

L1 AND L2 GLOSSES: THEIR EFFECTS ON INCIDENTAL VOCABULARY LEARNING

Makoto Yoshii

Prefectural University of Kumamoto

ABSTRACT

This study examined the effectiveness of L1 and L2 glosses on incidental vocabulary learning in a multimedia environment. The investigation included the effects of additional pictorial cues in L1 and L2 glosses, and how these additions affect vocabulary learning. The analyses of a mixed design repeated measures 2 (L1, L2) X 2 (picture, no picture) X 2 (immediate test, delayed test) analysis of variance (ANOVA) indicated no significant differences between L1 and L2 glosses for definition-supply and recognition tasks and showed significant differences between picture (text-plus-picture) and no-picture (text-only) glosses for definition-supply test only. The results also revealed significant interaction effects between languages and tests indicating that L1 and L2 groups showed different patterns of vocabulary retention over time. Findings suggest that both L1 and L2 glosses are effective for incidental vocabulary learning, but long-term retention may differ between the two types; and that the effect of additional visual cues on vocabulary learning may rely on the nature of the tasks given.

INTRODUCTION

Researchers have examined the effectiveness of glosses on incidental vocabulary learning (Hulstijn, 1992, 1993; Hulstijn, Hollander, & Greidanus, 1996; Jacobs, Dufon, & Hong, 1994; Knight, 1994; Laufer & Shumueli, 1997; Mondria & Wit-de Boer, 1991; Paribakht & Wesche 1996, 1997; Watanabe, 1997). Comparison of gloss conditions with non-gloss conditions has revealed the advantage of having glosses for enhancing incidental vocabulary learning. The question, therefore, has shifted from whether or not glosses are useful for incidental vocabulary learning to which gloss type is most effective; some researchers have further investigated the effectiveness of different formats in text glosses (Gettys, Imhof, & Kautz, 2001; Grace, 1998, 2000; Nagata, 1999; Watanabe, 1997). Watanabe (1997), for example, compared the effects of single glosses (regular gloss in L2) with multiple-choice glosses (two alternatives in L2) in his investigation of the effectiveness of different text modifications. He conducted his experiment with printed materials, and the study did not show any differences between the two gloss types. Nagata (1999) replicated Watanabe's study using computerized materials and found that multiple-choice glosses in L1 were better than single glosses in L1. Grace (1998, 2000) examined the effects of sentence-level L1 translations on incidental vocabulary learning, and, instead of typical word definitions or explanations, she used sentence-level information (translation of the sentence in which a certain target word appears). She discovered that the translation glosses were highly effective. Gettys et al. (2001) also compared the sentence-level L1 translation glosses with basic dictionary form L1 glosses and found that the basic dictionary glosses were more effective than sentence-level translation glosses. While these researchers examined the different text formats exclusively either in L1 or L2 conditions, others have looked into the comparison of L1 with L2 glosses.

LITERATURE REVIEW

L1 vs. L2

Attempts to compare the effectiveness of L1 and L2 glosses have brought mixed results, some indicating no difference between the two types and others suggesting the advantage of one gloss type over the other type (Chen, 2002; Jacobs et al., 1994; Miyasako, 2002). Jacobs et al. (1994), for instance, compared L1 with L2 glosses with 85 English-speaking participants who were studying Spanish as a second language. They read a Spanish text with 613 words under three conditions: (1) L1 (English) gloss; (2) L2 (Spanish) gloss; and (3) No gloss. After reading the text with 32 words or phrases glossed and presented in boldface, the participants received two vocabulary tests unexpectedly: one immediately after the reading and the other four weeks later. The results of the immediate test showed that the gloss conditions (either L1 or L2) were better than no gloss; however, the results did not indicate any significant difference between L1 and L2 glosses. The questionnaire also revealed that the participants preferred L2 glosses to L1 glosses. Chen (2002) examined L1 and L2 glosses with 85 college freshmen in Taiwan who were studying English as a second language. The participants were divided into three groups - (1) L1 (Chinese) gloss; (2) L2 (English) gloss; and (3) No gloss - and read a 193 word-English text with 20 target words being glossed. The results showed that the L2 group outperformed the no gloss group, and that the difference between L1 and L2 glosses was not significant. Chen also pointed out that the L2 gloss group took a longer time for reading the text than the L1 gloss group did.

While the two studies above showed no difference between L1 and L2 glosses, Miyasako (2002), on the other hand, revealed the advantage of one gloss type over the other. He looked into the effectiveness of multiple-choice glosses and single glosses and compared the effectiveness of L1 and L2 glosses. The research design contained six groups: (1) L2 (English) multiple-choice gloss; (2) L1 (Japanese) multiple-choice gloss; (3) L2 (English) single gloss; (4) L1 (Japanese) single gloss; (5) No gloss; and (6) Control (no reading). The participants were 187 Japanese high school students, and they read a 504-word text with 20 target words. The students took two vocabulary tests, one immediately after reading and the other 18 days later. The tests consisted of multiple-choice questions. Students saw each target word in context and had to find the most appropriate definition written in English out of four choices. Miyasako found that the L2 gloss groups (multiple-choice or single) outperformed the L1 gloss groups (multiple-choice or single) significantly for the immediate test. However, the multiple-choice and single gloss types did not differ in their effect on vocabulary learning. The researcher also found that L2 glosses tended to be more effective for higher-proficiency level learners, whereas L1 glosses were more effective for lower-proficiency learners.

Although Miyasako (2002) suggested the advantage of L2 over L1 glosses, especially for higher-proficiency learners, Jacobs et al. (1994) and Chen (2002) did not find any differences in the effectiveness of L1 and L2 glosses for enhancing incidental vocabulary learning. The results of the comparisons between L1 and L2 are still few and inconclusive, and we need to further investigate the effectiveness of L1 and L2 glosses and to examine which gloss type is more effective in what conditions and for which learners.

Effects of Multimedia Glosses

Advances and increased availability of computers have altered and expanded the possibilities of gloss formats. Glosses are no longer limited to only verbal forms but now multimedia forms such as pictures, videos, and sounds are also available. Researchers, therefore, have started looking into the effectiveness of multi-mode gloss types such as pictures and videos for reading comprehension (Aweiss, 1994; Chun & Plass, 1996; Davis & Lyman-Hager, 1997; Hong, 1997; Lomicka, 1998; Lyman-Hager & Davis, 1996; Taylor, 2006) and for incidental vocabulary learning (Al-Seghayer, 2001; Chun & Plass, 1996; Kost, Foss, & Lenzini, 1999; Plass, Chun, Mayer, & Leutner, 1998; Yeh & Wang, 2003; Yoshii & Flaitz, 2002). A theoretical base for these studies lies in the Dual Coding theory proposed by Paivio (1971, 1990)

and Sadoski & Paivio (2001). The theory assumes that two separate symbolic systems function in memory and cognition: "one specialized for the representation and processing of information concerning nonverbal objects and events, the other specialized for dealing with language" (Paivio, 1990, p. 53). The theory also assumes that the two systems can function independently, but that they are also interconnected. Paivio (1990) points out that the more learners associate target words with appropriate nonverbal referents (pictures, objects, events, and emotions), the richer and more meaningful become their interconnections between verbal and visual systems. As a result, learners have better recall and appropriate use of the words when they learn them coded dually (verbally and visually) than when the words are coded in a single manner (one or the other). Therefore, additional pictorial cues are effective and efficient in helping learners make associations between pictures and words (Oxford & Crookall, 1990, pp. 16-17) and the following multi-mode gloss studies have supported the theory.

A number of studies have investigated whether multi-mode glosses are better than single-mode, and if so, which combination of multi-mode glosses is most effective. Chun & Plass (1996), for instance, conducted a series of studies on multimedia glosses and vocabulary acquisition. The participants, 160 university students who were learning German as a second language, read a 762-word text on the computer and took a vocabulary test and a recall test. The research employed a within-subjects design and all the participants read the text using the same reading program with some words glossed with text-only, some with text and pictorial cues, and some with text and video clips; all the textual cues were written in L1. The researchers found that the combination of text and picture glosses was more effective than text-only or text-plus-video glosses.

The advantage of multi-mode glosses over single-mode glosses was confirmed in Plass et al. (1998). The study also examined the effects of gloss types in a multimedia context investigating whether learners are more likely to learn target words when they received both verbal (textual) and visual (picture) glosses than when they received one mode or none. The participants were 103 American university students who were studying German as a second language and they read the same 762-word text on computer as used in Chun & Plass (1996). The text contained 24 target words of which 12 had text and picture gloss options, and the other 12 had text and video options, all the textual cues being provided in L1. The vocabulary posttest required the learners to supply L1 translation for each target word. The researchers found that the performance was the best when the learners selected both visual and verbal modes of glosses, moderate when they selected only one mode, and worst when they selected none. The results also showed that the selection of text and picture was better than that of text and video for learning the words incidentally.

In order to confirm the effectiveness of the combination of text and picture glosses over that of text and video glosses, Al-Seghayer (2001) compared the two gloss combinations. The participants were 30 English as a Second Language (ESL) students and they read a 1,300-word text with 21 target words. Among them, seven words were glossed with text only, seven words with text and video, and seven with text and picture glosses; the textual cues were written in their L2 (English). The students received the immediate vocabulary tests with both recognition and production tasks. The results showed that the combination of text and video was more effective than that of text and picture.

We tend to associate multi-mode glosses with computers, but printed materials can certainly take advantage of such glosses with the use of pictures. Kost, Foss, & Lenzini (1999) specifically used printed texts and analyzed the three types of glosses: (1) text-only (L1) gloss; (2) picture-only gloss; and (3) text (L1)-plus-picture. The participants were 56 American university students who were studying German as a second language. They read a 272-word text which contained 20 glossed words under three treatment groups with different gloss types. The participants took two vocabulary tests on 14 target words, first immediately after and the second one two weeks later. The tests consisted of both recognition (multiple-choice of definitions) and production (supply definitions), and the overall results showed that the combination of text and picture glosses was better than the picture only or text only glosses.

The results were further confirmed in a computer environment by Yoshii & Flaitz (2002). Their work also focused on the three types of glosses which Kost et al. (1999) had previously examined, this time in a computer reading environment. 151 ESL students read a 392-word text with 20 word glosses among which 14 were the target words. The participants read the text under three gloss type groups: (1) text (L2) only; (2) picture only and (3) text (L2) and picture. The researchers administered both the immediate and delayed vocabulary tests unexpectedly, which consisted of both recognition and production tasks. The results confirmed the overall advantage of the combination of text and picture glosses over text-only and picture-only types.

Another study (Yeh & Wang, 2003) on multimedia glosses was unique in two ways: first, the study included sound as a component of multimedia glosses; secondly, the researchers used both L1 and L2 in textual glosses. The study investigated the effectiveness of three multimedia type glosses: (1) text only; (2) text plus picture; and (3) text, picture, and sound. This study with 82 university students in Taiwan included both L1 (Chinese translation) and L2 (English explanation) in the textual cues. The results showed that the combination of text and picture was the most effective type among the three types.

Gloss studies above have shown that the combination of textual and pictorial cues are more effective than text-only cues for enhancing incidental vocabulary learning (Chun & Plass, 1996; Plass et al., 1998; Kost et al., 1999; Yoshii & Flaitz, 2002; Yeh & Wang, 2003). All the studies which examined the combination of textual and pictorial glosses used textual cues either in L1 or L2 exclusively, except for Yeh and Wang (2003) in which each text gloss had both L1 and L2. Therefore, we need to further compare the effectiveness of L1 with L2 glosses, to investigate the effectiveness of visual cues and to study the interaction of factors.

Bilingual lexicon

To understand the effectiveness of L1 and L2 glosses better, it is important to examine how words and concepts are represented in L2 learners' minds. Potter, So, Eckhardt, and Feldman (1984) introduced two models of lexical and semantic representations in bilingual learners' minds: the word association model and the concept mediation model. The word association model claims that L2 is mediated through L1 since new L2 words are associated with L1 translations, while the concept mediation model suggests that L2 can be mediated through concepts without L1 translations. Studies have found a developmental shift from the word association model to the concept mediation model as L2 proficiency increases (Potter et al., 1984; Chen & Leung, 1989; Kroll & Curley, 1988). In other words, L2 learners rely on word-to-word links (lexical links) in early stages, but as their L2 proficiency develops, they link L2 directly to concepts (conceptual links). With these results, Kroll & Stewart (1994) revised their model to include both the lexical links of the word association model and the conceptual links of the concept mediation model, and delineated the developmental shift through differentiation of the strength of the conceptual links. [Figure 1](#) shows this revised model adding an "Image," which was part of the original model (Potter et al., 1984). This model suggests that L1 conceptual links are stronger than L2 conceptual links because L2 words are associated with their L1 translations for accessing meaning in early stages of L2 acquisition; and a direct link to concepts from L2 words can be possible only with increasing proficiency (Kroll & Sunderman, 2003:114).

The current research uses the model in [Figure 1](#) as a theoretical base, and through this model, one could expect that the L1 glosses would be more effective than L2 glosses for vocabulary learning since the word-to-concept connections are stronger for L1 than for L2 for the participants in this study who are regarded as intermediate learners. One can also recapture the concept of the dual-coding hypothesis in the model; images provide another source of conceptual linkage and consequently strengthen links between words and concepts.

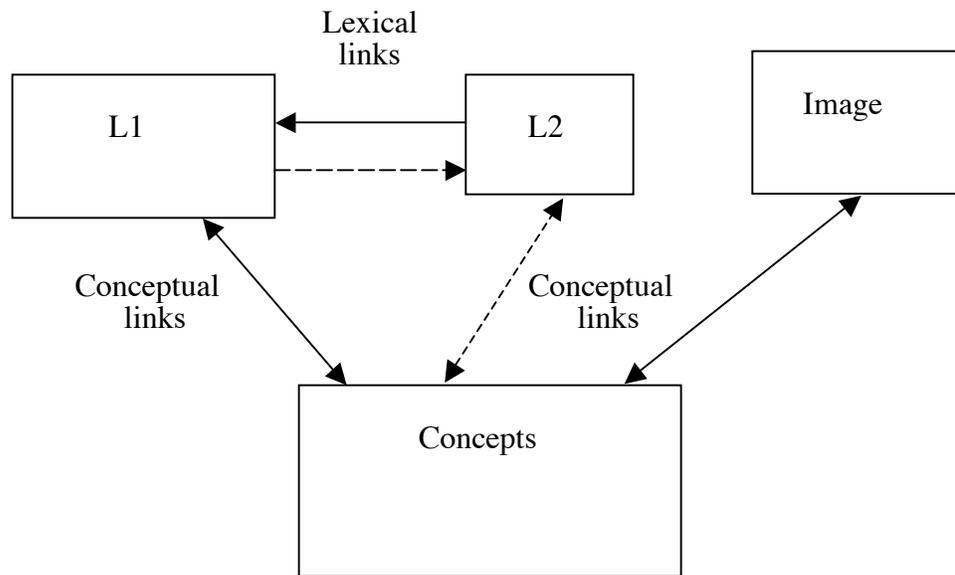


Figure 1. Revised hierarchical model with image
Source: Adapted and modified from Kroll & Stewart (1994)

RESEARCH QUESTIONS

This study, therefore, addresses the following research questions:

- 1) Do L1 and L2 glosses differ in their effectiveness on incidental vocabulary learning?
- 2) Do picture glosses (text-plus-picture) and no picture glosses (text-only) differ in their effectiveness on incidental vocabulary learning?
- 3) Are there any interaction effects among the three factors: languages (L1 or L2), pictures (presence or absence), and tests (immediate and delayed)?

METHOD

Participants and Design

A total of 195 students from two universities in Japan participated in the study: 124 students from one university and 71 from the other. The participants consisted of 130 freshmen, 29 sophomores, and 36 juniors who were learning English as a foreign language. The study started with 208 participants originally; however, only those who participated in both immediate and delayed tests were counted. Some students were also excluded from the study due to high scores on the pretest. A total of 195 participants remained in the study as a result. This study employed a mixed design repeated measures 2 (languages) X 2 (pictures) X 2 (tests) analysis of variance (ANOVA), thus including two between-subject factors – languages (L1, L2) and pictures (picture, no picture) – and one within-subject factor – tests (immediate, delayed).

Independent and Dependent Variables

The independent variable was the gloss type in the reading program on the Internet. The gloss types included: (1) L1 text only; (2) L2 text only; (3) L1 text plus picture; and (4) L2 text plus picture. [Figure 2](#) shows an example of the gloss types.

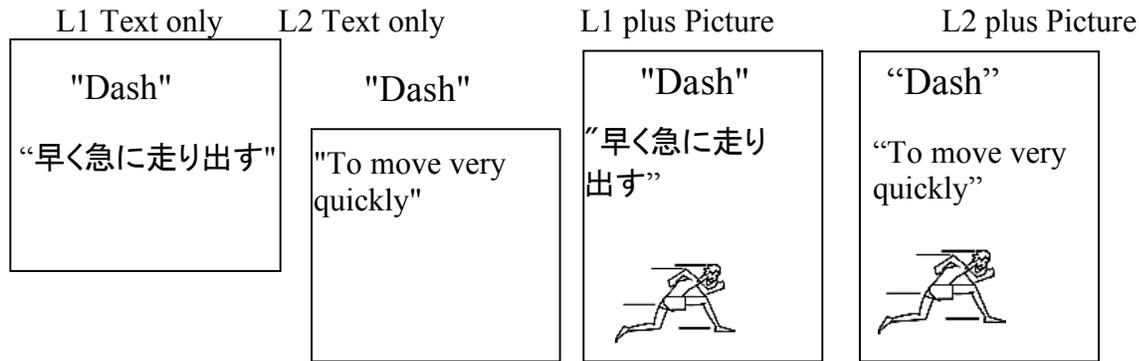


Figure 2. Gloss Types

The dependent variable was students' scores measured by the immediate and delayed vocabulary tests, and each test had both definition-supply and word recognition.

Instruments

Pretest

The participants took a pretest one week prior to the treatment. The test contained 14 target words and 10 additional distracters which were computer related words. The participants were instructed to put a check mark by any words they knew and provide a short written explanation in L1.

Reading Material and Target Words

The participants read a 390-word story including 14 target words and 6 distracters (familiar words) highlighted in the text. As the students clicked on a word, a gloss appeared on the right-hand side of the screen. In case of text plus picture glosses, the students saw both glosses at the same time on the screen. A log-file compiled the record of lookup behavior as the learners clicked on words. An example of the reading material is provided in [Appendix A](#). Studies in the past often focused on multiple parts of speech such as nouns, adjectives, adverbs, and verbs for target words. However, the current study focused on verbs exclusively to narrow the variables affecting the results of the study. Previous studies have shown that students' learning response may vary according to the difficulty level of different parts of speech (Laufer, 1990; Ludwig, 1984). Laufer (1990) points out, for example, "Nouns seem to be the easiest; adverbs -- most difficult; verbs and adjectives - somewhere between" (p. 298). This study used concrete verbs since the researcher needed to illustrate them effectively. The words were non-cognate to the participants.

Posttests

The participants received two vocabulary posttests: one immediately after the treatment and the other two weeks later. This study used a two-week span following similar studies (Chun & Plass, 1996; Kost et al., 1999; Yoshii & Flaitz, 2002). Each posttest contained two tests: a definition-supply and a recognition test. The participants did not know that they would be given a vocabulary test.

The format of the definition test was the same as the pretest in which the participants put a check mark by a word they remembered and supplied the meaning in their L1 (Japanese). The test included a total of 14 target words, and the students received one point for each correct answer, making the maximum score of the test 14 points. Two raters, the researcher himself and another high-proficient bilingual speaker, examined the answers. When discrepancies in scoring occurred, the two discussed them, then, upon agreement, a point was given or not given for the discussed item.

The recognition test also contained all 14 target words with four multiple-choice answers for each item. The participants received one point for each correct answer, which amounted to 14 maximum points in total for the test. The participants selected the most appropriate definition of a given word from four choices written in L2 (English). The definitions in the test came from the story but the phrasing was different from that used in the glosses to avoid participants' answering the questions from the mere memory of having seen them in the glosses without really having understood what they meant.

Both the immediate and delayed tests had the exact same format except for item order in each test. Samples of the posttests are provided in [Appendix B](#).

Procedure

The experiment took place during regular class periods. One week prior to the study, the students received a pretest on the target words along with the other computer related words as distracters. The students were randomly divided into four groups using four different gloss types: the first group read a story with L1 text-only glosses (n=47); the second group read it with L2 text-only (n=48); the third group used L1 text-plus-pictorial cues in the glosses (n=50); and the fourth group had as their glosses L2 text-plus-pictorial cues (n=50). The students did not know that they were in different groups under different gloss conditions. In the next class, the students read a story for comprehension and gave their feedback on the reading activity using the computer. The researchers gave a brief oral introduction to the class about the reading activity and the availability of glosses. The participants worked on this reading activity individually on their own computer and took a comprehension test. After the comprehension test, they unexpectedly had the immediate vocabulary tests which consisted of recognition and production tests. The students finished the production test first, supplying the definition in L1 to the 14 target words. Then, they received the recognition test, selecting the appropriate meaning out of the four choices written in L2. This order was chosen so that the recognition test would not give any hints for the definition-supply test. Finally, the students filled out a questionnaire form giving their feedback on the reading activity, though these results are not mentioned in this study as they pay no part in the focus of this study.

The students took the vocabulary tests two weeks later without any advanced notice. The test formats were exactly the same as the immediate tests with the definition-supply and recognition parts. The only difference lay in the order of the test items in each productive and recognition test.

Data Analysis

To examine the effectiveness of L1 and L2 glosses and the presence or absence of pictures on vocabulary test scores (immediate and delayed tests), a mixed design repeated measures 2 X 2 X 2 analysis of variance (ANOVA) was conducted on the definition-supply and recognition tests. The analysis included two between-subject factors – languages (L1, L2) and pictures (picture, no picture) – and one within-subject factor – tests (immediate, delayed). When the results showed significant differences for language and picture factors, post hoc analyses followed in order to further examine the differences.

RESULTS

Pretest Results

The results of the pretest ensured the equivalence of the groups' pre-knowledge of the target words. The cutoff point was at 30% (4 words out of 14 target words) in order for the average score of the pretest to be less than one point and this led to a removal of several students whose scores exceeded 5 points. As a result, the participants' mean score on the pretest turned out to be .76, keeping their pre-knowledge of the words to a minimum.

A one-way ANOVA was performed for the scores with a significance level set at .05. The results showed no significant differences among the groups ($F = 1.00, p = .39$) indicating that all four groups were essentially equal in their prior knowledge of the target words.

Posttest Results

Definition-Supply Test Results

The repeated measures ANOVA did not reveal a significant difference between L1 and L2 glosses ($F = 3.08, P > .05$) as seen in Table 2. The results, on the other hand, showed a significant difference between picture and no-picture groups ($F = 6.54, p < .05, \eta^2 = .03$). The mean scores of picture groups were higher than those of no picture groups as found in Table 1, indicating that glosses with pictures were significantly more effective than glosses without pictures. To further examine the differences between picture and no-picture groups, post hoc analyses were conducted and they disclosed a significantly higher score for the L1 plus picture group than that for the L2 (text-only) one; ($p = .01$) in the immediate test, and L2 plus picture scored significantly higher than L2 text-only group ($p = .01$) in the delayed test. The definition-supply tests also displayed a significant interaction effect between languages and tests ($F = 9.45, p < .05, \eta^2 = .05$), indicating that patterns of vocabulary retention differed for L1 and L2 gloss groups. Post hoc analyses revealed a significant difference in the decline rate of the scores between L1 plus picture and L2 plus picture groups ($p = .04$): L1 plus picture group had a significantly higher decline rate than the L2 plus picture group did.

Table 1. Means and Standard Deviations of Definition-Supply Tests.

Languages	L1				L2			
	Picture (50)		No Picture (47)		Picture (50)		No Picture (48)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Immediate	3.15	2.33	2.76	2.20	2.64	1.97	1.78	1.74
Delayed	2.16	1.67	1.91	1.69	2.42	1.55	1.44	1.35

Table 2. Repeated Measures ANOVA Summary for Languages, Pictures, and Tests in Definition-Supply Tests

Source	<i>DF</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>P</i>	η^2
Between-Subjects						
Languages	1	17.63	17.63	3.08	.08	.02
Pictures	1	37.48	37.48	6.54	.01*	.03
Lang X Pic	1	8.79	8.79	1.53	.22	.01
Error (Between)	191	1094.43	5.73			
Within-Subjects						
Tests	1	34.91	34.91	33.93	.00	.15
Lang X Tests	1	9.77	9.77	9.50	.00*	.05
Pic X Tests	1	.00	.00	.00	.95	.00
Lang X Pic X Tests	1	.46	.46	.44	.51	.00
Error (within)	191	196.48	1.03			

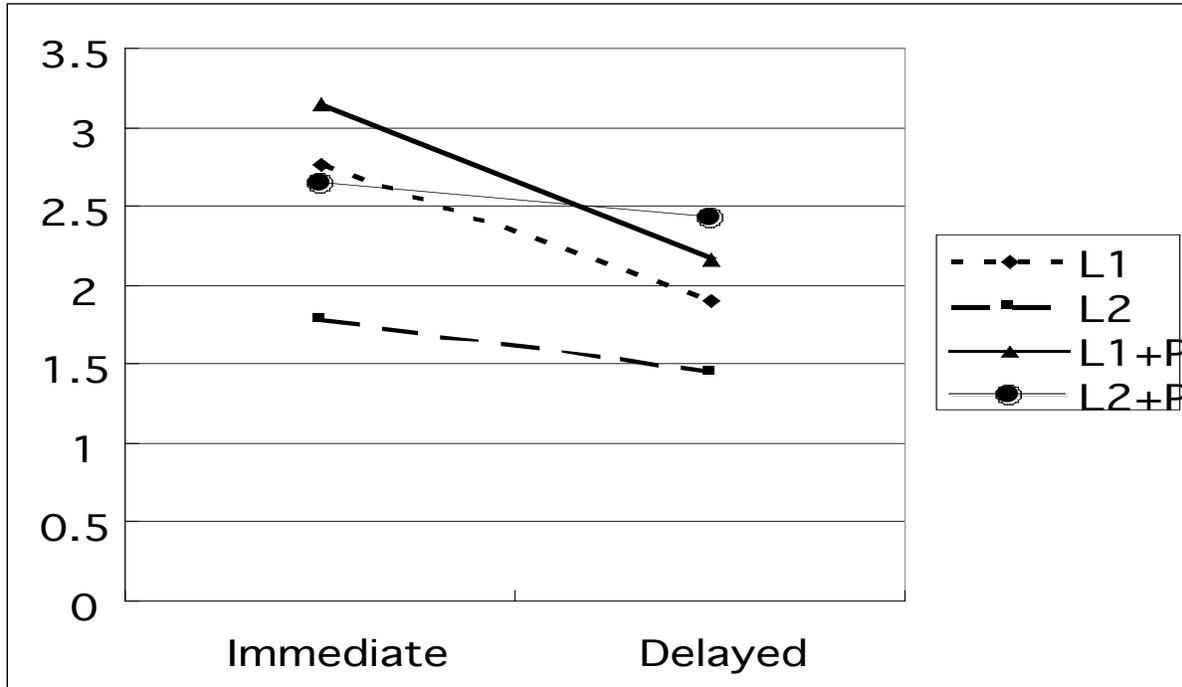


Figure 3. Immediate and delayed Definition-Supply test scores of picture and no picture in L1 and L2 gloss groups

Recognition Test Results

Table 3 depicts the descriptive statistics for the recognition test. The repeated measures ANOVA failed to reveal any significant difference between L1 and L2 glosses ($F = .08, p > .05$), or between picture and no-picture groups ($F = 3.02, p > .05$), as seen in Table 4. L1 and L2 glosses did not differ, neither did picture and no-picture glosses, in their effectiveness for recognizing words that the participants learned. The recognition tests manifested a significant interaction effect between languages and tests ($F = 6.21, p < .05, \eta^2 = .03$): patterns of vocabulary retention differed for L1 and L2 gloss groups. Figure 4 shows that the L1 text-only group displayed a rather unique pattern of retention over time: this group was able to sustain its scores, while the other three groups showed declines in their scores. Post hoc analyses showed a significant difference between L2 text-only and L1 text-only groups ($p = .04$) and between L2 plus picture and L1 text-only groups ($p = .01$). In other words, while L2 text-only and L2 plus picture groups suffered from their memory losses, participants in the L1 text-only group were able to remember the words.

Table 3. Means and Standard Deviations of Recognition Tests

Languages	L1				L2			
	Picture (50)		No Picture (47)		Picture (50)		No Picture (48)	
Pictures	M	SD	M	SD	M	SD	M	SD
Immediate	8.54	3.14	7.87	2.78	9.36	2.73	8.08	2.68
Delayed	7.60	3.22	7.98	2.81	8.02	2.78	6.96	2.80

Table 4. Repeated Measures ANOVA Summary for Languages, Pictures, and Tests in Recognition Tests

Source	DF	SS	MS	F	P	η^2
Between-Subjects						
Languages	1	1.13	1.13	.08	.78	.00
Pictures	1	42.03	42.03	3.02	.08	.02
Lang X Pic	1	25.58	25.58	1.84	.18	.01
Error (Between)	191	2661.84	13.94			
Within-Subjects						
Tests	1	66.26	66.26	25.37	.00	.12
Lang X Tests	1	16.21	16.21	6.21	.01*	.03
Pic X Tests	1	9.69	9.69	3.71	.06	.02
Lang X Pic X Tests	1	4.21	4.21	1.61	.21	.01
Error (within)	191	498.88	2.61			

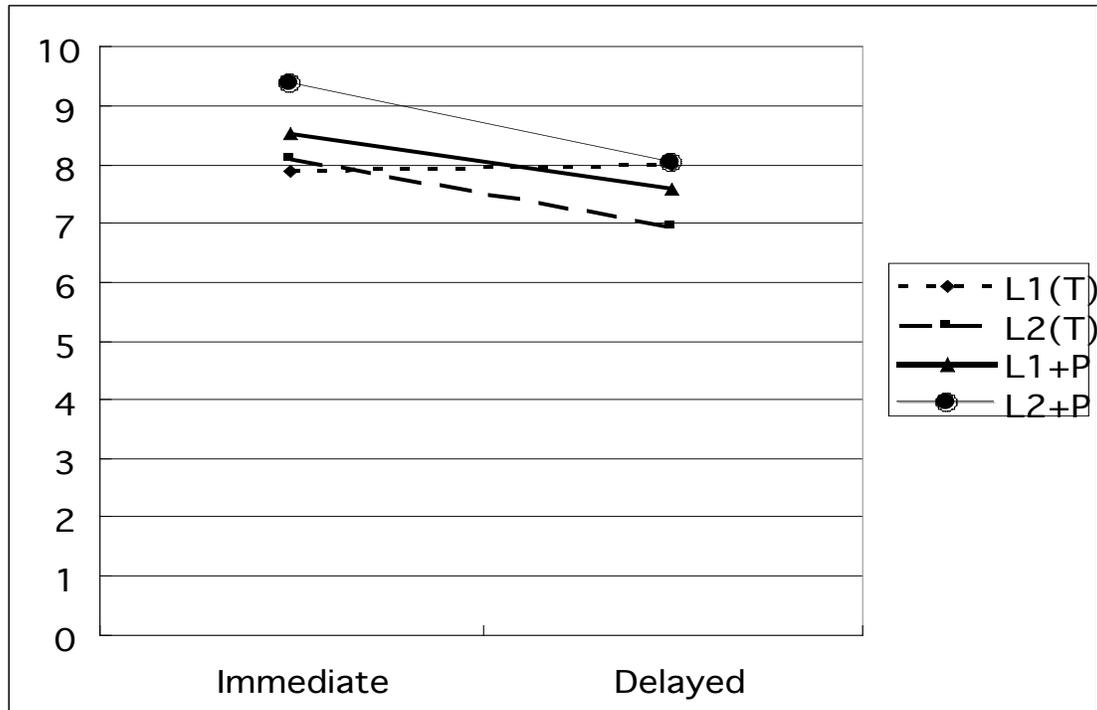


Figure 4. Immediate and Delayed Recognition Test scores of Picture and No Picture in L1 and L2 gloss groups

DISCUSSION

The first research question compared L1 with L2 glosses, and neither the definition-supply nor the recognition tests revealed significant differences between the two types. This result corresponded with those of previous studies (Chen, 2002; Jacobs et al., 1994) showing no significant differences between L1 and L2 text-only glosses. The effectiveness of glosses, either in L1 or L2, did not differ in terms of enhancing vocabulary learning. Therefore, the current study confirmed the usefulness of glosses, regardless of language types, on incidental vocabulary learning (Jacobs et al., 1994; Hulstijn, 1992, 1993; Hulstijn et al., 1996; Knight, 1994; Laufer & Shumueli, 1997; Mondria & Wit-de Boer, 1991; Paribakht

& Wesche 1996, 1997; Watanabe, 1997). One possible reason why L1 was not found to be more effective than L2 glosses might be related to the nature of incidental learning. Learners pick up the meanings incidentally as they encounter words; however, the rate of learning is still low (Nation, 2001, pp. 236-240). Even for the L1 group, it was still difficult to make conceptual links in such a short period of time on the first exposure. The fact that the L1 text-only group sustained their scores over time in the recognition tests in this study leads to a suspicion that using L1 glosses more often and over a longer period of time might reveal an advantage of L1 over L2 glosses, but the length/term of this study does not allow any clear conclusions in this respect. As the L1 text-only group sustained their scores over time in the recognition tests in this study, an advantage of L1 glosses may appear only at a later time. Another reason might have to do with the developmental shift of the bilingual lexicon from the word association model to the concept mediation model (Chen & Leung, 1989; Kroll & Curley, 1988; Potter et al., 1984) and the proficiency level of the learners of this study. We assumed that the participants were low-intermediate or intermediate levels of learners who were still in early stages where the conceptual links between L1 and concepts would be stronger than the conceptual links between L2 and concepts, and a direct link to concepts from L2 words would not be possible (Kroll & Stewart, 1994; Kroll & Sunderman, 2003). The participants might have had higher proficiency levels than expected or perhaps their levels varied, which may have contributed to the lack of differences between L1 and L2.

The second research question compared the picture (text-plus-picture) and no-picture (text-only glosses) groups; here, results differed according to the test types. The definition-supply tests implied significant difference between the two types ($F = 6.54, p < .05, \eta^2 = .03$) and thus supported the dual-coding hypothesis, but the recognition tests did not ($F = 3.02, p > .05$). We may find the reason for this inconsistency in the nature of the tasks themselves. In the definition-supply test, learners had to rely on their memory without any hints to recall the meanings, and the addition of pictures may have helped recall the meaning by assisting the memory trace. On the other hand, in the recognition test, learners were able to see the multiple choices and use them as hints for recalling the meanings. Therefore, the presence or absence of pictorial cues did not have as much effect as they did on the definition-supply tests and, thus, minimized the effect of the addition of pictures.

An interesting phenomenon emerged in this comparison of picture and no-picture glosses. As seen in [Figures 3 & 4](#), L2 seemed to benefit more than L1 did when a picture accompanied the glosses. For example, in the definition-supply test, the difference between the L2 (text-only) and L2-plus-picture groups seemed larger than that between the L1 (text-only) and L1-plus-picture groups. In fact, the post hoc analyses for the delayed definition test showed significant differences between the L2 (text-only) and L2-plus-picture ($p = .01$). In the recognition test, as seen in [Figure 3](#), the difference between L2-plus-picture and L2 (text-only) also appeared to be bigger than that between L1-plus-picture and L1 (text-only).

Why did the L2 tend to benefit from the additional pictures more than the L1 did? One possible explanation may lie in the nature of the combination of textual and visual cues. The pictorial cues of the glosses depicted the meaning of a word (a concrete verb), and the L1 textual cues were clear enough that the addition of the picture which conveyed the same information as the textual information may have not been necessary. Therefore, the addition of a picture did not appear to particularly enhance the learners' understanding of the words' meanings. However, the L2 textual information was not as easily processed as that of L1, and, as a result, the picture supplemented or strengthened the learning of the words. Therefore, answering the question of whether the combination of textual and visual information is redundant or supplemental might help in understanding the effectiveness of the text-plus picture glosses and their effect on incidental vocabulary learning. Even though these observations remain speculative, it would be worth investigating this tendency of pictures to benefit L2 glosses more than L1 glosses.

The third research question examined whether there were any significant interactions among the factors such as languages (L1 or L2), pictures (presence or absence), and tests (immediate and delayed). The

results revealed significant interaction effects between languages and tests for both the definition-supply and the recognition tests, which indicated that the scores for L1 and L2 glosses changed differently over time. Post hoc analyses disclosed that the L2 plus picture group had a significantly smaller decline rate than the L1 plus picture group ($p = .04$) did for the definition-supply tests. For the recognition tests, it was interesting to see similar declines among the three groups (the L2-plus-picture, L1-plus-picture, and L2 text-only groups) except for the L1 (text-only) group as seen in [Figure 4](#); the decline rate for the L1 text-only group was, in fact, significantly lower than that for the L2 text-only group ($p = .46$) and that of the L2 plus picture group ($p = .01$). The reason why the L1 text-only group was able to sustain their memory of words may have to do with the conceptual links the L1 group has; the advantage of having stronger links in L1 than L2 may have become more evident over time in the recognition task.

CONCLUSION

This research examined the effectiveness of different types of glosses on incidental vocabulary learning with particular focus on comparison of L1 and L2 types. We did not find differences between the two types, and this research therefore failed to support the bilingual lexicon model presented in [Figure 1](#). However, we did see some evidence for the strength of L1 in the interaction effects. We found that the groups displayed different forgetting rates over two weeks. The L1 text-only group remembered the words better than L2 text-only group or L2 text plus picture group did. Although we still do not know whether L1 or L2 glosses are better, we obtained further confirmation that glosses as a whole are useful. The study also compared the text plus picture and text-only glosses, and the advantage of having additional pictures manifested in the definition-supply test, thus partially supporting the dual-coding hypothesis.

The results of this study have implications for vocabulary learning and teaching. First, glosses are useful whether in L1 or L2 for enhancing learners' incidental vocabulary learning, and we should continue to utilize glosses in reading materials. As the significant interaction effects between languages and tests indicated, the effectiveness of L1 and L2 glosses may differ over time, and may also differ according to the kinds of instruments used (for example, in this study, either definition-supply or recognition tasks). We need to keep these factors in mind in conducting our teaching and research. We ought to examine longer-term effects and use sensitive instruments for detecting the differences. Thirdly, we would like to acknowledge the effectiveness of L1 textual cues. The L1 text-only group remembered words better than other groups did in the recognition task. This does not negate the dual-coding effect as we saw the evidence in the definition-supply test; however, it simply implies that L1 textual cues could be as effective as text-plus picture cues, as we saw in the recognition test results. We may need to recognize the strong conceptual links L1 brings and re-evaluate the efficacy of L1 textual glosses for vocabulary learning.

Future studies need to examine the effect of L1 and L2 glosses, taking the learners' proficiency level into consideration. Such studies need to investigate whether the effect of L1 and L2 glosses changes when we include learners' proficiency levels. We need to investigate whether higher level learners learn words better with L2 glosses than lower level learners and, conversely, whether lower level learners pick up words better with L1 glosses than higher level learners. Researchers need to examine how the effect of the additional pictorial cues with L1 and L2 glosses would differ according to participants' proficiency levels. The long-term effect of glosses needs to be investigated; we could administer delayed tests later than two weeks, for instance. Moreover, the definition of "long term" effect itself could be a topic of investigation. This study dealt only with concrete verbs; therefore, a future study needs to examine effects of glosses with other parts of speech and other types of words including abstract words. It would be interesting to see if there are any differences in the effectiveness of L1 or L2 glosses for abstract words and words that contain difficult concepts, as well as the effect of pictorial cues for these words. We ought to continue our investigation of glosses and examine when or in what conditions the addition of pictorial cues benefits

which learners.

APPENDIX A. STORY: "A SCARY NIGHT"

It's a cold winter night. It's midnight, and it's very quiet. I'm still awake and studying. I have a test tomorrow. I need to read two chapters. I finish one chapter and I read the next chapter. It's too difficult. I can't pass the test. What do I do? Shall I keep studying? Can I take the test some other time? Shall I give up? I'm **pondering** many things. I think my head is going to **burst**.

Suddenly, some noise **startles** me. Something **shattered** on the ground. I look at the window. Wait! What is that? I see a light across the street. It is from a new house. It's strange. Mr. & Mrs. Smith are on vacation now. They asked me to **rake** the lawn for them while they're gone. Nobody should be there. Oh, I see the light again.

Then, it **dawns on** me. Someone is **burglarizing** the house. I'm afraid. What do I do now? I have to call the police. I **dash** to the phone and call the police.

After ten minutes, the police arrive. They enter the house. As the police **search** the house, someone **hides** outside the house. The police **yell**, "Stop, right there!" But the man with a black mask runs into the woods near the house. Then, he **tumbles** down the hill in the woods. The police finally catch him. The police take off the mask. He **grins** first, then, starts to **sob**.

Two policemen come to my apartment. The first one looks very serious. He doesn't **greet** me. He just asks for my name. Then, he says, "Thank you for calling us about this problem." The other one is friendlier. He **inquires** about a couple of things. He wants to know when I first saw the light. He **scribbles** some notes.

The policemen are gone, and everything is quiet now. What a strange night! I'm glad this is over, but I am still **shivering** a little. So I **pour** some milk. This might help me. I can't study any longer and can't sleep right away.

I decide to read a book. I got it at a bookstore yesterday. The title is "American Short Stories." I look at the first chapter. And I **gape** at the title. It says, "My Life as a Burglar" by A Man with a Black Mask.

390 words in total

APPENDIX B. SAMPLES OF POSTTESTS

Definition-Supply Test

Directions: Please check any of these words you know. Please put [X] in the box. Please write the meanings either in English or your native language.

- [] sob _____
- [] shiver _____
- [] burst _____
- [] grin _____
- [] scribble _____
- [] ponder _____

Word Recognition Test

Directions: Please match the English word with the correct meaning. Put [X] in the box.

-
- | | |
|-------------------|--|
| 1. ponder | [] To study for a test
[] To think very carefully
[] To read something
[] To break open suddenly |
| <hr/> | |
| 2. shatter | [] To surprise someone
[] To fall suddenly
[] To break something into pieces
[] To look outside |
| <hr/> | |
| 3. grin | [] To cry loudly
[] To request information
[] To hide behind something
[] To have a big smile |
| <hr/> | |
| 4. gape | [] To open the mouth wide in surprise
[] To get very angry
[] To laugh loudly
[] To shake from cold or fear |
-

ACKNOWLEDGEMENTS

I am grateful to Dr. Linda Spetter at Baiko Gakuin University for her generous editorial support from the early stages of this paper. I am also thankful to my colleague, Rick Lavin at Prefectural University of Kumamoto for his help finalizing the revision of the paper. I would like to extend my appreciation to the anonymous reviewers for their insightful comments and guidance.

ABOUT THE AUTHOR

Makoto Yoshii is Associate Professor in the Faculty of Letters at Prefectural University of Kumamoto, Japan. He graduated from the University of South Florida with a Ph.D. in SLA & Instructional Technology. His interests include vocabulary learning and CALL.

E-mail: yoshii@pu-kumamoto.ac.jp

REFERENCES

- Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, 5(1), 202-232. *Language Learning & Technology*, 5(1), 202-232. Retrieved July 10, 2005, from <http://llt.msu.edu/vol5num1/alseghayer/default.html>
- Aweiss, S. (1994). Situated learning in technology: The case of computer-mediated reading supports. *Journal of Educational Technology Systems*, 23(1), 63-74.

- Chen, H. (2002). Investigating the effects of L1 and L2 glosses on foreign language reading comprehension and vocabulary retention. Paper presented at the annual meeting of the Computer-Assisted Language Instruction Consortium, Davis, CA.
- Chen, H.-C., & Leung, Y.-S. (1989). Patterns of lexical processing in a nonnative language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 316-25.
- Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, *80*(2), 183-198.
- Davis, J. F., & Lyman-Hager, M. (1997). Computers and L2 reading: Student performance, student attitudes. *Foreign Language Annals*, *30*(1), 58-72.
- Gettys, S., Imhof, L. A., & Kautz, J. O. (2001). Computer-assisted reading: The effect of glossing format on comprehension and vocabulary retention. *Foreign Language Annals*, *34*(2), 91-106.
- Grace, C. (1998). Retention of word meanings inferred from context and sentence-level translations: Implications for the design of beginning-level CALL software. *The Modern Language Journal*, *82*(4), 533-544.
- Grace, C. (2000). Gender differences: Vocabulary retention and access to translations for beginning language learners in CALL. *Modern Language Journal*, *84*(2), 214-24.
- Hong, W. (1997). Multimedia computer-assisted reading in business Chinese. *Foreign Language Annals*, *30*(3), 335-344.
- Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 113-125). London: Macmillan
- Hulstijn, J. H. (1993). When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *The Modern Language Journal*, *77*(2), 139-147.
- Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, *80*(3), 327-339.
- Jacobs, G., Dufon, P., & Hong, F. (1994). L1 and L2 vocabulary glosses in L2 reading passages: Their effects for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading*, *17*(1), 19-28.
- Knight, S. (1994). Dictionary: The tool of last resort in foreign language reading? A new perspective. *The Modern Language Journal*, *78*, 285-299.
- Kost, C. R., Foss, P., & Lenzini, J. J. (1999). Textual and pictorial glosses: Effectiveness of incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals*, *32*(1), 89-113.
- Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical Aspects of Memory: Current Research and Issues*, Vol. 2. (pp. 389-395). London: John Wiley & Sons.
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149-74.
- Kroll, J. F., & Sunderman, G. (2003). Cognitive processes in second language learners and bilinguals: The development of lexical and conceptual representations. In C. J. Doughty, & M. H. Long (Eds.) *The Handbook of SLA* (pp. 104-29). Malden, MA: Blackwell Publishing.

- Laufer, B. (1990). Why are some words more difficult than others? Some intralexical factors that affect the learning of words. *International Review of Applied Linguistics in Language Teaching*, 28(4), 293-307.
- Laufer, B., & Shumueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal*, 28(1), 89-108.
- Lomicka, L. (1998). "To gloss or not to gloss": An investigation of reading comprehension online. *Language Learning & Technology*, 1(2), 41-50. Retrieved May 1, 2005, from <http://llt.msu.edu/vol1num2/article2/default.html>
- Ludwig, J. (1984). Vocabulary acquisition as a function of word characteristics. *The Canadian Modern Language Review*, 40(5), 552-562.
- Lyman-Hager, M., & Davis, J. F. (1996). The case for computer-mediated reading: Une vide de boy. *The French Review*, 69(5), 775-790.
- Miyasako, N. (2002). Does text-glossing have any effects on incidental vocabulary learning through reading for Japanese senior high school students? *Language Education & Technology*, 39: 1-20.
- Mondria, J., & Wit-de-Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics*, 12(3), 249-267.
- Nagata, N. (1999). The effectiveness of computer-assisted interactive glosses. *Foreign Language Annals*, 32(4), 469-479.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press.
- Oxford, R., & Crookall, D. (1990). Vocabulary learning: A critical analysis of techniques. *TESL Canada Journal*, 7(2), 9-30.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston.
- Paivio, A. (1990). *Mental representations: A dual coding approach*. New York: Oxford University Press.
- Paribakht, T., & Wesche, M. (1996). Enhancing vocabulary acquisition through reading: A hierarchy of text-related exercise types. *The Canadian Modern Language Review*, 52(2), 155-178.
- Paribakht, T., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady, & T. Huckin, (Eds.) *Second Language Vocabulary Acquisition* (pp. 174-200). Cambridge: Cambridge University Press.
- Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology*, 90(1), 25-36.
- Potter, M. C., So, K., Eckardt, V., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23(1), 23-38.
- Sadoski, M., & Paivio, A. (2001). *Imagery and text: A dual coding theory of reading and writing*. Mahwah, NJ: Lawrence Erlbaum Associate.
- Taylor, A. (2006). The effects of CALL versus traditional L1 glosses on L2 reading comprehension. *CALICO Journal*, 23(2), 309-318.
- Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, 19(3), 287-307.
- Yeh, Y., & Wang, C. (2003). Effects of multimedia vocabulary annotations and learning styles on vocabulary learning. *CALICO Journal* 21(1), 131-144.

Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of picture and annotation types. *CALICO Journal*, 20(1), 33-58.