary Acquisition*

In this part, the research questions put forward in Chapter Three are to be investigated. In a word, the general research question is which kind of glossing, among glossing in Chinese, glossing in English and glossing in both English and Chinese, is the most effective one in enhancing incidental vocabulary acquisition through reading. The question should be answered from the perspectives of both the breadth and the depth of vocabulary knowledge and at the same time, in terms of immediate and delayed retention.

1. Effects of glossing on immediate retention

How different types of glossing, i.e. glossing in Chinese (C), glossing in English (E) and glossing in both Chinese and English (CE), exert effects on the breadth and the depth of word knowledge in the process of incidental vocabulary acquisition in the immediate retention testing part is to be analyzed and studied in this part.

1.1 Effects of glossing on the breadth of word knowledge in immediate retention

The following table (Table 4-6) presents us the results of three different types of glossing in BT1:

TABLE 4-6  
DESCRIPTIVE STATISTICS OF BT1

|    | N  | Mean | S.D. | Min. | Max. |
|----|----|------|------|------|------|
| CE | 60 | .406 | .300 | .00  | 1.00 |
| E  | 60 | .386 | .321 | .00  | 1.00 |
| C  | 60 | .531 | .303 | .00  | 1.00 |

BT1=the First Breadth Test

CE= (average of items Glossed in) Chinese and English

E= (average of items Glossed in) English

C= (average of items Glossed in) Chinese

Averages are used again in this part, for it is more convenient to illustrate the results of the test. The full score is one point for each item and from the table above it could be found that the mean of the scores in glossing in Chinese reaches 0.531, which is the highest among the three types and the means of the scores (i.e. 0.406 and 0.386 respectively) in other two types are very close to each other. On the other hand, the standard deviations of the three tests are close to one another, which demonstrates that the subjects did not differ much in the performance in each of the three types of glossing.

To compare the different effects of the three glossing types, Friedman Test for several related samples in the SPSS has been employed in this test.

The following table (Table 4-7) is taken directly from the result of Friedman Test.

TABLE 4-7  
TEST STATISTICS IN BT1

|             |       |
|-------------|-------|
| N           | 60    |
| Chi-Square  | 7.031 |
| Df          | 2     |
| Asymp. Sig. | .030  |

(Friedman Test)

This table demonstrates that the scores of the three types of glossing are significantly different from one another because the significance value reaches a level of .030, which is smaller than .05. As a result, it could be concluded that the effects of the three different types of glossing differ significantly from one another.

Taking into consideration what we have discussed about Table 4-6, among the three types of glossing, glossing in Chinese (C) is the most effective in enhancing the breadth of word knowledge in incidental vocabulary acquisition and also it reaches a significant level. Glossing in both Chinese and English (CE) comes next. The least effective is the type of glossing in English (E).

The reason why the type of glossing in both Chinese and English (CE) failed to obtain the highest retention rate in terms of the breadth of word knowledge may be owing to various reasons. First of all, some students did not read carefully or even just ignored the English definition on encountering this type of glossing, which is consistent with the finding from the interview with students after the first testing. English definitions, in general, are much longer than Chinese ones, or their Chinese "equivalents," more exactly. As a consequence, subjects were in such a hurry to finish the reading task that they may just scan the Chinese definition following its English counterpart swiftly and hastily. This is in line with what has been discovered in previous research (Xu, 2010). As has been claimed by Prince (1996: 489), this phenomenon is "one of attitude rather than of processing skill itself." He has further pointed out that it is needed to "remedy the situation not only by practice at developing the appropriate skills but also by working upon learners' perceptions of long-term aims and of the ways the task at hand contributes to them (ibid)."

On the other hand, for most of the subjects glossing in mother tongue is easier to understand and more convenient to memorize. Apart from this reason, it might be probable that without the long English definition preceding the short Chinese definition, subjects may experience a lower degree of anxiety while reading the gloss.

Glossing in English (E) is the least effective one. For one thing, just as has been mentioned above, English definitions are usually much longer than its Chinese counterparts. Therefore, it is more demanding for subjects to understand the English definitions and even more difficult to commit them to memory. For another, this might increase the degree of anxiety of subjects, thus forming a "vicious circle".

In a word, in the breadth test of immediate retention of incidental vocabulary acquisition, glossing in Chinese (C) is the most effective, and glossing in both Chinese and English (CE) is the second most effective, and the poorest is glossing in English (E).

#### 1.2 Effects of glossing on the depth of word knowledge in immediate retention

Following the same pattern, another aspect, i.e. the aspect of the depth of word knowledge, is to be examined in this part. First of all, the results from the depth test of word knowledge (with four points as its full score) and those from Friedman Test are presented respectively in the following two tables (Table 4-8 and Table 4-9):

TABLE 4-8  
DESCRIPTIVE STATISTICS OF DT1

|    | N  | Mean  | S.D. | Min. | Max. |
|----|----|-------|------|------|------|
| CE | 60 | 1.99  | .541 | .17  | 2.67 |
| E  | 60 | 1.825 | .606 | .50  | 2.83 |
| C  | 60 | 1.689 | .761 | .17  | 2.67 |

DT1=the First Depth Test

CE= (average of items Glossing in) Chinese and English

E= (average of items Glossing in) English

C= (average of items Glossing in) Chinese

TABLE 4-9  
TEST STATISTICS IN DT1

|             |        |
|-------------|--------|
| N           | 60     |
| Chi-Square  | 10.545 |
| Df          | 2      |
| Asymp. Sig. | .005   |

(Friedman Test)

As could be seen from Table 4-8, the type of glossing in both Chinese and English (CE) is the highest in mean, next is the glossing-in-English (E) type and glossing in Chinese (C) is in the last place. From Table 4-9, it could be further concluded that this difference is rather significant, for the significance value reaches as high as  $.005 < .01$ .

Put it in other words, glossing in both Chinese and English (CE) is the most effective in enhancing the depth of word knowledge in the course of incidental vocabulary acquisition. And glossing in English (E) is less effective than glossing in both Chinese and English (CE), but more effective than glossing in Chinese (C) at a significant level.

The reasons for these results may be various. In the first place, mother tongue is easy to understand and also helpful in committing things to memory and using mother tongue is certainly the most rapid way (Prince, 1996). The acquisition of L2 words has been said to more often than not involve a mapping of the new word form onto pre-existing conceptual meanings or onto L1 translation equivalents as approximations (Ellis, 2002). Secondly, English definitions could provide an exact, clearer and rather full explanation for the target words, thus helping subjects build a network of words in the semantic field and acquiring more aspects of the target words. Also glossing in both Chinese and English (CE) is conducive to understanding the text and also the target words. What is more important, just as Albert and Obler (1978 cited in Swan, 2002) have argued, words in one language, and their translation equivalents in the other, if there are any, are connected in the brain in a nonrandom way, much as a word and its synonym within the same language may be related in an associational network. That is to say, the mental lexicon of L1 and L2 may be connected with each other and together enhance the acquisition of words being learned. As a consequence, glossing in both Chinese and English (CE) is the most effective in enhancing the depth of word knowledge acquisition in the process of reading the text.

Glossing in English (E) ranks second in terms of the depth of word knowledge acquisition, for there is no mother tongue to help them with comprehension and neither connection nor association has occurred between L1 mental lexicon and L2 mental lexicon.

However, on the other hand, the so-called Chinese definitions of the target words are in most cases just their Chinese “equivalents”. These equivalents could not fully convey even the meaning of the target words, to say nothing of the knowledge about other aspects of target words. What is even worse, subject may misunderstand some aspects of the target words, thus hindering the acquisition of the depth of word knowledge. In addition, persistent reliance on L1 has been claimed to be probably one of a complex cluster of factors that lead to ineffective L2 learning and that this reliance stems largely from a desire to understand quickly (Prince, 1996). Therefore, it could be easily understood why words that have been glossed solely in Chinese (C) have been worst acquired in terms of the depth of word knowledge.

In conclusion, glossing in both Chinese and English (CE) is the most effective one in enhancing the depth of word knowledge in terms of the immediate retention, next is glossing in English (E), and glossing in Chinese (C) is the poorest of the three. In addition, this difference in effectiveness has reaches a rather significant level.

## 2. Effects of glossing on delayed retention

How the three different types of glossing affect the delayed retention of the two respective aspects of the breadth and the depth of word knowledge is to be explored in detail in this part.

### 2.1 Effects of glossing on the breadth of word knowledge in delayed retention

Table 4-10 and Table 4-11 in the below demonstrate the results of the second breadth test of word knowledge (with one point as its full score) and the results from Friedman Test.

TABLE 4-10  
DESCRIPTIVE STATISTICS OF BT2

|    | N  | Mean | S.D. | Min. | Max. |
|----|----|------|------|------|------|
| CE | 60 | .267 | .209 | .00  | .83  |
| E  | 60 | .233 | .228 | .00  | .83  |
| C  | 60 | .253 | .256 | .00  | .83  |

BT2=the Second Breadth Test

CE= (average of items Glossing in) Chinese and English

E= (average of items Glossing in) English

C= (average of items Glossing in) Chinese

TABLE 4-11  
TEST STATISTICS IN BT2

|             |      |
|-------------|------|
| N           | 60   |
| Chi-Square  | .497 |
| Df          | 2    |
| Asymp. Sig. | .780 |

(Friedman Test)

From Table 4-10, it could be seen that different from what has been found in the first breadth test or the immediate breadth test, the mean of the scores for the words glossed in both Chinese and English (CE) is the highest among the three. The type of glossing solely in Chinese (E) falls to the second place. And the same as the results in the immediate testing session, glossing in English (E) is the least effective one among the three. However, the difference in the breadth of word knowledge in delayed retention testing is not significant, for the significance value is .780, much higher than .05.

In other words, there is no significant difference among the three types of glossing in the delayed test of the breadth of word knowledge.

## 2.2 Effects of glossing on the depth of word knowledge in delayed retention

Following the same pattern as the previous one, this part is to explore how the other aspect, the aspect of the depth, is affected by the three different types of glossing.

Table 4-12 in the below demonstrate the results of the second depth test of word knowledge (with four points as its full score), and Table 4-13 the results from Friedman Test.

TABLE 4-12  
DESCRIPTIVE STATISTICS OF DT2

|    | N  | Mean  | S.D. | Min. | Max. |
|----|----|-------|------|------|------|
| CE | 60 | 1.386 | .765 | .00  | 2.83 |
| E  | 60 | 1.119 | .892 | .00  | 2.83 |
| C  | 60 | .925  | .968 | .00  | 2.67 |

DT2=the Second Depth Test

CE= (average of items Glossing in) Chinese and English

E= (average of items Glossing in) English

C= (average of items Glossing in) Chinese

TABLE 4-13  
TEST STATISTICS IN DT2

|             |        |
|-------------|--------|
| N           | 60     |
| Chi-Square  | 26.995 |
| Df          | 2      |
| Asymp. Sig. | .000   |

(Friedman Test)

From Table 4-10, we could see the standard deviation of the type of glossing in both Chinese and English (CE) is the lowest, which entails that students have varied least in the performance of the depth of word knowledge test. And in addition, at the same time, the mean of CE is also the highest. And the same with the first depth test (DT1), glossing in English (E) is placed in the second place, and glossing solely in Chinese (C) is the last one. And this difference is very significant, for the significant value has reaches a level much lower than .01.

In other words, glossing in both Chinese and English (CE) is the most effective in the delayed retention of the depth of word knowledge at a quite significant level. In addition, glossing in English (E) is the second most effective one in enhancing this aspect of word knowledge in delayed testing. Furthermore, glossing in Chinese (C) is the least effective one.

The reasons for the most effectiveness of glossing in both Chinese and English (CE) may be explained from several aspects. First of all, just as has been mentioned in the part discussing the first depth test of word knowledge, it is very beneficial to have both English and Chinese, the mother tongue, in command to better understand the text and the target words. In the second place, the definitions in the two languages may compensate with each other, one easy to understand and remember and the other exact, precise and accurate, thus enhancing the acquisition of more aspects of

word knowledge. Last but not least, when two languages are related or associated with each other in some way, the L1 and L2 may interact with each other, thus connecting the two languages together and strengthening the network of words within the target language and between the two languages.

As to the type of glossing in English (E), it could be said that since only English is available to them, subjects could only rely on it. Some may not understand the meaning of some English definitions very well. On the other hand, networks of the semantic field could be built, but only within the English language, rather than between the English and the Chinese languages. Thus the connections of any kind may not be as strong and firm as those in the case of glossing in both Chinese and English (CE). As a result, glossing solely in English (E) is less effective than glossing in both Chinese and English (CE) in enhancing retention rate of the depth of word knowledge.

The reason why glossing solely in Chinese (C) is the least effective one lies in the fact that though the mother tongue is the quickest and easiest way to learn the aspect of meaning of new words, and in most cases the superficial meaning, it is far from accurate and exact in explaining the new words. Subjects may miss some points or misunderstand the meaning of the new word. What is more, since it is very easy to commit meanings of new words in Chinese to memory, little effort has been made, thus increasing the possibility of forgetting. As a consequence, the type of glossing in Chinese is the poorest in enhancing delayed retention of the depth of word knowledge.

In a word, from the aspect of the depth of word knowledge in the delayed testing, one aspect of our second research question has been answered, i.e. the type of glossing in both Chinese and English (CE) is the most effective one in enhancing the delayed retention of the depth of word knowledge, and glossing in English (E) ranks second, and glossing in Chinese is the poorest in this aspect.

## V. CONCLUSION

The major findings of the present study are illustrated as follows:

First of all, a better understating of the reading passage does contribute to a higher rate of incidental vocabulary acquisition in terms of the breadth of word knowledge. However, a better understanding of the text does not guarantee a higher rate of the development of the depth of word knowledge in the course of incidental vocabulary acquisition.

Secondly, glossing in both Chinese and English (CE) is the most effective in enhancing the depth of word knowledge in both the immediate retention and in the delayed retention.

Furthermore, glossing in Chinese (C) is the most effective in the immediate retention of the breadth of word knowledge. But this type is the least effective in enhancing the depth of word knowledge in both the immediate and the delayed retention.

Moreover, glossing in English (E) is more effective than glossing in Chinese (C) in the depth of word knowledge in both the immediate and the delayed retention. However, the effect of this type of glossing is the poorest in enhancing the breadth of word knowledge in the immediate retention.

Lastly, no significant difference has been found in the delayed retention of the breadth of word knowledge among these three different types of glossing.

As could be concluded from the major findings, the type of glossing in Chinese is the most effective in enhancing the immediate retention of the breadth of word knowledge; and the type of glossing in both Chinese and English is the most effective in enhancing the depth of word knowledge in both the immediate and the delayed retention.

Therefore, here come the theoretical implications from the present study. First of all, this study sheds some light on the study of the depth of word knowledge and on how to enhance the incidental vocabulary acquisition through reading. Moreover, the English language and the Chinese language have the potential to help build networks between them, and similar with the view of the similarity in structure of L1 and L2 mental lexicon found by Wolter (2001), L1 and L2 mental lexicon may be connected, or related in some way, with each other and facilitate the learning of new words.

On the other hand, there also exist practical implications for editors in the field of English textbook or English reading material publishing and for teachers in foreign language teaching.

For editors of English textbooks, it is necessary to specify different purposes of each part. For example, as for fast-reading parts in textbooks, the purpose is to enhance students' reading speed and train their reading skills. And a higher rate of the immediate retention of the breadth of word knowledge is essential for fast-reading parts. Therefore, for the fast-reading part in textbooks, the type of glossing in Chinese is recommended. As with the intensive reading part, with its main purpose to develop students' depth of word knowledge and productive word knowledge in a long run, it is important to provide glossing in both Chinese and English, for this type of glossing has been proved by this present study to be the most effective in enhancing the depth of word knowledge in both the immediate and the delayed retention. For the extensive reading part or after-class reading part, if its purpose is to improve students reading ability, glossing in Chinese is preferred; if its purpose is to enlarge the receptive vocabulary of students, glossing in both Chinese and English is recommended, for it is the most effective in long-term retention of the breadth of word knowledge.

The same is true for the glossing type in various reading materials in English, including novels, magazines and newspapers, it is important for editors to know the purpose of the reading material. If it is intended to train students' reading skills, glossing in Chinese is enough. If its purpose is to enlarge student's receptive vocabulary or develop students' depth of word knowledge, it is recommended to gloss in both Chinese and English.

Similarly, for English teachers, the types of explaining new words also depend on different purposes. For example, when making an explanation of important words, whose productive use is required, English teachers should make full use of definitions in both languages to help students build semantic networks between words. In addition, they should try to help students cultivate a sound habit and a correct attitude toward the English learning, and help them get rid of their habit of overdependence on the mother tongue in the long run (Prince, 1996; Swan, 2002). If only receptive knowledge of some words is required, it is also necessary to give an explanation of these words in both Chinese and English to enhance students' long-term retention of words. If the purpose is just for immediate text comprehension, a brief explanation in Chinese is enough.

In spite of the fact that great efforts have been made and much importance has been attached to the present study, there are several limitations to the present study due to the limitation of time, manpower, and the feasibility of carrying out the experiment. Consequently, in the future research, larger numbers of subjects should be included in the future experiment to improve the generalizability of the conclusion. Furthermore, subjects of different proficiency in English, including college students of different grades, junior and senior high school English learners, might be included. Last but not least, the time span between the testing sections should be lengthened to thoroughly investigate the long-term effects of different types of treatments (i.e. different types of glossing in the present case).

#### REFERENCES

- [1] Alderson, J. Charles (2007). Judging the Frequency of English Words. *Applied Linguistics*, 28, 3: 383-409.
- [2] Chen, C. and J. Truscott (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, 31, 693-713.
- [3] Chun, Dorothy M. and Jan L. Plass (1996). Effects of Multimedia Annotations on Vocabulary Acquisition. *Modern Language Journal*, 80, 2: 183-198.
- [4] Coady, J. (2001). L2 vocabulary acquisition: A synthesis of the research. In J. Coady and T. Huckin (Eds.) *Second Language Vocabulary Acquisition*. Shanghai: Shanghai Foreign Language Education Press, 273-290.
- [5] Ellis, Nick C. (2002). Vocabulary acquisition: word structure, collocation, word-class, and meaning. In N. Schmitt and M. McCarthy (Eds.) *Vocabulary: Description, Acquisition and Pedagogy*. Shanghai: Shanghai Foreign Language Education Press, 122-139.
- [6] Grabe, William and Fredricka L. Stoller (2001). Reading and vocabulary development in a second language: A case study. In J. Coady and T. Huckin (Eds.) *Second Language Vocabulary Acquisition*. Shanghai: Shanghai Foreign Language Education Press, 98-122.
- [7] Gui, Bao (2015). Task type effects on English as a Foreign Language learners' acquisition of receptive and productive vocabulary knowledge. *System*, 53, 84-95.
- [8] Laufer, B. (2001). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady and T. Huckin (Eds.) *Second Language Vocabulary Acquisition*. Shanghai: Shanghai Foreign Language Education Press, 20-34.
- [9] Laufer, B. (2002). What's in a Word That Makes it Hard or Easy: Some Intralexical Factors that Affect the learning of Words. In N. Schmitt and M. McCarthy (Eds.) *Vocabulary: Description, Acquisition and Pedagogy*. Shanghai: Shanghai Foreign Language Education Press, 140-155.
- [10] Nation, I. S. P. (1990). Teaching and learning vocabulary. Boston: Heinle & Heinle.
- [11] Prince, P. (1996). Second Language Vocabulary Learning: The Role of Context versus Translations as a Function of Proficiency. *Modern Language Journal*, 80, 4: 478-493.
- [12] Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic performance: an assessment perspective. *Language Learning*, 52, 3: 513-536.
- [13] Qian, D. and Mary Schedl (2004). Evaluation of in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, 21: 28-52.
- [14] Saz, Oscar, Lin, Yibin and Maxine Eskenazi (2015). Measuring the impact of translation on the accuracy and fluency of vocabulary acquisition of English. *Computer Speech and Language*, 31, 1: 49-64.
- [15] Swan, M. (2002). The influence of the mother tongue on second language vocabulary acquisition. In N. Schmitt and M. McCarthy (Eds.) *Vocabulary: Description, Acquisition and Pedagogy*. Shanghai: Shanghai Foreign Language Education Press, 156-180.
- [16] Tanaka, Mitsuko (2017). Examining EFL vocabulary learning motivation in a demotivating learning environment. *System*, 65:130-138.
- [17] Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27: 33-52.
- [18] Webb, S. (2007). The Effects of Repetition on Vocabulary Knowledge. *Applied Linguistics*, 28, 1: 46-65.
- [19] Wesche, M. B. and T. S. Paribakht (2000). Reading-Based Exercises in Second Language Vocabulary Learning: An Introspective Study. *The Modern Language Journal*, 84, 2: 196-213.
- [20] Wolter, B. (2001). Comparing the L1 and L2 Mental Lexicon. *SSLA*, 23: 41-69.
- [21] Xu, Xiaohui. (2010). The Effects of Glosses on Incidental Vocabulary Acquisition in Reading. *Journal of Language Teaching and Research*, 1, 2: 117-120.
- [22] Zeeland, Hilde van and Norbert Schmitt (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, 41: 609-624.

**Shan Liu** was born in Kaifeng, China in 1983. She received her master's degree in foreign linguistics and applied linguistics from Zhejiang University, China in 2008.

She is currently a lecturer in Hangzhou College of Commerce, Zhejiang Gongshang University, Hangzhou, China. Her research interests include second language acquisition, foreign language teaching and research, translation research and the study of the history of the United Kingdom.