THE EFFECTS OF NON-CALL GLOSSES AND CALL GLOSSES ON INCIDENTAL VOCABULARY LEARNING OF L2 JAPANESE
non-CALL 語注と CALL 語注による L2 日本語学習者の付随的語彙学習

Hungyi Lee, Ochanomizu University
李虹儀, お茶の水女子大学

1. INTRODUCTION

During the past decades, a great deal of research has advocated that vocabulary is a key aspect in second language acquisition (SLA), especially when it comes to incidental vocabulary learning (Restrepo Ramos, 2015). Incidental vocabulary learning means that learners comprehend the meaning of a text as the main purpose and learn vocabulary as a by-product (Huckin & Coady, 1999). Research has shown that second language (L2) learners may incidentally acquire vocabulary knowledge through reading (Webb, 2008).

However, incidental vocabulary learning has some problems, such as the wrong inference of unknown words from the context or neglect of noticing the presence of unfamiliar words. One technique used to solve these problems and promote incidental vocabulary learning is glossing (Hulstijn, Hollander & Greidanus, 1996), that is, providing explanations of the meanings of words; typically glosses are located in the margin of texts (Lomicka, 1998). Recently, Computer Assisted Language Learning (CALL) has been applied to the glosses. In this study, the glosses with CALL are considered "CALL glosses" and those without CALL are considered "non-CALL glosses". CALL glosses are vocabulary annotations in the hypertext format that contain information about a word; CALL glosses show up on the same screen as the text only when the hypertext is clicked (Lenders, 2008). Also, CALL glosses only present one word at a time. Non-CALL glosses are similar to traditional glosses, where all of the glossed words appear simultaneously along with the reading text.

To the best of my knowledge, only three experimental studies have compared the effects of CALL glosses and non-CALL glosses on incidental vocabulary learning (Lyman-Hager, Davis, Burnett & Chennault, 1993; Bowles, 2004; Jalali & Neiriz, 2012), but the studies failed to reach consensus. Moreover, those results are obtained based on western languages. No previous research has been conducted on Japanese language learning with CALL glosses. Japanese features a unique orthography, consisting of Kanji, Hiragana, and Katakana. It is distinct from western languages. Furthermore, studies about the effects and impact of technology on language learning are necessary, but there are few that focus on Japanese learning (Honda, Iwata, Yoshinaga & Watanabe, 2014). To fill the gap, this study aimed to compare the effects of CALL glosses and non-CALL glosses on incidental vocabulary learning of L2 Japanese learners.

2. LITERATURE REVIEW

The three studies mentioned in the previous section are the only researches comparing the effects of CALL to non-CALL in L2 vocabulary learning. Also, Bowles (2004) pointed out that the learners' proficiency might have the impact on the usage of

CALL and should be in future studies. Hence, the previous research on the effect of glosses versus learner's proficiency was included.

2.1 CALL GLOSSES VERSUS NON-CALL GLOSSES

Lyman-Hager et al. (1993) investigated the effects of CALL glosses as compared to non-CALL glosses on L2 reading comprehension and vocabulary retention. The participants were 262 college students from intermediate-level French courses. All participants were randomly assigned to either a CALL or a non-CALL condition and were asked to read a text (660 words) with 20 glosses. Three weeks later, the participants took a vocabulary quiz on the 20 glossed words. The results revealed that the mean score of the CALL group was higher than the non-CALL group, and determined that CALL glosses enabled participants to retain vocabulary better than non-CALL glosses.

Bowles (2004) empirically studied the effects of CALL glosses and non-CALL glosses on L2 vocabulary acquisition. 50 freshmen majoring in Spanish were randomly assigned to a CALL group, a non-CALL group, or a control group without any glossing. They were asked to read a text (568 words) with 21 glossed words. After reading, all participants were asked to complete an immediate and a delayed vocabulary posttest, which included recognition and production tests. The results indicated that the CALL glosses and the non-CALL glosses groups were significantly better than the control group, but there was no significant difference between the CALL glosses group and the non-CALL glosses group.

The comparison of CALL and non-CALL glosses was also made by Jalali and Neiriz (2011). There were 180 English learners from Iran as participants, who were randomly assigned to different groups: some with CALL glosses and some with non-CALL glosses. They were asked to read two texts, and each text had 350 words and 20 glossed words. After reading, they were asked to complete an immediate and a delayed vocabulary posttest, which included recognition and production tests. According to the experiment results, Jalali and Neiriz concluded that the usage of CALL or non-CALL glosses made no difference in vocabulary acquisition.

2.2 GLOSSING AND LEARNERS' PROFICIENCY

Cheng and Good (2009) examined the interaction between glossing and learner's proficiency. The participants were 135 Taiwanese undergraduate students of business or engineering backgrounds. They were grouped into four English proficiency levels: four for the highest proficiency while one for the lowest. All participants were asked to complete one vocabulary pretest, one immediate vocabulary posttest, and two delayed vocabulary posttests. The study revealed that the language proficiency was indeed a factor affecting the gloss effects: the benefits of glossing were only found in participants of level two and three. That is, the highest and lowest participants did not appear to benefit from glossing.

Yusuf, Sim, and Awab (2014) investigated whether the use of CALL glosses aided L2 vocabulary learning of participants with different proficiency levels. The participants were 99 English learners from Malaysia. Based on their English proficiency, the participants were grouped into three levels: high, intermediate, and low. They were asked to complete an immediate and a delayed vocabulary posttest, which included

recognition and production tests. The results revealed that high proficiency learners made the most gains followed by mid and low proficiency learners.

2.3   SUMMARY OF LITERATURE

According to the previous research, there was no clear consensus on the effects of CALL glosses and non-CALL glosses on incidental vocabulary learning. In addition, there are some limitations to the findings of those studies which threaten their validity. Lyman-Hager et al. (1993) did not conduct immediate vocabulary test; therefore, the immediate effects of glossing were unknown. Moreover, in the studies by Lyman-Hager et al. (1993), Bowles (2004), Jalali and Neiriz (2011), when they grouped the leaners into different conditions, they did not consider learners' computer literacy. Finally, Cheng and Good (2009), Yusuf, Sim, and Awab (2014) revealed that the language proficiency might influence the gloss effects; however, previous studies did not take learners' proficiency into account while comparing CALL glosses to non-CALL glosses.

3.   RESEARCH QUESTIONS

This study investigates the effects of non-CALL glosses and CALL glosses on incidental vocabulary learning with learners of different language proficiency. In order to achieve the goal of this study, the following research questions will be addressed:

1. In L2 reading comprehension, is there any significant difference between CALL glosses and non-CALL glosses with different learners' proficiency in the immediate vocabulary posttest?

2. Comparing the immediate and delayed vocabulary posttests, is there any significant difference between CALL and non-CALL glosses?

4.   METHOD
4.1   PARTICIPANTS

The participants for this study consisted of 83 native Chinese speakers, who are third year Japanese majors. Based on their language proficiency, the participants were put into two groups: high and low. Their language proficiency was assessed by an internationally recognized standardized language test, SPOT[1] (Simple Performance-Oriented Test), developed by Tsukuba University. The participants were categorized into two gloss conditions: CALL and non-CALL; this was done by a computer literacy survey, which was about their experience reading on the computer and doing online reading activities. Among all 83 participants, 17 were selected (8 from High Proficiency Group and 9 from Low Proficiency Group) for a pilot study; the remaining 66 were the participants in the present study and categorized into four groups according to their SPOT

---

[1] The SPOT is similar to a cloze test in format, but is unique in that it employs audio stimuli that contain full sentences, including answers, spoken at natural speed. The learner's task is to fill in each blank in a sentence with one hiragana character, which is one component of the Japanese writing system. Thus, the SPOT involves not only grammatical knowledge, reading and listening ability, but also the immediate or automatic processing of aural and visual stimuli, as test takers must keep pace with the audio stimuli (Eda, Itomitsu & Noda, 2008).

score and computer literacy survey: (1) High, CALL (n=18) (2) Low, CALL (n=17) (3) High, non-CALL (n=16) (4) Low, non-CALL (n=15)

4.2   READING MATERALS AND GLOSSED WORDS

A week before the present study, a pilot study was conducted to determine the reading materials and glossed words. The 17 participants were asked to read three texts[2] and underline unknown words in the texts. Then, they were asked to do the reading comprehension test. After finishing the test, they would complete a vocabulary checklist and a pilot study questionnaire. The vocabulary checklist contained words that showed up in the texts and were potentially unknown to the participant. If a word was unknown to a participant, he would have to mark this word in the vocabulary checklist. This was used to avoid the situation where the participants overlooked unknown words while reading texts. If a word was underlined in the text or marked in the vocabulary checklist by more than half of the participants, it would be glossed in the present study. The pilot study questionnaire asked the participants about their perceived difficulty of reading texts and reading comprehensive test, how interesting the texts were, and their Japanese learning background. Among three reading texts in the pilot study, the text for the present study was selected for the following reasons: the reading comprehension test result of the selected text showed no significant difference between High and Low groups. Also, the marked or underlined words were the least diverse among the three texts. Furthermore, none of the participants considered the selected text too easy or too difficult.

According to the pilot study, the text "電車内、まなざしの行方 (How People Gaze on the Train)" with 955 words was selected, and 13 words were targeted for glossing in the two different gloss conditions. In both gloss conditions, the meaning of the word in the L1 (first language, i.e., Chinese) was placed on the right side margin of the reading text. Also, it was counted how many times a participant used glossing while reading texts. For CALL, computers would record it automatically; for non-CALL, participants were asked to use tally marks beside the glosses.

4.3   VOCABULARY TESTS

The vocabulary tests were undertaken in two phases: immediate posttest and delayed posttest. For both phases, two tests were taken. One was a form recognition test (form test) and the other was a meaning comprehension test (meaning test); both tests had 20 multiple-choice questions. Out of those 20 questions, there were 13 questions that included the targeted words while the other 7 were unrelated to the reading materials as distractors. Each multiple-choice question had four options: one correct answer, two distractors, and one "I don't know". In the form test (Fig 1), participants were asked to complete a Japanese fill-in-the-blank quiz; on the other hand, in the meaning test (Fig 2), participants were given a word in Japanese and asked to choose its equivalent in Chinese. The form and meaning tests were used to measure participants' receptive vocabulary knowledge.

---

[2] These three texts were selected from "Advanced Readings in Japanese".

In the immediate posttest phase, the test was followed by a word list and a present study questionnaire; the former was used to check the participants' prior vocabulary knowledge while the latter was mainly about their perception of glossing usage and their personal language learning background. Furthermore, the present study questionnaire for CALL groups asked some extra questions about the participants' computer literacy and their preference between computer-based and paper-based reading. After the delayed posttest, the participants were asked to check only a word list. This list aimed to ask their exposure to those targeted glossed words between two posttest phases.

( ) 電車には目を閉じて、＿＿＿＿をする人がいる。
①タスキねいり　②タヌキねいり　③タヌギねいり　④分からない

Figure 1　The sample of form test.

( ) タヌキねいり：
①像狐狸一樣睡覺　②呼呼大睡　③假裝睡覺　④不知道

Figure 2　The sample of meaning test.

4.4　PROCEDURE

Regarding the procedure of this study, SPOT and the computer literacy survey were undertaken as a prerequisite. Based on the results, the participants were put into four groups as mentioned before. Then, a reading text and target words were selected based on the pilot study. One week after the pilot study, the present study was conducted. The participants of the present study were instructed to read a text; at this time, they were unaware of any vocabulary test. The reason was that knowing the presence of those tests in advance might direct their attention to vocabulary instead of reading comprehension, which would deviate from the nature of incidental vocabulary learning. After reading the text, they were asked to complete an immediate posttest (form test and meaning test), a present study questionnaire, and a word list without time constraint. In order to test the long-term vocabulary retention and avoid the short-term memory bias, the delayed posttest along with a word list were conducted after two weeks. Because the content and structure of the delayed posttest were the same as the immediate one, the presence of the targeted words in the delayed posttest was rearranged in a scrambled order to avoid potential relational memory bias. Another thing worth noting is that the meaning test will accidentally result in additional exposure of the targeted glossed words, which might introduce the bias during the form test; hence, the participants would take the form test first to eliminate this potential bias.

5.　RESULTS

Table 1 shows the results of descriptive statistics from the immediate and delayed form tests; this includes the sample number (*N*), the mean (*M*), and the standard deviation (*SD*) of the test score from all four groups, and Table 2 shows those for the meaning test.

In order to provide a better visualization of the results, the mean scores from all test scenarios are illustrated in Fig 3.

The results implied that "High・CALL" group outperformed the other three groups in all four tests and is followed by "High・non-CALL" group, "Low・CALL" group, and "Low・non-CALL" group. With respect to the effect of CALL, Fig 3 illustrates that the scores of CALL group were higher than those of non-CALL group in all test scenarios. Also, all groups showed a score decrease in the delayed posttests compared to the immediate posttests.

Table 1　The results of descriptive statistics for immediate and delayed form test scores

Proficiency	Glosses	Immediate form test			Delayed form test			Mean difference[3]
		N	M	SD	N	M	SD	
High	CALL	18	10.44	2.33	18	9.00	1.88	1.44
	non-CALL	16	8.63	2.09	16	6.38	2.36	2.25
Low	CALL	17	7.65	2.40	17	6.06	1.64	1.59
	non-CALL	15	6.80	2.27	15	4.93	2.31	1.87

Note. perfect score is 13.

Table 2　The results of descriptive statistics for immediate and delayed meaning test scores

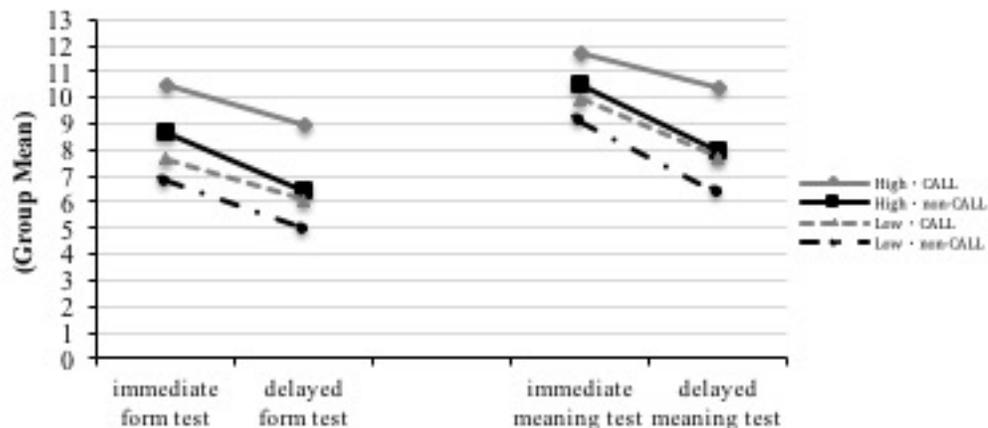| Proficiency	Glosses	Immediate meaning test			Delayed meaning test			Mean difference
		N	M	SD	N	M	SD	
High	CALL	18	11.67	1.65	18	10.39	1.77	1.28
	non-CALL	16	10.44	1.83	16	7.94	2.24	2.50
Low	CALL	17	9.94	1.95	17	7.71	1.97	2.23
	non-CALL	15	9.07	2.34	15	6.33	2.92	2.74

Note. perfect score is 13.

Figure 3　Group means from all tests over time.

---

[3] "Mean difference" is the difference between immediate posttest and delayed posttest.

The three-way Multivariate Analysis of Variance (MANOVA) was applied. It revealed that there were significant effects for gloss condition, $F(1,124)=9.66$, $p<.01$, learner's proficiency, $F(1,124)=17.79$, $p<.01$, and the timing of the tests $F(1,124)=17.77$, $p<.01$. However, no statistically significant difference was found for any interaction effect. It indicated that CALL groups significantly outperformed non-CALL groups, high proficiency learners performed significantly better than low proficiency ones, and the score of the immediate tests was significantly higher than that of the delayed ones. However, MANOVA cannot provide meaningful comparison when combing two or more significant main effects, such as the comparison between "High・non-CALL" group and "Low・CALL" group. Hence, the one-way Analysis of Variance (ANOVA) and Fisher's Least Significant Difference (LSD) test are necessary for this situation where two or more significant main effects are taken into consideration.

ANOVA was used to detect whether there was a significant difference between the four groups (High・CALL, High・non-CALL, Low・CALL, and Low・non-CALL). The ANOVA result confirmed that there was a significant effect for group in all test scenarios (Table 3), and the results of post hoc LSD are shown in Table 4.

Table 3　The results of one-way ANOVA

		Sum of Squares	df	Mean Square	F	p
Immediate meaning test	Between Groups	58.86	3	19.62	5.20	.00**
	Within Groups	233.81	62	3.77		
	Total	292.67	65			
Immediate form test	Between Groups	123.89	3	41.30	7.94	.00**
	Within Groups	322.48	62	5.20		
	Total	446.36	65			
Delayed meaning test	Between Groups	143.74	3	47.91	9.68	.00**
	Within Groups	310.08	62	5.00		
	Total	453.82	65			
Delayed form test	Between Groups	150.69	3	50.23	11.90	.00*
	Within Groups	261.63	62	4.22		
	Total	412.32	65			

Note. * = $p<.05$, **=$p<.01$

Table 4　Significant contrasts detected by LSD post hoc comparisons of posttest measures

	Form test			Meaning test		
Immediate posttest	High・CALL	>	High・non-CALL Low・CALL Low・non-CALL	High・CALL	>	Low・CALL Low・non-CALL
	High・non-CALL	>	Low・non-CALL			
Delayed posttest	High・CALL	>	High・non-CALL Low・CALL Low・non-CALL	High・CALL	>	High・non-CALL Low・CALL Low・non-CALL

Basically, in all the test scenarios, LSD implied that "High・CALL" group performed significantly better than all the other groups except for the immediate meaning posttest, where there was no significant difference between "High・CALL" and "High・

non-CALL" groups. Another exception was that the score of "High・non-CALL" was significantly higher than that of "Low・non-CALL" only on the immediate form posttest but not on the other three tests. The above findings were the key to the investigation of the effect of CALL and non-CALL in vocabulary learning (short term) and vocabulary retention (long term), which will be further addressed in the Discussion section.

SUMMARY OF RESULTS

RQ1 asked: In L2 reading comprehension, is there any significant difference between CALL glosses and non-CALL glosses with different learners' proficiency in the immediate vocabulary posttest?

According to Table 5, among learners with high proficiency, the CALL group scored significantly higher than the non-CALL group in the form test but not in the meaning test. On the other hand, there was no significant difference between "Low・CALL" and "Low・non-CALL" for both tests. Hence, at least for the form test, learners with high proficiency performed significantly better with the help of CALL glossing.

RQ2 asked: Comparing the immediate and delayed vocabulary posttests, is there any significant difference between CALL and non-CALL glosses?

In the form test, the significant difference between "High・non-CALL" and "Low・non-CALL" only existed in the immediate posttest but not in the delayed posttest. On the other hand, the meaning test result showed that there was no significant difference between "High・CALL" and "High・non-CALL" in the immediate posttest, which was not the case in the delayed posttest.

6. DISCUSSION

By further analyzing the results, intriguing findings fell into the following three domains: the effect of glossing conditions, the effect of CALL and non-CALL glosses in vocabulary retention, and the effect of CALL and non-CALL glosses with different learner's proficiency. Each domain will be further addressed below.

With respect to the glossing condition, it was found that in the immediate meaning test, there was no significant difference between "High・CALL" and "High・non-CALL" groups; similarly, no such difference was observed when comparing "Low・CALL" and "Low・non-CALL" group. This implied that the glossing condition makes no difference for the short-term meaning learning for participants with both High and Low proficiency levels, which agreed partially with the existing studies (Bowles 2004; Jalali & Neriz 2011) where they concluded that a glossing condition of CALL or non-CALL was irrelevant in the immediate meaning test. However, in the delayed meaning test, we noticed that there was a difference between "High・CALL" and "High・non-CALL" groups; this was not observed in the experiments conducted by Bowles and Jalali et al.

To address the effect of CALL and non-CALL glosses in vocabulary retention, we would have to compare the results of immediate posttests and delayed posttests. In the form test, "High・non-CALL" group outperformed "Low・non-CALL" group only in the immediate posttests, which implies that non-CALL glosses only provide relatively short-term help in vocabulary learning to learners with high proficiency. Also, in the meaning tests, a significant difference between "High・CALL" and "High・non-CALL"

group could only be observed in the delayed posttest; this indicates that for high proficiency learners, CALL glosses are more beneficial to vocabulary retention than non-CALL. By combining the above two findings, it could be said that for leaners with higher proficiency, CALL gloss is clearly more helpful than non-CALL gloss in vocabulary retention. The hypothesis is that CALL glossing shows only one word at once where learners can pay more attention to the form and meaning of the word they clicked on, which helps them to memorize the word better.

Also, according to the score of the tests, "High・CALL" group significantly outperformed "High・non-CALL" group in most of the tests except for the immediate meaning test; on the other hand, "Low・CALL" and "Low・non-CALL" groups showed no difference in all four tests. This fact indirectly indicates that learners with high proficiency could leverage CALL glosses much better than those with low proficiency. This fact was revealed in Yusuf, Sim and Awab's (2014) study where they found that high proficiency learners made the most gains from CALL glosses. This can be further interpreted to mean that the effect of glossing usage largely depends on the proficiency of the learners. Hence, it is important to consider learners' proficiency levels when designing gloss conditions since the selection of glossing depends on the proficiency level of the learners.

7. CONCLUSION

This study aimed to compare the effects of CALL glosses and non-CALL glosses on L2 incidental vocabulary learning and vocabulary retention. One of the main results of this study was that CALL glosses enable learners to retain vocabulary longer than non-CALL glosses, which differs from the studies of Bowles (2004) and Jalali and Neiriz (2011), where they revealed that CALL glossing had no effect on the performance of the learners in vocabulary learning. There are some potential reasons to justify the findings in this study. One is that learners with CALL condition can control the behavior of glossing, including appearance and duration of a certain word; this will eliminate the distraction of other unclicked words. Another explanation is that the CALL glossing only shows one word at a time, which helps learners concentrate on each word, leading to greater vocabulary retention. Moreover, participants' computer literacy was not considered in the previous studies. This might result in the situation where a participant was assigned to a CALL group while he was not comfortable using computers.

Notwithstanding the potential benefits of CALL glosses, we found that high proficiency learners benefit more from CALL glosses than low proficiency learners. Also, their computer literacy would surely influence the effect of CALL glossing. Hence, the selection of CALL or non-CALL should depend on the learner's language proficiency, and their computer literacy.

In the future, there are still some issues that should be addressed in future studies. In this study, the focus was on the receptive vocabulary knowledge instead of productive knowledge; hence, the effect of glossing on both can be further studied. It is also interesting to correlate learner's language proficiency to the vocabulary glossing usage. The present study questionnaire showed that some high proficiency learners inferred the

word meaning before using glosses and they used glosses mainly for confirmation. Therefore, the relation between the effect of CALL glossing and a learning strategy of different language proficiency levels needs to be studied in the future. Furthermore, one of the possible explanations supporting CALL is that showing one glossed word at a time can help learners focus on each word more and retain it better. However, there was no evidence to prove this hypothesis. Hence, eye-tracking technology, which will trace the movement of gaze during reading, can be introduced into future studies in order to investigate this issue.

REFERENCES

Bowles, M. A. (2004). L2 glossing: To CALL or not to CALL, *Hispania*, *87*(3), 541-552.

Cheng, Y. H., & Good, R. L. (2009). L1 glosses: Effects on EFL learners' reading comprehension and vocabulary retention. *Reading in a Foreign Language*, *21*(2), 119-142.

Eda, S., Itomitsu, M., & Noda, M. (2008). The Japanese skills test as an on-demand placement test: Validity comparisons and reliability. *Foreign Language Annals*, *41*(2), 218-236.

Honda, H., Iwata, M., Yoshinaga T., & Watanabe, T. (2014). *Gait of Japanese education: Studies guide for beginners(translation)*. Osaka: Osaka University.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language, *Studies in Second Language Acquisition*, *21*(2), 181-193.

Hulstijn, J. H., 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.

Jalali, S., & Neiriz, A. (2012). Computer-based versus traditional L1 and L2 glosses, *International Journal of Philosophy and Social Sciences*, *2*(9), 188-217.

Kakikura, Y., Suzuki, S., Mikami, K., & Yamagata, M. (2008). *Advanced readings in Japanese*. Tokyo: ALC.

Lenders, O. (2008). Electronic glossing: Is it worth the effort?, *Computer Assisted Language Learning*, *21*(5), 457-481.

Lomicka, L. L. (1998). "To gloss or not to gloss": An investigation of reading comprehension online, *Language Learning & Technology*, *1*(2), 41-50.

Lyman-Hager, M., Davis, J.N., Burnett, J., & Chennault, R. (1993). Une vie de boy: Interactive reading in French. In F.L. Borchardt & E. M. T. Johnson (Eds.), *Proceeding of the CALICO 1993 Annual Symposium on "Assessment"*, Durham, NC: Duke University. 93-97.

Restrepo Ramos, F. D. (2015). Incidental vocabulary learning in second language acquisition: A literature review. *PROFILE Issues in Teachers' Professional Development*, *17*(1), 157-166.

Yusuf, M. A., Sim, T. S., & Awab, S. (2014). Students' proficiency and textual computer gloss use in facilitating vocabulary knowledge, *English Language Teaching*, *7*(11), 99-107.

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*(1), 232-245.