Language Learning & Technology
ISSN 1094-3501

June 2017, Volume 21, Issue 2
pp. 32–51

ARTICLE

The effects of concordance-based electronic glosses on L2 vocabulary learning

Hansol Lee, University of California, Irvine and Korea Military Academy

Mark Warschauer, University of California, Irvine

Jang Ho Lee, Chung-Ang University

Abstract

The present study investigates the effects of two different vocabulary learning conditions in digital reading environments equipped with electronic textual glossing. The first condition presents the concordance lines of a target lexical item, thereby making learners infer its meaning by reading the referenced sentences. The second condition additionally offers the definition of a target lexical item after learners consult the concordance lines, thus enabling learners to confirm their meaning inference. A total of 138 English as a Foreign Language students completed a meaning-recall vocabulary pre-test, and three different reading tasks, which were followed by meaning-recall vocabulary post-tests in a repeated measures design with a control condition. Overall, the findings showed that the second condition resulted in higher vocabulary gains than both the first condition and the control condition. Yet, a closer look at the interactions of (a) the participants' clicking behaviors, (b) the difficulty of selected concordance lines, (c) the surrounding contexts around target lexical items, and (d) the participants' prior knowledge of the target lexical items showed that each target lexical item may require different treatments for it to be recalled most efficiently and effectively. Through this investigation, the present study suggests that glossary information, such as concordance lines, may involve more complex and unexpected learner interactions.

Keywords: *Corpus, Literacy, Multimodal Texts, Reading, Vocabulary*

Language(s) Learned in this Study: *English*

APA Citation: Lee, H., Warschauer, M., & Lee, J. H. (2017). The effects of concordance-based electronic glosses on L2 vocabulary learning. *Language Learning & Technology*, *21*(2), 32–51. Retrieved from http://llt.msu.edu/issues/june2017/leewarschauerlee.pdf

Introduction

Unprecedented technological change is transforming classroom environments, often leading students to read electronic texts on computer screens (i.e., digital reading) instead of paper-based textbooks. Digital reading may offer some potential advantages. For example, vocabulary learning through reading could benefit from multimedia environments that provide textual (e.g., synonyms, definitions), audio (e.g., pronunciation, sound effects), or visual supports (e.g., pictures, videos; see Anderson-Inman & Horney, 2007; Nation, 2009; Yanguas, 2009). Among these digital scaffolding tools, the focus of this study is on electronic glosses (e-glosses) for textual supports. Given that digital reading environments are more versatile and dynamic than their paper-based counterparts, the potential of e-glosses has been a subject of scholarly interest for second language (L2) vocabulary and reading research (Abraham, 2008; Chun, 2011).

Traditional glosses, which provide supplementary information for vocabulary in reading texts, have been highlighted as an effective tool for vocabulary learning, particularly in learning meanings of unfamiliar words when reading a lexically challenging text (Nation, 2009; Schmitt, 2008). On the other hand, e-glosses may have different *formats* on the computer screen (e.g., AbuSeileek, 2011; Chen & Yen, 2013; Lee & Lee, 2015) or may be filled with different *types* of glossary information (e.g., Lee & Lee, 2015; Poole, 2012).

This means that digital reading can include various types of glossary information for its target vocabulary, regardless of length. In this study, we endeavored to adopt a new type of glossary information: concordance lines. As shown in Figure 1, these lines are collections of example sentences extracted from a corpus, which is a set of large and structured language data. For instance, Line 7 could be a good example of concordance lines for the target lexical item (TLI) *in the vicinity of*: "…travels Walton had noticed that large retail chains tended to locate their stores *in the vicinity of* large population centres, in the belief that small-town outlets could not be profitable…"

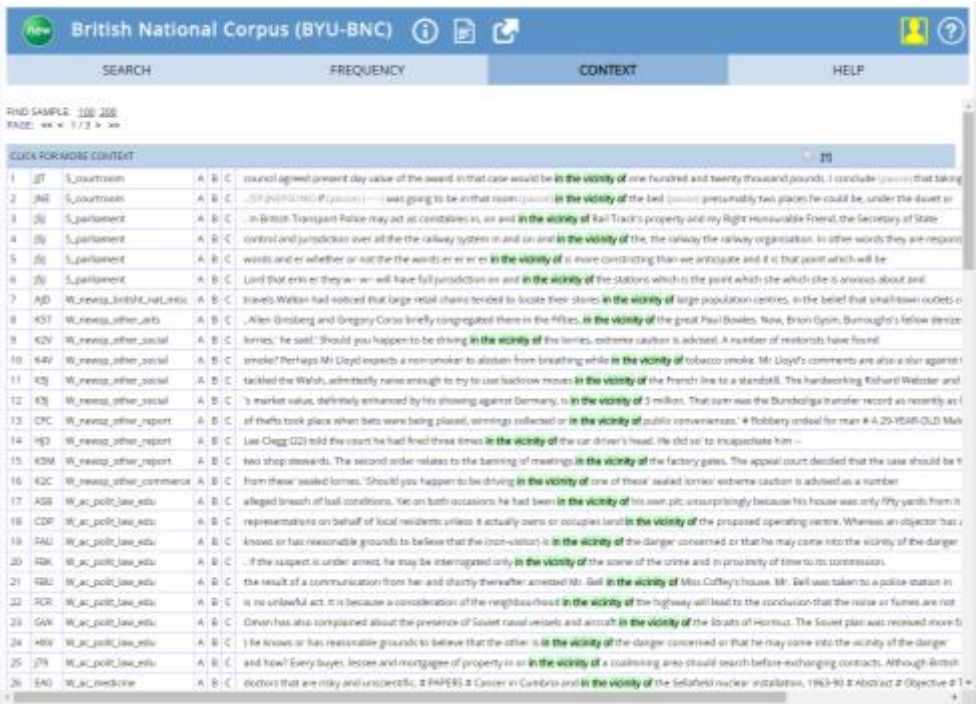

Figure 1. A snapshot of concordance lines from the British National Corpus (Davies, 2004).

It is surprising that there have been only a limited number of empirical studies on this issue to date (e.g., Lee & Lee, 2015; Poole, 2012), considering the strong theoretical justification for the idea of using this type of information for L2 vocabulary learning, such as input enhancement (i.e., when key words are salient in each sentence; see Chapelle, 2003), the noticing hypothesis (i.e., learners will notice a TLI while exposed to its occurrences in multiple contexts; see Schmidt, 2001), and the involvement load hypothesis (i.e., readers will be involved in meaning inferences; see Laufer & Hulstijn, 2001).

The issue we attempt to address by examining the value of concordance lines as effective glossary information for L2 learners' acquisition of word meaning is the gap between the theoretical supports and empirical evidence that should be bridged for finding a more effective pedagogical approach to L2 vocabulary learning. To this end, we tested the effects of two different types of e-glosses, with the first type providing the concordance lines of a TLI only and the second type providing the concordance lines plus the definition of a TLI, under a repeated measures design (i.e., within-subject). We also analyzed log data related to the participants' interactions with e-glosses in order to gauge the extent to which they consulted the glossed items and comprehended concordance glossary information. Along with results from the experimental phase, interview data with a subset of the participants and the record of their implementation of e-glosses aided in understanding of the learners' complex interactions with e-glosses.

Background

The conceptual foundation of our study flows from three interrelated topics: (a) the role of e-glosses in

digital reading environments, (b) the benefits and limitations of using concordance lines for vocabulary learning, and (c) prior research on learners' interactions with e-glosses and its implication for using concordance lines as a type of glossary information in e-glosses.

Digital Reading and E-Glosses

Anderson-Inman and Horney (2007) suggest that digital reading would offer promising opportunities for readers in terms of accessibility and supportiveness by providing various types of digital scaffolding tools alongside text. In their view, e-glosses can serve as effective supports for transforming a plain electronic text into a "supported eText" (p. 153). In a similar sense, e-glosses have been supported within several theoretical and pedagogical backgrounds (e.g., Chun, 2011; Lee & Lee, 2015). Above all, digital reading increases the likelihood of TLIs being noticed by readers because these items, which are hyperlinked to e-glosses, can be made visually salient in a variety of styles (Chapelle, 2003; Chun, 2001). Therefore, e-glosses have the potential to contribute to a reader's learning of unfamiliar vocabulary when reading electronic texts (i.e., noticing hypothesis; see Schmidt, 2001). Furthermore, by giving readers more control over their reading processes (Leu, Kinzer, Coiro, Castek, & Henry, 2013), digital reading constructs an "interaction" among readers, texts, and scaffolding materials (Chapelle, 2003, p. 25). Lastly, unlike the print form, digital reading is not limited by spatial restrictions (Lee & Lee, 2015). Thus, digital platforms may have e-glosses filled with an abundance of lexical information such as multiple concordance lines for TLIs (Nation, 2009).

In light of these virtues of digital environments, a number of empirical studies have demonstrated the positive effect of e-glosses on L2 learners' vocabulary learning (for a meta-review, see Abraham, 2008). Furthermore, the interest of the research community has recently shifted toward the format of glossing (e.g., AbuSeileek, 2011; Chen & Yen, 2013) as well as the type of glossary information (e.g., Lee & Lee, 2015; Poole, 2012). Regarding the former, a small number of empirical studies (e.g., AbuSeileek, 2011; Chen & Yen, 2013) have been conducted on comparing the effects of different formats of e-glosses (e.g., pop-up type, marginal type), but the findings of these studies have not been consistent. Research on different types of glossary information for e-glosses is even scarcer, particularly regarding the use of concordance lines as glossary information. In the following section, we will first review the literature dedicated to the use of corpora for vocabulary learning, and then introduce two studies that have used concordance data in e-gloss format.

Concordance Lines as Vocabulary Learning Resources

The use of corpora in L2 vocabulary learning has attracted the interest of the research community for a number of reasons. First, inferring the meaning of an unfamiliar word is considered an effective strategy for learning vocabulary (e.g., Fraser, 1999; Schmitt, 1997). Learners, in theory, are supposed to make a more informed and accurate guess of the meaning of an unfamiliar word when exposed to multiple contextual instances surrounding a target word (Johns, 1986). Second, allowing learners to infer meaning from examples is thought to generate a high level of learner involvement, which may lead to greater retention (Laufer & Hulstijn, 2001). Third, providing multiple instances of TLIs in a wide range of sentential contexts is believed to enhance learners' awareness of TLIs, thus accelerating their vocabulary acquisition (Chapelle, 2003; Schmidt, 2001).

Although they did not utilize a corpus analysis, early empirical studies attempted to confirm the effect of using example sentences excerpted from corpus-based dictionaries on L2 vocabulary learning (e.g., Laufer, 1993; Summers, 1988). For example, Summers (1988) examined the effects of example sentences for the participants' vocabulary comprehension and production. She designed three different conditions, with types of information selected from dictionaries: definitions, example sentences, and both definitions and example sentences. Although the experimental conditions led participants to have better results than the control condition, there was no significant difference across the two treatment conditions with respect to the participants' vocabulary comprehension and production. With a similar research objective, Laufer (1993) tested the use and comprehension of 18 TLIs with 43 English as a Foreign Language (EFL) undergraduate

students, providing four different conditions (i.e., definitions, examples, definitions followed by examples, and examples followed by definitions). The results indicated that the combinations of definitions and examples were more effective than definitions only or examples only for the participants' vocabulary use and comprehension. Moreover, Laufer found that definitions might contribute more to improving comprehension than examples, whereas the contributions of these two components were similar for the production counterpart.

Recent studies showed more positive results, probably because of more diverse experimental conditions thanks to the state-of-the-art corpus technology (e.g., Chan & Liou, 2005; Cobb, 1999; Frankenberg-Garcia, 2012, 2014). For example, Cobb (1999) conducted an empirical study with two different vocabulary learning conditions: concordance-based vocabulary learning (e.g., the use of a concordance program) and traditional vocabulary learning (e.g., the use of dictionaries and word lists). The results of Cobb's study with 20 adult Chinese EFL students showed that the former treatment yielded more gains in terms of the learners' knowledge of vocabulary. In Chan and Liou (2005), 32 Taiwan college students completed web-based practice units incorporated with a bilingual concordancer, and the results showed a significant collocation improvement with the use of a corpus example, as well as an on-line concordance program during their vocabulary practice. Recently, studies by Frankenberg-Garcia (2012, 2014) confirmed the positive effects of concordance lines on L2 vocabulary learning. Taking into consideration that examples should include enough contextual clues for comprehension, she carefully selected concordance lines from multiple corpora. The results indicated that these examples were effective for EFL students in terms of both productive (e.g., correcting typical L2 mistakes, 2012; and writing sentences using TLIs, 2014) and receptive knowledge (e.g., understanding TLIs, 2012; 2014).

However, when it comes to L2 vocabulary learning in the e-gloss format, empirical findings have been inconclusive (e.g., Lee & Lee, 2015; Poole, 2012). For example, Poole (2012) compared the effects of syntactically modified concordance lines and dictionary definitions of glossed words as two different types of glossary information, and could not find any statistical difference between these two types in improving the participants' vocabulary acquisition. Similarly, Lee and Lee's (2015) study re-examined the effects of these two different types of glossary information. Unlike Poole's (2012) study, the authors did not modify concordance lines to the level of the participants. The results showed that participants who received dictionary definitions made higher vocabulary gains than those who had concordance lines as their glossary information.

These inconclusive results concerning the value of concordance lines may be explained by previous suggestions that learners inferencing meaning from context may do so ineffectively (e.g., Schmitt, 2008), and thus retain wrongly inferred meanings in their lexicons (Mondria, 2003). Similarly, it should also be noted that the learners might not be able to understand all the given concordance lines to successfully elicit the meaning of a TLI. In short, the use of concordance lines, which inevitably involves the inference of meaning, has not received the attention it deserves.

One way of overcoming the limitation of concordance lines is to enable learners to confirm the meaning they inferred from contexts (e.g., Cobb, Greaves, & Horst, 2001; Fraser, 1999). As Godwin-Jones (2001) suggested, if the learners' meaning inferences can later be confirmed, their inaccurate inferences will be minimized. However, to date, there has been no empirical study to support this suggestion—in particular, none in digital reading environments.

Learners' Interactions with E-Glosses

Along with the learners' inaccurate meaning inferences, another practical issue to consider in using concordance lines as glossary information is the learners' interactions with e-glosses, including implementation rate of clicking e-glosses. By tracking user behavior, previous studies have focused on conditions that made learners click e-glosses (e.g., Chun, 2001; Laufer, 2000). One of the major findings from these studies was that learners largely preferred a type of lexical information that required a relatively low level of cognitive load. For example, Laufer (2000) found that participants did not make use of example

sentences of a target word as a type of glossary information in digital reading environments; rather, they opted for word definitions. It is likely that concordance lines as glossary information require a relatively high level of cognitive load for learners to process. It can be expected that this type of glossary information may not be overwhelmingly favored by learners. Hence, understanding the learners' interactions with e-glosses seems to be an important issue in examining the effects of e-glosses for vocabulary learning.

However, there have been few empirical efforts to assess implementation rates and observe the specific behaviors of learners' consultations of concordance lines in e-glosses. For example, Poole (2012) did not include any research methods to figure out how and to what extent his participants interacted with glossary information for their understanding and learning. This is a problem in other studies on e-glosses, which are limited to focusing only on the results of vocabulary tests, based on the assumption that treatments have been ideally employed without properly understanding learners' behaviors. Along the same line, it is noteworthy that the aforementioned theoretical supports for using concordance lines as glossary information are based on the premise that learners would be likely to devote their full attention to that kind of lexical information.

In light of this gap in the literature, we not only examine the effects of providing a confirmation process along with concordance lines for the meaning inferences, but also observe the learners' implementation rate of consulting glossary information as well as their clicking behaviors with e-glosses. We hypothesize that this will be a significant step toward overcoming the limitations of concordance lines as glossary information, and propose future pedagogical directions for L2 vocabulary learning.

Research Questions

The following research questions guided our study:

1. What effects do the two different vocabulary learning conditions in digital reading environments have on the meaning-recall of TLIs for EFL adult learners?
   A. How significant are the effects of receiving different treatments on the meaning-recall of TLIs?
   B. How significant are the effects of providing the treatments in different orders?
2. How do the participants interact with e-glosses when reading, and what are the pedagogical implications of these interactions?
   A. What are the average implementation rates of and amount of time spent on consulting e-glosses during reading?
   B. What are their relations to the results of a meaning-recall post-test?
   C. How are the participants' clicking behaviors different across TLIs?

Methods

To answer the research questions, we carried out an experiment that was based on a repeated measures design. This enabled us to deliver different reading conditions and measure their effects on the participants' meaning-recall of TLIs in a controlled way with a reliable level of confidence. This section discusses the different aspects of the research method of the present study: the description of the participants, experimental design, target reading materials, and an outline of the procedure and data analysis.

Participants

A total of 138 undergraduate South Korean EFL learners participated in the study. All of them were 21 years old and had 10 years of English learning experience in formal school contexts. The average score of the participants on the Test of English for International Communication was 732, indicating that they were independent English users (B1–B2) based on the Common European Framework of Reference for Languages, according to the testing publishers (Educational Testing Service, 2012). The present study was conducted during their enrollment in a mandatory English course. The participants were from nine intact

classes, taught by three different instructors with the same textbook and curriculum at the time of the study.

Experimental Conditions and Design

This study first developed three different versions of the digital reading materials, one with e-glosses for concordance lines (CONC) and another with e-glosses for concordance data supplemented with dictionary definitions (CODI), and the other without any e-glosses, which served as a control (CTRL). A pilot study was conducted to determine the appropriate number of concordance lines for glossary information; in general, students in the study (n = 45) pointed to three examples as the most manageable number for inferring the meaning of target items, without being distracted (M = 2.91, SD = 0.73). In this way, the first author of this study and the aforementioned three instructors carefully selected three concordance lines for each TLI from multiple corpora (e.g., Open American National Corpus, British National Corpus, Brown Corpus) for the participants' effective meaning inferences in light of their proficiency levels (for similar efforts, see Frankenberg-Garcia, 2012, 2014). Appendix A further explicates how we chose the most appropriate three examples for the word *inflection* in order to highlight its specific meaning used within its context (i.e., a change in the pitch or tone of a person's voice).

For the dictionary definitions of TLIs, this study opted to use the Merriam-Webster Online Dictionary, which had been widely used in the participants' institution. The rationale behind choosing L2 dictionary definitions for glossary information instead of first language (L1) was twofold: the participants were intermediate-level learners with a vocabulary of more than 2,000 words (i.e., learners with a vocabulary of less than 2,000 often have comprehension problems with L2 glosses; see Nation, 2009) and L2 dictionary definitions were considered to be more appropriate language input in this experimental condition, where L2 concordance lines were provided as glossary information.

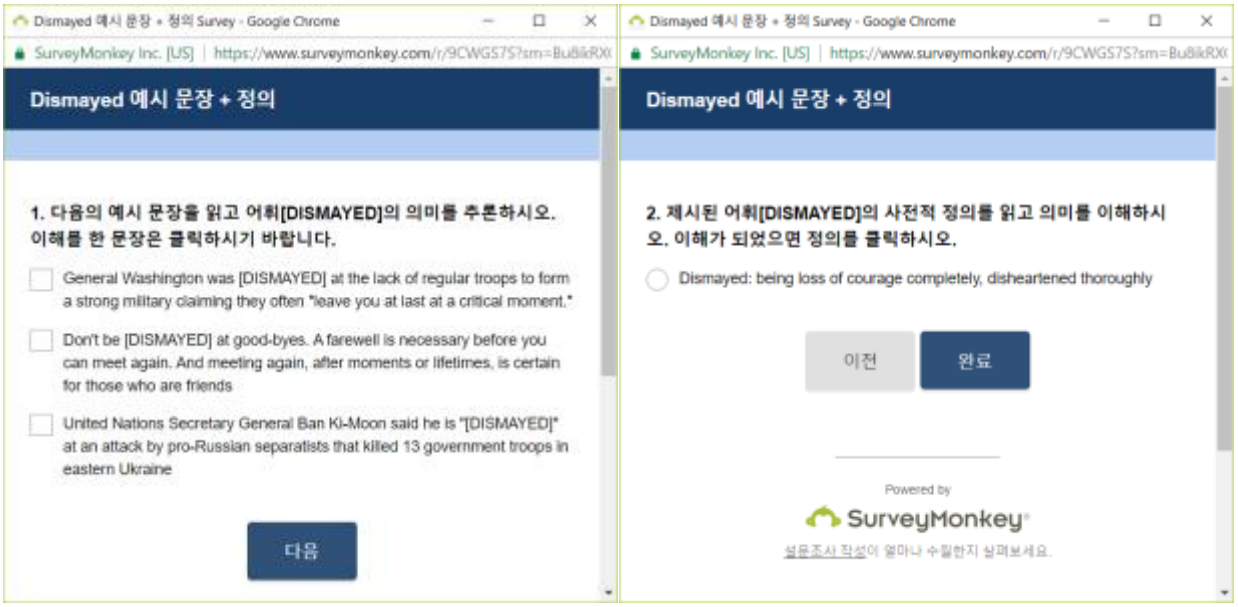

Figure 2. Glossary information: Concordance lines (left) and dictionary definitions (right).

To further obtain data on the participants' behaviors with e-glosses and glossary information, this study used a free online survey tool (i.e., SurveyMonkey) to provide glossary information in a popup window format. Specifically, every glossed lexical item was hyperlinked to a pop-up style window, which presented concordance lines of the item upon the participants' activation. For the CODI condition, an additional window was designed to provide the definition of a target item, which was activated when a user clicked the "Next" button after reading concordance lines in the previous window (see Figure 2). In this manner, data regarding the participants' clicks on glossed items were collected to analyze the implementation rate of and length of time they spent with each e-gloss until closing the window. Furthermore, a checkbox was

included in front of each concordance sentence so that the participants could report which of the concordance lines they understood. This e-gloss format was designed in such a way that this study could measure how and to what extent the participants interacted with the glossed words and whether they understood or consulted glossary information during their reading.

In order to expose the participants to all the conditions, a repeated measures design was adopted. This design allows each participant to experience all the conditions, including the control condition. In repeated measures experiments, it is important to confirm that any order effects (i.e., the effects that the order of presenting three different conditions might have on the results) do not exist. As part of this effort, counter-balancing was taken into consideration when designing the group formation, as shown in Table 1. For example, there were six possible orders to consider all the possible combinations of the three conditions: (a) CTRL, CODI, CONC; (b) CODI, CONC, CTRL; (c) CONC, CTRL, CODI; (d) CTRL, CONC, CODI; (e) CONC, CODI, CTRL; and (f) CODI, CTRL, CONC. Since there were nine intact classrooms in total, there was a random assignment of six different orders to the six classrooms, and the remaining three classrooms were randomly assigned to one of those six orders. The results of the data analyses confirmed that there were no significant order effects from the experimental design (see non-significant coefficients of the order effect variable in Table 3).

Table 1. *Study Design*

Classroom	Order	Trial 1	Trial 2	Trial 3
1, 2	Order 1 (n = 29)	CTRL	CODI	CONC
3, 4	Order 2 (n = 29)	CODI	CONC	CTRL
5, 6	Order 3 (n = 28)	CONC	CTRL	CODI
7	Order 4 (n = 15)	CTRL	CONC	CODI
8	Order 5 (n = 15)	CONC	CODI	CTRL
9	Order 6 (n = 16)	CODI	CTRL	CONC

Material

Three reading texts were extracted from Cutting Edge Advanced (Cunningham, Moor, & Carr, 2003). The length of each reading was 459, 479, and 519 words, respectively. We chose 30 potential TLIs from these texts and selected ten TLIs per text based on the results of a pilot test with 45 students similar in profile to the participants in the study (for a list of TLIs, see Appendix B; for hyperlinks to one of these texts, see Appendix C).

Testing and Scoring

A total of four meaning-recall tests of vocabulary were conducted, at the beginning of the study as well as after each reading activity. In these tests, participants were asked to write down the meaning of a target word either in English or their L1 (i.e., Korean). When scoring, a total of two points were allotted for each item. One point was given when students gave a partially correct meaning of each TLI, while two points were given for a completely correct meaning. The first author and one of the three instructors graded the vocabulary test. Both raters scored fifteen percent of the testing sheets for the purpose of checking inter-rater reliability; and the reliability was found to be 0.93 (Cohen's Kappa, $p < 0.05$). Any discrepancies were resolved through discussion with the third author.

Procedure

At the beginning of the study, the participants were given a pre-test with all 30 TLIs from the three target texts; as previously mentioned, the TLIs were selected from a pilot study. Then, a computer workshop was given by the first author after the pre-test to give the participants some basic knowledge concerning the definition of concordance lines and how they could infer the meaning of a glossed word by consulting the

given lines.

The main reading tasks were conducted two weeks after the pre-test, with the presence of the first author of the present study in the classrooms for the sake of treatment fidelity. As part of the effort to minimize the potential impact of instructors, all the materials were designed in a way in which each individual could complete all the activities without any further guidance or instruction. Each task was performed weekly to prevent any possible carry-on effects (i.e., effects that carry over from one condition to another). In each reading session, the participants were asked to read the text with their own laptops for 15 mins, and this reading was followed by an immediate post-test for 5 mins on a different web page.

Interviews

After the experiment, in order to understand how the participants interact with e-glosses across the TLIs, interviews were carried out with a purposely stratified sample of three participants: one student of advanced proficiency (C1), one at the upper intermediate level (B2), and one intermediate user (B1). In the interview, these participants were presented with the three different texts they read, and were asked to give their opinions about each TLI, and how much additional glossary information was needed in comprehending its meaning. The interviews were conducted in their L1, audio-recorded, and partially transcribed and translated.

Data Analysis

Statistical analyses were performed using STATA (Version 14.0). Prior to the regression analyses, correlations between predictor variables (i.e., independent variables in the regression equations) were examined in order to control for multicollinearity. The results showed that predictor variables were not strongly related ($r < 0.8$). Then, residualized change regression analyses with the Huber-White standard errors (i.e., controlling for heteroscedasticity) including the cluster adjustment (i.e., ensembling multiple test results at the student level) were conducted for the first research question, followed by additional analyses for the robustness checks (i.e., fixed-effect adjustments, simple change regression analyses; for the variables and equations for these regression models and details, see Appendix D).

For the second research question, the number of clicks of all the glossed words and the length of time spent consulting glossary information were analyzed, along with the participants' reports on the number of concordance lines they had comprehended. In particular, the amount of time the participants spent on consulting glossary information was analyzed by excluding potential outliers (e.g., those who did not spend enough time on making meaning inferences or those who spent too much time on each TLI, for example, if they left the pop-up window open). The interquartile range (IQR) rule was applied in this case. For example, the first quartile (Q1) and the third quartile (Q3) were calculated, based on the time the participants spent on each of the TLIs. Then the IQR was calculated (IQR = Q3 - Q1), and the lower boundary (Q1 - 1.5 × IQR) and the upper boundary (Q3 + 1.5 × IQR) were computed. If the time one spent on consulting glossary information was outside this range, then this click was considered an outlier. Combining all of the above, the implementation rates of the participants' clicking the TLIs and consulting glossary information were analyzed. A paired t-test was further conducted to compare the mean difference between implementation rates for the two experimental conditions (i.e., CODI and CONC). Moreover, a multiple regression analysis with Huber-White standard errors was conducted in order to confirm possible associations between the participants' clicking behaviors and the meaning-recall rate of the TLIs.

Limitations

There are two limitations of the present study. First, delayed tests were not conducted because of the participants' limited availability. Within a brief time period, we decided to provide them with all the conditions without employing delayed tests, rather than to randomly assign them into one of the three conditions (i.e., CTRL, CONC, and CODI) with delayed tests. So this study was not able to assess retention of vocabulary. Second, while the inclusion of an experimental condition with definitions alone would have allowed us to measure the effects of dictionary definitions in the CODI condition more accurately,

scheduling considerations (i.e., participant availability) made the use of a control group more feasible than a definition-only group. This allowed us to examine the effects of concordance lines as glossary information, as well as the effects of the confirmation of meaning inferences through dictionary definitions. These effects have been examined only to a limited extent in the previous literature (unlike definition-only, which has received considerable attention). While the experimental design of the present study was suitable for our goals, future research with a definition-only condition will be valuable for illuminating the pedagogical implications of exposure to different types of glossary information.

Results

This section presents the findings in two parts, with the first part reporting the results of vocabulary tests based on a set of multiple regression models, and the second part presenting the results concerning the participants' clicking behaviors in different experimental conditions along with the interactions of (a) individual lexical items, (b) their recall rates, and (c) their clicking behaviors.

Results of Vocabulary Recall Tests

Table 2 provides descriptive statistics for the scores of the three conditions on the vocabulary recall tests. Out of 138 participants, a total of six could not complete all three reading tasks, and thus were excluded in the analyses. Overall, the participants demonstrated significant gains in learning vocabulary for all the conditions, according to the results of paired *t*-tests ($p < 0.001$).

Table 2. *Means and Standard Deviations for the Vocabulary Tests and Results of the Paired t-tests*

	Pre-test		Post-test			
Conditions	***M***	***SD***	***M***	***SD***	***t*-test**	***t* value**
CTRL (n = 132)	0.27	0.64	3.97	3.17	Pre < Post	13.45***
CONC (n = 132)	0.20	0.44	6.24	4.17	Pre < Post	17.00***
CODI (n = 132)	0.14	0.43	8.89	4.60	Pre < Post	22.19***

Note. *** $p < 0.001$

Regarding the first research question, the results from the residualized change model, as shown in Model 1 in Table 3, revealed that there was a significant treatment effect depending on the different conditions (b = 2.51, $p < 0.001$) when controlling for three learner variables (i.e., pre-test scores, English proficiency, and gender).

For the next step, dummy variables for the three different conditions were plugged into the regression model to compare the participants' post-test scores under these conditions (see Model 2 in Table 3). The estimated coefficients for the dummy variables implied that the CTRL condition would, on average, lead a participant to get a 2.32 lower vocabulary score than the CONC condition (b = -2.32, $p < 0.001$), and that one would, on average, get a 2.69 higher vocabulary score if the participant were assigned to the CODI condition rather than to the CONC condition (b = 2.69, $p < 0.001$). As a result, it can be interpreted that one would, on average, learn about one more TLI out of 10 total, or partially learn about two more TLIs if the participants were given CODI rather than CONC.

In addition, when the order effect product term (i.e., the interaction effect of providing different experimental conditions in different orders; conditions × trial) was added to the model, no significant effect was found (b = 0.25, $p > 0.05$), with the treatment effect depending on the different conditions remaining statistically significant, as shown in Model 3 in Table 3.

The additional analyses for the robustness checks also confirmed the aforementioned findings. The full results regarding these analyses, including fixed-effect adjustments (i.e., Model 4 in Table 3) and simple change models (i.e., Models 1, 2, and 3 in Table 6), are described in Appendix E.

Table 3. *Regression Models of the Vocabulary Tests (Residualized Models)*

Independent variables	Dependent Variable: Vocabulary Recall Post-test (*n* = 132 Participants × 3 Conditions = 396 Observations)			
	Model 1: Residualized Model	**Model 2: Dummy Variables**	**Model 3: Order Effect Added**	**Model 4: Classroom Fixed-Effects**
Conditions	2.51***			
	(0.21)			
CTRL vs. CONC		-2.32***	-1.84**	-1.97*
		(0.41)	(0.66)	(0.88)
CODI vs. CONC		2.69***	2.21***	2.34**
		(0.44)	(0.62)	(0.88)
Trial			0.86	1.00
			(0.50)	(0.79)
Order Effect			0.25	0.18
Trial × Condition			(0.28)	(0.38)
Pre-test	0.40*	0.40*	0.41*	0.41*
	(0.20)	(0.20)	(0.18)	(0.19)
English Proficiency	0.01***	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Gender	-0.32	-0.32	-0.39	-0.70
	(1.54)	(1.54)	(1.55)	(0.96)

*Notes. Standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

Results Concerning the Participants' Clicking Behavior

As for the second research question, we first examined whether the CODI and CONC conditions resulted in different implementation rates. As shown in Table 4, the participants, on average, showed an implementation rate of about 83%. Despite the different amounts of vocabulary gains between the CODI and CONC conditions (t = 3.41, $p < 0.001$), the implementation rates were not significantly different between the two conditions (t = 0.92, $p > 0.05$). Moreover, the participants spent similar amounts of time looking up glossary information under both the CODI and CONC conditions (t = 0.66, $p > 0.05$). Lastly, the number of concordance lines of which the participants reported comprehension was not substantially different between the conditions (t = 0.52, $p > 0.05$).

Further regression analysis was performed to explore a relationship between the participants' clicking behaviors and their recall rates of target vocabulary. The dependent variable was the meaning-recall scores of the TLIs, whereas the independent variables included (a) the average amount of time (in secs) spent on each piece of glossary information, (b) the average number of concordance lines each participant reported to have comprehended, (c) the rates of clicking for each e-gloss, and (d) the condition variable (dummies for CODI and CONC; see Table 5). The results indicated that the condition variable had a significant effect on predicting the recall score (b = 11.70, $p < 0.01$), whereas the number of concordance lines each participant clicked did not (b = -15.86, $p > 0.05$). On the other hand, the amount of time spent on glossaries had a negative effect on predicting the dependent variable, albeit a very weak one (b = -0.28, $p < 0.05$).

Table 4. *Differences in Recall Scores and Clicking Behaviors between CONC and CODI*

Conditions	**CONC (*n* = 30)**	**CODI (*n* = 30)**	***t*-test**	***t* value**
Vocabulary Recall Test Score	27.57 (12.33)	39.23 (14.10)	CONC < CODI	3.41***
Rates of Clicking Each E-gloss	0.82 (0.14)	0.85 (0.11)	CONC = CODI	0.92^{ns}
Average Time Spent	37.22 (21.82)	40.88 (20.99)	CONC = CODI	0.66^{ns}
Number of Concordance Lines Comprehended	1.52 (0.37)	1.57 (0.32)	CONC = CODI	0.52^{ns}

Notes. Standard deviations are in parentheses. ^{ns} *p > 0.05, *** p < 0.001*

Table 5. *Influences of Participants' Clicking Behaviors on Meaning-Recall Test*

Independent Variables	**Coefficients**	***SE***
Average Time Spent	-0.28*	0.11
Rates of Clicking Each E-gloss	60.69	34.74
Number of Concordance Lines Comprehended	-15.86	8.73
Condition CODI vs. CONC	11.70**	3.43

*Notes. * p < 0.05, ** p < 0.01*

The results of the interview revealed a complex picture of the participants' vocabulary learning, as described in Figure 3. In other words, the close analyses of each TLI, along with the participants' clicking behaviors and recall scores, pointed to complex interactions of (a) participants' clicking behaviors, (b) the nature of selected concordance lines, (c) the surrounding context of a TLI, and (d) the participants' prior knowledge of each TLI.

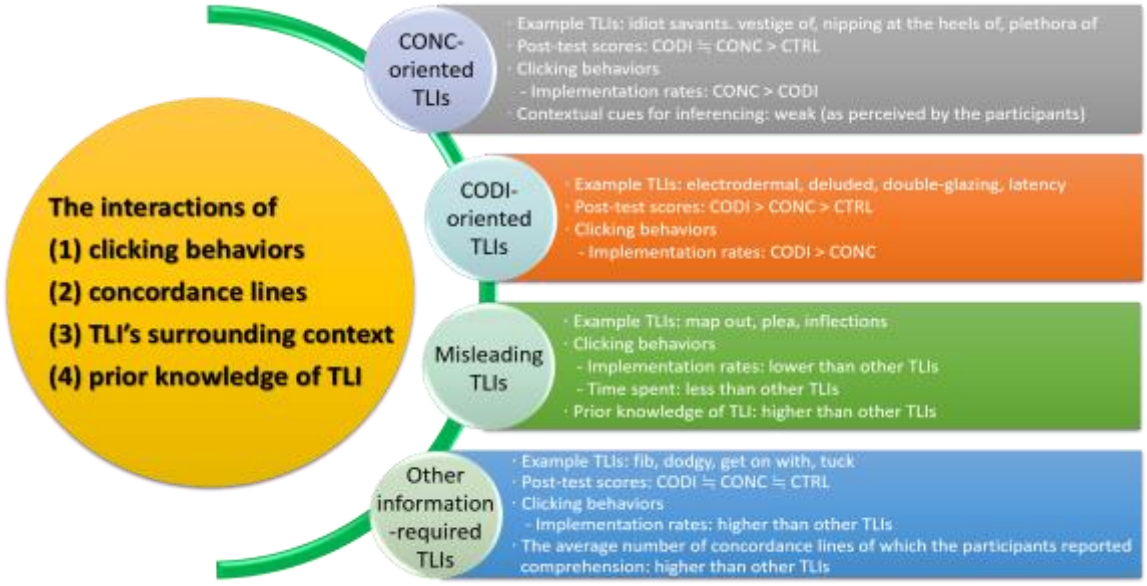

Figure 3. Four emerging patterns of TLIs from the participants' interactions with e-glosses.

The first group of the TLIs with similar patterns included *idiot savants*, *vestige of*, *nipping at the heels of*, and *plethora of*. These patterns were as follows: (a) the participants' meaning-recall scores in the CTRL condition were on average close to zero (indicating that the participants in this condition failed to infer the meaning of these TLIs), and there were only small differences in the meaning-recall score between CONC and CODI; (b) the participants' implementation rates were higher on average in CONC than CODI; and (c) the participants lacked contextual cues in inferring the meaning of a TLI, as can be seen from the response to the TLI *plethora of*:

> C1 (the interviewees' names are replaced by their proficiency levels: *C1*, *B2*, and *B1*): To be honest, I don't know about this item … I think I can guess its meaning from the previous paragraph … but not from sentences surrounding this item.
>
> B2: I think I can guess what it means … but I am highly uncertain.

With these TLIs, the participants were thus not able to infer their meanings properly, but concordance lines significantly contributed to their meaning inferences, whereas the confirmation of their meanings through dictionary information was not obligatory for most of the participants. Having seen the patterns of the first group, the words that fall into this category may be called *CONC-oriented TLIs*, meaning that concordance lines as glossary information are not only beneficial, but also sufficient for accurate meaning inferences.

The second set of TLIs that showed consistent patterns included *electrodermal*, *under duress*, *in the vicinity of*, *deluded*, *double-glazing*, and *latency*. For these TLIs, three observations were made. First, the participants' average meaning-recall score was highest for the CODI condition, lower for CONC, and lowest for CTRL. Second, the participants' implementation rates were higher in the CODI than in the CONC condition. Third, the participants were able to make some inferences about the meaning of a TLI based on the surrounding context and the part of the TLI (i.e., morpheme), as can be seen from the interviewees' comments on *electrodermal*:

> C1: When I see this word, *electrodermal*… it reminds me of the word *electronic*. Considering previous words, such as *blood pressure* and *breathing rate*… this word could be related to the physical signs of the human body.
>
> B2: I see this word consists of *electro-* and *dermal*… and I know both of these words. After reading the previous and next sentences … I was able to figure out that this word may indicate a sort of electronic sign from human skin.
>
> B1: I think this word is highly related to the term *electronic*. I don't see much of a problem for guessing the meaning of this word.

The interviewees also made a similar response to *latency*, indicating that they were able to make inferences based on the surrounding context. However, the comparison of the total meaning-recall scores between the CODI (total score = 36) and CONC conditions (total score = 24) suggests that the confirmation of the meaning of this word enhanced the participants' comprehension. Thus, these words may be called *CODI-oriented TLIs*, indicating that meaning inferences followed by the confirmation of correct meaning inferences would result in the most positive learning outcome for these TLIs.

The third group of the TLIs, which included *map out*, *plea*, and *inflections*, had the following patterns: (a) the participants' implementation rates were 10% lower for these items when compared with the rates of all other TLIs (which was about 80%); (b) the participants spent much less time on reading glossary information of these TLIs (on average, 23 secs, compared with the average of 45 secs for the other set of the TLIs); and (c) the participants were fairly confident in their inferred meanings of these TLIs, as can be seen from their responses to the target item, *map out*:

> C1: I already knew this expression… *to map* is to draw a map… to me, there is no need to have extra help for this easy phrase.
>
> B2: I do not see any necessity for accessing additional information for *map out*.

> B1: I am not sure about this phrase… but it seems straightforward... I think… it is to draw something. I don't think I need more information for this word.

It was found that the interviewees did not attempt to find contextual meanings of the expression *map out*, which is defined as *to plan* in the target context. It appeared that the three interviewees knew about the word *map*, but did not go further to explore the meaning of *map out*. The results of the post-test also support the comments of the interviewees. In particular, the majority of the participants' wrong answers were related to *a map* or to *drawing a map* (47 out of 97 wrong responses). In light of the rather low implementation rate for these TLIs, and the interviewees' confidence and misjudgment, we call those kind of terms, *misleading TLIs*. These items require particular attention from instructors, who will need to make sure their learners would not make wrong meaning inferences.

The final set of the TLIs included *fib*, *dodgy*, *get on with*, and *tuck* and had the following pattern: (a) the participants' average implementation rate was high (approximately 88%), (b) the average number of concordance lines for which each participant reported to have comprehended their meanings (M = 1.92) was higher than that for the rest of the TLIs combined (M = 1.72), and (c) the CONC and CODI conditions did not result in higher recall scores than did the CTRL condition. In other words, these TLIs were highly consulted, and their concordance lines were comprehensible to the participants. However, the CONC and CODI conditions were not necessarily more beneficial to the participants' meaning-recall than was the CTRL condition. So, it seems that these TLIs required other glossary information that was not provided in this study (e.g., L1 equivalents) for higher meaning-recall rates. Based on this insight, these words are called *other information-required TLIs*.

Discussion

The first research question of the present study investigated whether two different vocabulary learning conditions (i.e., CONC and CODI) would make any differences in undergraduate EFL learners' meaning-recall knowledge of target vocabulary. With regard to this research aim, the results showed that the participants fared better under the CONC condition than the CTRL condition. The finding here supports several theoretical hypotheses that would confirm the use of concordance lines for vocabulary learning, such as the noticing hypothesis (Schmidt, 2001) and the involvement load hypothesis (Laufer & Hulstijn, 2001). In light of previous concerns about using concordance lines as glossary information (e.g., Cobb et al., 2001; Fraser, 1999; Godwin-Jones, 2001), we cautiously suggest that a few steps undertaken in the present study may account for the aforementioned positive results. That is, through a carefully planned pilot study, we examined the most appropriate number of concordance lines for their meaning inferences (i.e., three), and had multiple discussions with the instructors of the target classes in selecting example sentences from concordance lines, which were deemed fine-tuned to the participants' level of English proficiency (for similar efforts, see Frankenberg-Garcia, 2012, 2014).

Moreover, the CODI condition was more beneficial to the participants' meaning-recall than the CONC condition, supporting prior findings that the additional confirmation of an inferred meaning contributes to students' making more accurate meaning inferences from concordance lines (e.g., Cobb et al., 2001; Fraser, 1999). This finding also accords with Laufer's (1993) study, in which the combination of definition and example sentences resulted in the highest comprehension gains.

The results related to the second research question showed that a holistic account of the participants' meaning-recall is complex, after a close analysis of the interactions concerning (a) the participants' clicking behaviors, (b) the difficulty of selected concordance lines, (c) the context surrounding the TLIs, and (d) the participants' prior knowledge of each TLI. In particular, we have shown that the participants interacted rather differently with each set of TLIs. That is, a majority of the TLIs (e.g., *electrodermal*, *latency*) were best recalled in CODI, in accordance with the results of the first research question. The superiority of CODI over the other two conditions discussed above may be attributable to the fact that most of our TLIs fall into this group. On the other hand, some TLIs (e.g., *idiot savants*, *vestige of*) were recalled fairly well even

without the aforementioned confirmation process. These items were concordance-oriented; if concordance lines were judiciously selected for them, then their recall could be guaranteed. Another group of the TLIs (e.g., *map out, inflections*) misled learners into thinking that their meanings were easy to infer or were already known to them. In such a case, learners may easily make a wrong inference. These are the lexical items that should be dealt with very carefully by an instructor, as a wrongly inferred meaning could be retained in the learners' vocabulary system (Mondria, 2003). Finally, there was a small number of words (e.g., *fib, dodgy*) that were not recalled well even with concordance lines and dictionary definitions. It can be assumed that these words may be better retained by learners if other lexical information is provided.

The aforementioned categorization of the TLIs may not be equally applicable in other contexts. It is highly likely that learners from different pedagogical contexts, even with the same level of English proficiency as those in the present study, may interact differently with the aforementioned TLIs. Our intention was to raise researchers' awareness of the possibility of the dynamic interactions of (a) the learners' prior knowledge of target vocabulary, (b) the comprehensibility of glossary information, and (c) their actual utilization of such glossary information. As an example, the participants' implementation rates in this study ranged from about 50% to 100%, depending on the TLI. Through the interviews, it was found that some participants may opt not to use the given glossary information by mistakenly thinking that they already know the meaning of a TLI. On the other hand, the implementation rate and the participants' self-reported understanding level of concordance lines did not always correlate with the recall rates, thereby implying that L2 vocabulary learning may be subject to the aforementioned dynamic interactions. We believe that the innovation of the present study lies in demonstrating that some glossary information, such as concordance lines, may involve more unexpected interactions with L2 learners when compared with traditional dictionary information.

Conclusion

The present study investigated the effects of, and clicking behaviors related to, two different vocabulary learning conditions in digital reading environments: one providing concordance lines only and the other providing concordance lines along with definitions as glossary information. Based on the findings of this study, we conclude that providing concordance lines along with the subsequent confirmation of the inferred meanings is more effective than providing concordance lines only, which in turn results in better meaning-recall than no glossary information. Furthermore, we have shown that a particular lexical item may need different treatments for it to be recalled most efficiently and effectively through the close analyses of the interactions of (a) the participants' clicking behaviors, (b) the difficulty of selected concordance lines, (c) the context surrounding the target vocabulary, and (d) the participants' prior knowledge of the target vocabulary. While our findings should not be interpreted as leading to a prescriptive method for teaching these TLIs, they nevertheless can provide important guidelines for future L2 vocabulary research and teaching. One promising direction for future research would be to compare the effects of CODI with an experimental condition of dictionary definitions alone on the meaning acquisition of different sets of L2 lexical items. Future research may also benefit from the use of vocabulary measures other than meaning-focused tests, which may further reveal the effectiveness of concordance lines as glossary information for improvements in more productive aspects of lexical competence.

Acknowledgements

The authors appreciate Greg J. Duncan, Glenn Levine, and Robin Scarcella from the University of California, Irvine for their helpful comments on an earlier version of this study. They would also like to thank the editors of the journal and the anonymous reviewers for their careful reading of the manuscript and their insightful and constructive suggestions.

References

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, *21*(3), 199–226.

AbuSeileek, A. F. (2011). Hypermedia annotation presentation: The effect of location and type on the EFL learners' achievement in reading comprehension and vocabulary acquisition. *Computers & Education*, *57*(1), 1281–1291.

Anderson-Inman, L., & Horney, M. A. (2007). Supported eText: Assistive technology through text transformations. *Reading Research Quarterly*, *42*(1), 153–160.

Chan, T. P., & Liou, H. C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb–noun collocations. *Computer Assisted Language Learning*, *18*(3), 231–251.

Chapelle, C. A. (2003). *English language learning and technology*. Amsterdam, Netherlands: John Benjamins.

Chen, I.-J., & Yen, J.-C. (2013). Hypertext annotation: Effects of presentation formats and learner proficiency on reading comprehension and vocabulary learning in foreign languages. *Computers & Education*, *63*, 416–423.

Chun, D. (2001). L2 reading on the web: Strategies for accessing information in hypermedia. *Computer Assisted Language Learning*, *14*(5), 367–403.

Chun, D. (2011). CALL technologies for L2 reading post Web 2.0. In L. Ducate & N. Arnold (Eds.), *Calling on CALL: Theory and research to new directions in foreign language teaching (2nd ed.)* (pp. 131–170). San Marcos, TX: CALICO.

Cobb, T. (1999). Applying constructivism: A test for the learner-as-scientist. *Educational Technology Research and Development*, *47*(3), 15–31.

Cobb, T., Greaves, C., & Horst, M. (2001). Can the rate of lexical acquisition from reading be increased? An experiment in reading French with a suite of on-line resources. In P. Raymond & C. Cornaire (Eds.), *Regards sur la didactique des langues seconds* (pp. 133–153). Montréal, Canada: Éditions logique.

Cunningham, S., Moor, P., & Carr, J. C. (2003). *Cutting edge advanced with phrase builder*. Harlow, UK: Pearson Education.

Davies, M. (2004). *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). Available online at http://corpus.byu.edu/bnc/

Educational Testing Service. (2012). *Mapping the TOEIC® Listening and Reading onto the CEFR*. Retrieved from https://www.etsglobal.org/

Frankenberg-Garcia, A. (2012). Learners' use of corpus examples. *International Journal of Lexicography*, *25*(3), 273–296.

Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL*, *26*(2), 128–146.

Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition*, *21*(2), 225–241.

Godwin-Jones, R. (2001). Tools and trends in corpora use for teaching and learning. *Language Learning & Technology*, *5*(3), 7–12. Retrieved from http://llt.msu.edu/vol5num3/pdf/emerging.pdf

Johns, T. (1986). Micro-concord: A language learner's research tool. *System*, *14*(2), 151–162.

Laufer, B. (1993). The effects of dictionary definitions and examples on the use comprehension of new L2 words. *Cahiers de Lexocologie*, *63*, 131–142.

Laufer, B. (2000). Electronic dictionaries and incidental vocabulary acquisition: Does technology make a difference? In U. Heid, S. Evert, E. Lehmann, & C. Rohrer (Eds.), *Proceedings of the Ninth EURALEX International Congress* (pp. 849–854). Stuttgart, Germany: Stuttgart University.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1–26.

Lee, H., & Lee, J. H. (2015). The effects of electronic glossing types on foreign language vocabulary learning: Different types of format and glossary information. *The Asia-Pacific Education Researcher*, *24*(4), 591–601.

Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2013). New literacies: A dual-level theory of the changing nature of literacy, instruction, and assessment. In D. E. Alvermann, N. J. Unrau, & R. B. Ruddell (Eds.), *Theoretical models and processes of reading (6th ed.)* (pp. 1150–1181). Newark, Delaware: International Reading Association.

Mondria, J. A. (2003). The effects of inferring, verifying, and memorizing on the retention of L2 word meanings. *Studies in Second Language Acquisition*, *25*(4), 473–499.

Nation, I. S. P. (2009). New roles for FL vocabulary? In L. Wei & V. Cook (Eds.), *Contemporary applied linguistics volume 1: Language teaching and learning* (pp. 99–116). London, UK: Continuum.

Poole, R. (2012). Concordance-based glosses for academic vocabulary acquisition. *CALICO Journal*, *29*(4), 679–693.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge, UK: Cambridge University Press.

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 199–228). Cambridge, UK: Cambridge University Press.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363.

Summers, D. (1988). The role of dictionaries in language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 111–125). New York, NY: Routledge.

Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning & Technology*, *13*(2), 48–67. Retrieved from http://llt.msu.edu/vol13num2/yanguas.pdf

Appendix A. The Process of Selecting Example Concordance Lines for TLIs

The following illustrates the process of selecting example concordance lines for one of the TLIs, *inflection*, which has the meaning of "a change in the pitch or tone of a person's voice" in the target context. Below are sample concordance lines selected from BNC, OANC, and Brown Corpus.

i. It was purely to bring his ear reverentially into line with the mouth of whoever was speaking. "Exactly," he murmured. "Exactly." And Dyson knew from the depth of humility and reverence in his "INFLECTIONS" that he was getting a larger fee than even Lord Boddy (from the BNC).

ii. When you deal with customers over the phone, you have a whole new set of etiquette rules. The minute you pick up the phone, body language disappears, and your "INFLECTIONS" and tone of voice, and the words you use become the entire story (from the OANC).

iii. Godunov, it is the consistency with which every person on the stage—including the chorus—comes alive in the music. Much of this lifelike quality results from Mussorgsky's care in basing his vocal

line on natural speech "INFLECTIONS" (from the Brown Corpus).

The first author and instructors made the following decision in terms of selecting concordance lines for the target word *inflection*:

1. The concordance line (i.) was excluded. Its surrounding context requires further information to be comprehended, and there are many unfamiliar words and phrases, such as "the depth of humility" and "reverence," along with the key word. Furthermore, there is no obvious clue for inferring the meaning of *inflection* from the context.
2. The concordance line (ii.) was selected because the surrounding words and structures are not only comprehensible to the participants, but also clearly indicate the meaning of *inflection* as a modulation of intonation in the voice.
3. The concordance line (iii.) was not selected, although there are some words that allow for the meaning inference, such as "music," "vocal line," or "speech." The reason is that the given clues are not strong enough to provide the aforementioned definition of *inflection*, but may induce faulty meaning inferences.

Appendix B. List of TLIs and Their Definitions

First Reading Text

TLI	Definition
endowments	an attribute of the mind or body; natural talents or qualities.
idiot savants	a mentally defective person with an exceptional skill or talent in a special field.
fib	a small or trivial lie; minor falsehood.
corroborated	being supported to be more certain; be confirmed.
are beset with	being harassed by something; being attacked on all sides.
misnomer	a misapplied or inappropriate name or designation.
a vestige of	visible evidence of something that is no longer present or in existence.
dismayed	being loss of courage completely, disheartened thoroughly
nipping at the heels of	trying to be almost as good as someone that you are competing with.
a plethora of	overabundance; excess; too many; a lot of.

Second Reading Text

TLI	Definition
dodgy	untruthfully tricky; uncertain or unreliable.
tuck	to put into a small, close, or concealing place.
lumbering	moving clumsily or heavily.
get on with	to proceed with; to begin or continue; to work with.
cracked	pass through (a barrier); break through.
deluded	deceived; misguided; the mind or judgment is mislead.
mucky	filthy, dirty, or slimy.
traipse	to walk or go aimlessly or idly.
grannies	a grandmother; an elderly woman.
double-glazing	two panes of glass in a window.

Third Reading Text

TLI	Definition
interrogators	one who asks questions of (someone, especially a suspect or prisoner) closely, aggressively or formally.
polygraphs	a machine designed to detect and record changes in physiological characteristics, such as a person's pulse breathing rates; used as a lied detector.
rationale	A strong reason to support for something.
under duress	under pressure; forcibly restraint or restricted.
electrodermal	related to electrical properties of the skin.
plea	serious and emotional request for something.
in the vicinity of	the area around or near a particular place.
inflections	a change in the pitch or tone of a person's voice.
latency	the state of being inactive or late.
map out	to plan or sketch.

Appendix C. Hyperlinks to the Texts

Text #1: CONC / CODI / CTRL

Text #2: CONC / CODI / CTRL

Text #3: CONC / CODI / CTRL

Appendix D. Variables and Equations for the Regression Models

The first residualized change model (Model 1 in Table 3) included variables for treatment conditions. The independent variables are the participants' pre-test results, prior English proficiency (participants' official TOEIC scores, as developed by Educational Testing Service), and gender (male = 0 and female = 1), whereas their post-test results are the dependent variable. In addition, to detect any difference in the treatment effects between the three conditions, three dummy variables were generated that identified different conditions (CONC is the reference group among the CTRL, CONC, and CODI conditions), in the second regression equation (i.e., Model 2 in Table 3). The third equation includes trials (the first = 1, the second = 2, and the third = 3, among three different trials), and order effect (the interaction effect of delivering different conditions in different orders or trials) as additional independent variables, in order to detect any order effects (i.e., Model 3 in Table 3). The equation of Model 3 is as follows:

(1) $(Post\text{-}test)_{ij} = a + b_1(CTRL)_{ij} + b_2(CODI)_{ij} + b_3(Trial)_{ij} + b_4(Order\ effect)_{ij}$

$$+ b_5(Pre\text{-}test)_{ij} + b_6(Eng\_proficiency)_{ij} + b_7(Gender)_{ij} + \varepsilon_{ij.}$$

Second, an additional regression analysis, including the classroom fixed-effects ($\delta_{classroom}$), was conducted to only capitalize on within-classroom differences after removing between-classroom differences that could bias the estimation of the treatment effects. This approach was part of an effort to eliminate any possible discrepancies between the participants' intact classrooms. The equation of Model 4 is as follows:

(2) $(Post\text{-}test)_{ij} = a + b_1(CTRL)_{ij} + b_2(CODI)_{ij} + b_3(Trial)_{ij} + b_4(Order\ effect)_{ij} +$

$$b_5(Pre\text{-}test)_{ij} + b_6(Eng\_proficiency)_{ij} + b_7(Gender)_{ij} + \delta_{classroom} + \varepsilon_{ij.}$$

Third, we employed simple change models (i.e., Models 1, 2, and 3 in Table 6) in order to check the

robustness of the results of the aforementioned residualized change models. These equations considered the participants' meaning-recall knowledge gains per each condition (calculated by subtracting pre-test scores from post-test scores) as the dependent variable, and included all the independent variables, except for the pre-test results variable. For example, the equation of the third simple change model is as follows:

(3) $(\Delta Score;\ Post\text{-}test - Pre\text{-}test)_{ij} = \Delta a + b_1(CTRL)_{ij} + b_2(CODI)_{ij} + b_3(Trial)_{ij}$

$$+ b_4(Order\ effect)_{ij} + b_5(Eng\_proficiency)_{ij} + b_6(Gender)_{ij} + \delta_{classroom} + \Delta\varepsilon_{ij.}$$

Appendix E. Results Regarding Model 1 in Table 3 and Models 1, 2, 3 in Table 6

To find a more accurate and robust estimation of the effects of different conditions on the participants' meaning-recall test scores, we conducted an additional regression analysis with classroom fixed-effects (see Model 4 in Table 3), in addition to the residualized change models (Models 1, 2, and 3 in Table 3). The results showed that coefficients did change but by a very small amount relative to their standard errors. In other words, it appeared that the fixed effects adjustments produced small changes in the coefficients.

As a part of the efforts to check the robustness of the findings, we additionally conducted simple change regression analyses. Although participants received, on average, nearly 0 for their pre-tests (see Table 2), it should be noted that everyone may have experienced different amounts of gains (i.e., post-test–pre-test) throughout the experiment. Since the residualized change models only focus on the within-group mean, which may imply a regression towards the mean between groups, simple change models make more sense in this case by focusing on the participants' individual gains across the experiment.

Table 6. *Regression Models of the Vocabulary Tests (Simple Change Models)*

	Dependent Variable: Vocabulary Gains (Post-test–Pre-test; *n* = 396)		
Independent variables	**Model 1: Simple Change Model**	**Model 2: Dummy Variables**	**Model 3: Classroom Fixed-Effects**
Conditions	2.02***		
	(0.54)		
CTRL vs. CONC		-1.83**	-1.97*
		(0.66)	(0.88)
CODI vs. CONC		2.20***	2.34**
		(0.62)	(0.87)
Trial	0.82	0.82	0.96
	(0.50)	(0.50)	(0.79)
Order Effect	0.25	0.25	0.18
Trial × Condition	(0.28)	(0.28)	(0.38)
English Proficiency	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)
Gender	-0.39	-0.39	-0.71
	(1.53)	(1.53)	(0.95)

*Notes. Standard errors are in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001*

Models 1, 2, and 3 in Table 6 included change scores (i.e., post-test–pre-test per each condition) as the dependent variable, instead of post-test scores, without having the pre-test variable as one of the

independent variables in its regression equation. The results revealed that different treatment conditions are still significant predictors of the participants' vocabulary gains (b = 2.02, p < 0.001; see Model 1 in Table 6). In particular, CTRL would, on average, lead a participant to gain a 1.83 lower vocabulary score than CONC (b = – 1.83, p < 0.01), and participants assigned to the CODI condition, on average, would gain a 2.20 higher score than those assigned to CONC (b = 2.20, p < 0.01). Furthermore, it was re-confirmed that there was no order effect (b = 0.25, p > 0.05), in accordance with receiving different conditions in different orders (see Model 2 in Table 6).

Finally, we conducted an additional regression analysis with classroom fixed-effects (see Model 3 in Table 6). The results showed a similar pattern to that of the fourth residualized change model (i.e., Model 4 in Table 3), in which most of the standard errors got larger and the coefficients on the key predictors increased as well—that is, the main effect of providing different conditions became even larger after removing variations across classrooms.

About the Authors

Hansol Lee is a PhD candidate in the School of Education at the University of California, Irvine, and an Assistant Professor in the Department of English at Korea Military Academy. He has a wide range of research interests across applied linguistics, such as computer-assisted language learning, corpus linguistics, and language assessment.

E-mail: hansol3@uci.edu or hansol6461@gmail.com

Mark Warschauer is a Professor of Education and Informatics at the University of California, Irvine, where he directs the Digital Learning Lab and the Teaching and Learning Research Center, and the Editor of *AERA Open* journal. His research focuses on digital media and learning.

E-mail: markw@uci.edu

Jang Ho Lee received his PhD in education from the University of Oxford. He is presently an Assistant Professor in the Department of English Education at Chung-Ang University. His areas of interest are teachers' code-switching in English classrooms, learners' attitudes towards teachers' language use, and the use of mobile technology for language learners. All correspondence regarding this publication should be addressed to him.

E-mail: jangholee@cau.ac.kr