

The Effectiveness of Using Marginal Annotations in Reading Comprehension for FL Learners

Name: Dr. Elham Sweilam Ahmad Desouky
Lecturer of Curriculum & Instruction (TEFL),
Faculty of Education, October 6 University
Email: elesweilam@gmail.com

Abstract

Current research randomly allocated 104 FL learners to investigate whether there is any positive effect of marginal annotations on their reading comprehension performance and some other issues related to their learning. Learners were divided into two groups: the experimental group, which used marginal annotations, and the control group, which did not use marginal annotations. To achieve the goal of the study, both a pre-test and immediate post-test were used in two sessions, while the other two sessions used the delayed post-test design. English unaccusativity and 10 pseudo-word items were the goal features; thus, the current study utilized meaning recognition test, a word form and grammaticality judgment tests to evaluate the target features learning. The findings revealed that marginal annotations have no significant impact on reading comprehension outcomes, but they facilitate learning the target features to a great extent.

Keywords: Marginal Annotations; Reading Comprehension; FL Learners'

المخلص العربي

فاعلية استخدام الشروح الهامشية في الفهم القرائي لمتعلمي اللغة الانجليزية كلغة اجنبية
إستخدم البحث الحالي مجموعه عشوائيه مكونه من 104 من متعلمي اللغة الانجليزية وذلك للتحقق من التأثير الإيجابي للشروح الهامشية علي أداء هؤلاء الطلاب فيما يخص الفهم القرائي وكذلك بعض الأمور الاخرى المتعلقة بتعلمهم. تم تقسيم المتعلمين إلى مجموعتين: المجموعة التجريبية ، التي استخدمت التعليقات التوضيحية الهامشية ، والمجموعة الضابطة ، التي لم تستخدم التعليقات التوضيحية الهامشية. لتحقيق هدف الدراسة ، تم استخدام كل من الاختبار القبلي والاختبار البعدي المباشر في دورتين ، بينما استخدمت الدورتان الأخرى تصميم الاختبار البعدي المتأخر. واعتمادا علي ان الافعال الانجليزية المتداخلة وكذلك 10 كلمات الزائفة يمثلان هدف التعلم فقد استخدمت الدراسة الحاليه اختبار التعرف على المعنى ، وشكل الكلمة واختبارات الحكم النحوية لتقييم التعلم المستهدف. كشفت النتائج أن التعليقات التوضيحية الهامشية ليس لها تأثير كبير على نتائج الفهم القرائي ، لكنها تسهل تعلم الميزات المستهدفة إلى حد كبير.

The Effectiveness of Using Marginal Annotations in Reading Comprehension for FL Learners'

Eskey (2005) claimed that reading is one of the language skills in which all FL learners seek to improve their abilities. Moreover, if the learners were able to read the foreign language, they can understand its input and acquire it. Krashen (1982) argued that these two parts of FL reading are engaging in a symbiotic relation and are interlinked, which means that if the learners gain more knowledge of FL, they will be more skillful in dealing with texts, which leads to better FL reading comprehension and increased efficiency. Therefore, raising comprehension capacity and increase efficiency should be taken into consideration when preparing and applying FL reading instruction. In fact, many researchers (Grabe, 2009; Koda, 2012) have neglected reinforcing FL development during reading, despite the importance of comprehension as the central construct in FL reading instruction.

Van Patten (2012) revealed that learners' tendency to treat input to get meaning had enhanced through concentrating on comprehension. This means that it may prevent the appearance of the types of syntactic treatment needed for the acquisition of FL. Studies by Leow (2009) and Jung (2016) provided some pedagogical techniques to support FL acquisition that centered on FL reading instruction and keeping comprehension as the initial focal point at the same time. These studies offered those approaches after being aware of the restrictions of comprehension-limited instruction. The current study found that the types of textual amendment approaches that were used in some studies are just some instantiations of techniques

such as marginal annotations in Jung's study (2016) and textual input reinforcement in Sharwood's research (1993). Some studies (Robinson et al., 2012; Godfroid, 2013) showed that these approaches have one joint basic supposition, which is to direct the learners' attention to the target language form- meaning constructions during reading to improve comprehension. Despite this, previous studies have not offered a clear vision concerning the role of marginal annotations in FL reading comprehension and FL learning, which lack stimulated the researcher to examine this topic.

Review of Pertinent Literature

The main target of marginal annotations is to limit linguistic ambiguity by giving a translation, definition, or synonym of a difficult or unknown linguistic item. This step will facilitate reading comprehension as well as help learners to understand the meaning of the text. Many researches utilized different approaches to marginal annotations and compared them to measure efficacy. Some of those studies were those of Ko (2012), who used FL marginal annotations versus first language, Al-Seghayer (2001), who employed multimedia marginal annotations versus textual, and Bowles (2004), who applied paper-based marginal annotations versus computerized. The previous studies were not only different and varied in dealing with the different techniques of marginal annotations, but also differed in the levels of clarity in giving marginal annotations. A study by Al-Seghayer (2001) combined translation with related video clips or pictures, while the merge between the definition and its use in a model sentence was used in a study by Hulstijn and Laufer (2001). Guidi's (2009) research was different, where the type of provided marginal annotations

was direct and simple by using a synonym for the ambiguous word or giving a straightforward definition.

Results of the previous studies were inconclusive in terms of the impacts of marginal annotations on FL reading comprehension and FL learning. In studying the effects of marginal annotations, some studies (Ko, 2012; Martinez-Fernandez, 2010) supported the positive effect of marginal annotations on reading comprehension. By contrast, some other studies (Bell & LeBlanc, 2000; Davis & Lyman-Hager, 1997) showed little interest in the role of marginal annotations on facilitating reading comprehension and learning. Therefore, differences in previous results may be due to using dissimilar ways to measure comprehension as well as diversity of types of marginal annotation approaches. Therefore, the results of the previous studies were maladjusted, which can be noticed in Martinez-Fernandez's (2010) study where comprehension questionnaires were used, free recall tasks in L1 in Bell and LeBlanc's research (2000) and multiple-choice comprehension items in Ko's study (2005). Building on the presumption that marginal annotated lexical items may gain heightened attention from the learners, which led to incidental acquisition of the lexical items in terms of form and meaning, the recent studies about marginal annotations have addressed their impact on FL vocabulary learning.

Many studies (Nagata, 1999; Bowles, 2004) have revealed that the positive effect of the marginal annotations on FL vocabulary acquisition was small, but it still exists. At the same time, the proportional efficiency of the different types of marginal annotations in different studies (Guidi, 2009; Rott, 2005) may be varied. Regardless of the small positive effects

of marginal annotations on the acquisition of FL vocabulary, some researchers (Boers, 2000; Hulstijn & Laufer, 2001) considered that the act of giving the learners ways to facilitate grasping the meanings of the words may have no significant effects on learning FL vocabulary, as the FL learners were not urged to make a mental effort to obtain the meanings of words. Resulting from this comes Involvement Load Theory, which was suggested by Laufer and Hulstijn (2001) to complete the reading task and assess the semantic and grammatical fit of the concluded meaning in context, as well as to infer the meaning of an ambiguous word. Therefore, the absence or presence of those three components affect the learners in terms of involvement in varied levels of engagement, which may affect the learners' ability to retain the words' meanings over the long term.

Some studies (Watanabe, 1997; Gettys et al., 2001) argued that learners' inferring abilities are largely varied and different from one another. In addition, the learners may be exposed to inadequate or misleading data that make the learners' inferences inaccurate. The studies of Martinez-Fernandez (2010), Rott and Williams (2003), and Rott (2005) had mixed results when they posed an experimental question to find out whether involvement in inference/search operations—in Hulstijn and Laufer's (2001) words, greater engagement load—has a stronger effect than marginal annotations on FL vocabulary learning. The current study discovered that the previous studies on the effect of annotations on FL grammatical development is still meager (i.e., Guidi, 2009; Martinez-Fernandez, 2010) comparing to studies conducted on FL vocabulary learning. The results of Guidi's (2009) study showed that the impacts of marginal annotations was restricted, as it affected only learning of Spanish

present perfect and impersonal, which enhances possible interaction between the impacts of marginal annotations and inherent characteristics of linguistic features of the target forms, like the abstractness of meaning-referent. In line with Guidi's findings, Martinez-Fernandez's (2010) study showed that the two elements used in the study (fill-in tasks and marginal annotations) were not successful in enhancing the learning of the Spanish subjunctive, while they were dynamic in learning the target items of FL Spanish vocabulary.

Contrary to previous studies, Nagata's (1999) study found positive effects for marginal annotations when used with Japanese learners for the purpose of learning three Japanese grammatical structures. As a result of the deficiency in the number of studies that search for the impacts of the marginal annotations on learning of FL grammatical items, there are methodological differences among those studies. It can be concluded that the researcher of the present study was enthusiastic about studying the effects of marginal annotations on reading comprehension for FL learners, as well as some other issues related to their learning (vocabulary and grammar). To achieve the goal of the study, the researcher developed research questions to help ensure that the study is more accurate. These questions were formulated as follows:

- 1- If there is any positive effect for marginal annotations on the extent to which target FL grammatical structures are gained, what is this effect?
- 2- If there is any positive effect for marginal annotations on the extent to which target FL new vocabulary is gained, what is this effect?
- 3- If there is any positive effect for marginal annotations on FL reading comprehension, what is this effect?

Method

Participants

The learners who participated in this study were 104 (28 males and 76 females) who ranged in age between 22 and 23 years old. All participants were in their fourth year of college in the English department of the faculty of education in a university in Egypt. The researcher chose participants randomly without any plan or bias and divided them into two groups: 52 in the marginal annotations group and 52 in the control group without using marginal annotations. The participants did not receive any direct instructions about the target constructions before participating in this research. In line with Jung's (2016) study, the current study used a 'Reading and use of English' section of a practice Cambridge Proficiency: English (CPE) test (Cambridge English Language Assessment, 2013) (Cronbach's alpha: 1951, 702) to calculate the participants' English proficiency level. The role of this step was to guarantee symmetry among the English proficiency capacities of the participants. The next step was using the scores to build on when applying the random samples to ensure equivalence in English proficiency between the groups, as well as minimizing sampling errors.

Design of the study

As mentioned before, the study divided the participants into two groups that took part in four sessions. The researcher not only divided the groups, but also divided the time into four weeks and created a plan for each week. In the first week, both groups completed the pre-test (which included a grammaticality judgment test only), background questionnaire, and FL proficiency test. The reason for using a grammaticality judgment

test only in the pre-test and avoid utilizing the target words was that the target words were pseudo-word items, which means they didn't need a pre-test. In the second week, the researcher employed treatment 1 and a post-reading questionnaire for groups and utilized treatment 2, a post-reading questionnaire, and an immediate post-test in the third week. Finally, in week four, the participants in both groups used a delayed post-test and exit questionnaire.

Tools

The tools used in this study included: targeted FL features, texts, and questionnaires. These parts will be clarified as follows:

Targeted FL features

The current study used 10 pseudo-words to explore whether marginal annotations have effects on incidental learning of lexical items, as well as employed English unaccusativity, a grammatical feature. Previous studies (Shin, 2011; Jung, 2016) showed that the learners of FL experience difficulty in gaining English unaccusativity. The Unaccusativity Hypothesis was presented for the first time by Perlmutter (1978), which is a special kind of intransitive verb divided into either unergatives (i.e., John dined) or unaccusatives (i.e., the window broke). Semantically, unergative verbs have a subject perceived as actively initiating or actively responsible for the action expressed by the verb while the subject of an unaccusative verb does not actively initiate or is not actively responsible for the action of the verb, or it treats the argument like the accusative argument of a transitive verb.

The learners of FL always make errors when dealing with unaccusative verbs, as mentioned by some studies (Zobl, 1989; Croft,

1995). Hwang (2001) indicated that the reasons for these errors may be varied, as FL learners may tend to make too many errors if an unaccusative verb has its transitive counterpart (i.e., boat, change, shut, fracture), whereas the FL learners with non-alternating unaccusative verbs (i.e., occur, outcome, come, vanish) may have less of a tendency to make errors. Some other researchers (No & Chung, 2006; Lee et al., 2008) have identified some other elements that may raise the difficulty of gaining English unaccusativity for FL learners. These elements are transfer from or to the first language, low input frequency, and the absence of a conceptualizable factor.

In terms of the impacts of input frequency, the results of previous studies (Lee et al., 2008) showed that English unaccusative verbs might be gained in an item-based manner, as in lexis-specific learning. Thence, the learning may stretch to more abstract constructions (i.e., construction-based generalization of models) and low-frequency factors. Two treatment texts were used in the current research and have specified 17 English unaccusative verbs to work as the target features (see Appendix A). In both texts, all the unaccusative verbs appeared just once and selected six verbs only without alternating transitive counterparts, while the other verbs had just one alternating transitive counterpart. The frequency of the target verbs was lower than 11,000 per 450 million, according to the Corpus of Contemporary American English (COCA: Davies, 2008).

Concerning the pseudo-words, it was previously mentioned that the current study used 10 pseudo-words to serve as the lexical target features after replacing 10 lexical items in the treatment texts (see Appendix A). The present study benefited from Pulido's (2007) conditions in choosing

the lexical items. Firstly, to control the part of speech, the word should be a noun; secondly, to control the frequency, the word should appear just once in the text. The third condition was about whether the word was in a context that permits learners to deduce the meaning of the word (the Egyptian learners of English were checked to ensure their ability to infer the meaning of the word). Pulido (2007) added that the pseudo-words should follow English orthographic and morphological rules; each of the words used as pseudo-words should include seven letters and consist of two syllables. In addition, plurality was also marked in the pseudo-words when the original word was in its plural form.

Texts

The current study depended on two standards to select the texts: firstly, choosing uncommon and/or difficult topics to show to the learners; secondly, a sufficient number of occurrences of the target grammatical feature. The two texts were selected from passages used in TOEFL tests; both of them were different, but at the same time they were comparable in terms of readability and length (see Appendix A). The first text was discussing the formation, extraction, and refinement of natural gas resources and challenges that appeared in its use. The second text was about the Ediacaran Period and reviews macroscopic fossil evidence of soft-bodied organisms that can be found in a few localities around the world, confirming Darwin's expectations. The previous studies (Senter & Smith, 1967; Coleman & Liau, 1975; Gunning, 1952; McLaughlin, 1969; Flesch, 1948) were beneficial in helping the present research to measure the average readability from various readability indicators, including the

Automated Readability index, Coleman–Liau index, Gunning– Fog score, SMOG index, and Flesch–Kincaid grade level.

Questionnaires

The present research used three different types of questionnaire with different purposes; firstly, to collect data about the participants' learning experiences in English and know their demographic characteristics, a background questionnaire was used. Secondly, the researcher asked the learners to give information about their perceived standards when facing familiar or difficult topics in reading a text by giving them a post-reading questionnaire. The last questionnaire was an exit questionnaire that aimed to make the participants talk about the treatment sessions and provide backdated comments.

Treatment tasks

What participants do when dealing with a reading text of a TOEFL test was the based on which treatment task relies in the current study. The researcher embedded reading comprehension measures into the learning task to be able to calculate the learners' understanding level as well as the fulfillment of the task. The present research benefited from Educational Testing Services (2012) in two ways; firstly, using the multiple-choice reading comprehension items that were taken from the TOEFL tests, then developed and tested by ETS; secondly, identifying the shape and the instructions of the reading comprehension items that demanded direct or indirect treatment of the target constructions in this study. The participants in this study found that the reading comprehension items asked them to identify references, comprehend rhetorical targets, recognize the meaning of vocabulary words that appeared in the text, put sentences into a

paragraph, pick out main ideas of the text, facilitate/rewriting a sentence, and determine realistic/negative data. The researcher divided the texts into five parts; each part consisted of one or two paragraphs, and the learners had the right to keep the text while answering the related reading comprehension questions that came directly after the paragraph. In terms of the scores of the reading comprehension, the current research followed Cronbach's alpha (1951) in making 15 the maximum score for each test. The participants were given Arabic translations of the pseudo-word items and the target unaccusative verbs in the texts as marginal annotations.

Assessment tasks

The current study used both multiple-choice and meaning recognition tests, as well as the grammaticality judgment test; the first test was used to evaluate learning the pseudo-word items, while the second test was used to measure learning of the unaccusative verbs. To know whether the learning gained is automatized, the present research followed the study of Jung (2016) into accomplish these measures with reaction time data. At the same time this study benefited from the study of Rebuschat and Williams (2012) in checking the solidity of the knowledge obtained via collecting twofold confidence grades. In the grammaticality judgment test, this study used 80 randomized sentences shown on a computer screen. As previously mentioned, this research used 17 target unaccusative verbs, and for each of them the researcher constructed one grammatical and one ungrammatical passive sentence until they reached 34 sentences (i.e., the moon was soon vanished vs. the stress soon vanished).

This study has get help from Cobb's (2011) Complete Lexical Tutor version 6.2 by using the list of 2,000 most frequently used English words

to choose another 8 extra verbs to be able to investigate whether the learning from the treatment has transmitted to other unaccusative verbs (see Appendix A). The unaccusative eight verbs were divided into two parts; four of them were alternating while the other four were nonalternating, but these verbs reached 16 sentences after using one grammatical and one ungrammatical sentence. The study of Schneider, Eschman, and Zuccolotto (2012) was beneficial for utilizing E-Prime 2.0 for reaction time analysis to construct the grammaticality judgment test. Another 30 extra sentences were used as distracters and were divided into two parts; half of them were grammatical and the other half were ungrammatical sentences.

The researcher asked the participants to answer very quickly after the sentences were presented on the computer screen by clicking on the button 'g' if the sentence was grammatical and the button 'u' if the sentence was ungrammatical. The researcher asked the learners after responding to rate the level of confidence in their response ('high confidence' vs. 'low confidence) to make a twofold resolution. The time used in the test ranged from ten to twelve minutes, and the maximum score was 60. The study of Cronbach's alpha (1951) presented .665 for novel sentences and .802 for target sentences (as mentioned in Jung, 2016).

The second test used in this study was multiple-choice and meaning recognition test, including forty items derived by using E-Prime 2.0 to gauge the learning of the target pseudo-words. The forty items were divided into three parts; part one followed the model used in Leeser's study (2007) where the target was to measure the capacity of the learners to recognize the target pseudo-word forms using 20 items. In part two, there

were ten items as the target pseudo-words, while part three included the other ten items to be distracters employed in the form recognition task. The learners were given four choices and were asked to choose the right Arabic translation of the given target word (Martinez-Fernandez, 2010). The researcher asked the participants to answer very quickly after the items were presented on the computer screen by clicking on the button 'g' if they remember seeing the word in the texts and the button 'u' if they didn't remember. The researcher asked the learners after responding to rate the level of confidence in their response ('high confidence' vs. 'low confidence) to make a twofold decision. The time used in the test ranged from six to seven minutes, and the maximum score for both the form and meaning recognition target test items were 20 scores. The study of Cronbach's alpha (1951) presented .452 for the meaning recognition test and .586 for the form recognition test (as mentioned in Jung, 2016).

Procedures

The time of the experiment has been divided into four sessions; all of them took place in the computer lab at the faculty of education. The time used in each session ranged from forty- five to sixty minutes. In the first session, the learners were asked to take the three tests: FL proficiency, background, and the questionnaire test. The participants took part in the treatment in both sessions 2 and 3: after that, they completed an immediate post-test in addition to a post-reading questionnaire. Finally, the researcher demanded the learners complete both the exit questionnaire and a delayed post-test in the fourth week; thus, the data over the four sessions were ready to collect.

Data Analysis

The current study followed IBM (2011) in using SPSS 20.0 in order to analysis the statistics. To sum up all the data collected over the four sessions, the researcher calculated the descriptive statistics for the tests. Alpha level of $p < .05$ was set as the significance level, as well as utilizing Cronbach's alpha to specify the inner reliability of the different tests. In addition, different forms of ANOVA were carried out to direct the questions of the study, and in order to check the mutual relations between the different test scores, the Pearson's coefficients were calculated.

Results

The researcher divided this section into five parts as follows:

Part (A) Introductory analysis

Part (B) FL grammar learning after using marginal annotations

Part (C) FL vocabulary learning after using marginal annotations

Part (D) Stability of the acquired knowledge

Part (E) FL reading comprehension after using marginal annotations

Part (A) Introductory analysis

The researcher conducted a one-way ANOVA analysis twice, firstly to examine the equivalence of English proficiency standard among the groups of participants ($F(3) = 0.002, p = .978$) which revealed that there was no significant difference between the two sets; secondly, on the grammaticality judgment pre-test to confirm that all the sets began at a statistically comparable developmental phase (one of the target elements) of the target elements. The findings showed that there was no significant difference among the sets in terms of their capacity to judge the grammaticality of the English unaccusative sentences in both sides; novel

items: $F(3) = 0.026$, $p = .910$ and Target items: $F(3) = 0.124$, $p = .804$ (see Table 1). Post-reading questionnaire items (i.e., I thought this topic of the reading was familiar; I had some background knowledge about the reading topic) were used to measure the impacts of the learners' acquaintance with every subject. The results in both texts were Text 1: $r(104) = 1.368$, $p < .001$, Text 2: $r(104) = 1.128$, $p < .001$, which means that the responses to the two items were significantly correlated with each other. A simple regression analysis with the sum of responses was utilized with the reading comprehension scores as the dependent variable and the two questionnaire items as the independent variable to examine the influence of topic familiarity on reading comprehension. The findings in both texts were Text 1: $F(3) = 0.406$, $p = .654$, Text 2: $F(3) \leq 0.02$, $p = .995$, which means that the learners' topic familiarity with the texts did not significantly influence their scores on the reading comprehension items.

Table 1. Illustrative statistics for the grammaticality judgment test

| Group | Test | N | Target items | | | Novel items | | |
|------------------------------|---------------------|----|--------------|------|------|-------------|-------|------|
| | | | M | M.G | SD | M | M.G | SD |
| With marginal annotations | Pre-test | 52 | 35 | – | 8.1 | 20.46 | – | 4.84 |
| | Immediate post-test | 52 | 40.24 | 5.24 | 9.36 | 20.38 | -0.08 | 5.96 |
| | Delayed post-test | 52 | 44.92 | 9.92 | 8.2 | 22.16 | 1.7 | 4.62 |
| Without marginal annotations | Pre-test | 52 | 35.54 | – | 7.44 | 20.62 | – | 4.9 |
| | Immediate post-test | 52 | 38.46 | 3.16 | 8.88 | 21.92 | 1.16 | 4.58 |
| | Delayed post-test | 52 | 39.62 | 3.84 | 9.08 | 22.62 | 2.16 | 4.1 |

Note. MG =mean gain, Maximum score for: novel items = 52, target items = 68.

Part (B) FL grammar learning after using marginal annotations

The previous table shows the descriptive statistics for the grammaticality judgment test scores by group, where the average gain scores were lower in the instantaneous post-test than in the delayed post-test. In addition, Table 2 clarifies that the mean accuracy scores were generally lower for the target items than those for the novel items. In the second step, a mixed-model ANOVA was computed with marginal annotations as the between-time and entrant element (i.e., pre-test, immediate post-test, and delayed post-test) as the within-partaker factor on the target grammaticality judgment test elements.

The major significant impact for time was $F(4,2) = 61.46, p < .001$, partial $\eta^2 = 0.762$, and for time and marginal annotations, $F(4,2) = 10.778, p = .006$, partial $\eta^2 = 0.194$. In terms of mixed-model ANOVAs the follow-up showed that there were significant differences in the target grammaticality judgment scores between immediate and delayed post-test, $F(3) = 9.078, p = .038$, partial $\eta^2 = 0.166$, and between pre-test and delayed post-test, $F(3) = 20.804, p = .002$, partial $\eta^2 = 0.344$. On the other hand, there was no significant difference between pre-test and immediate post-test, $F(3) = 2.95, p = .230$, partial $\eta^2 = 0.058$. Based on this, the findings indicated that marginal annotations had a positive effect on learning the target unaccusative elements, and the delayed post-test proved that. Additionally, the rates of accuracy were raised for grammatical and ungrammatical sentences; this can be seen in Table 2. A mixed-model ANOVA was combined with marginal annotations as the between-entrant element and time as the within-entrant factor for the sake of the novel

grammaticality judgment test items. The major significant impact for time was observed as $F(4,2) = 8.76, p = .015, \text{partial } \eta^2 = 0.162$, while the matter was different for time and marginal annotations, $F(4,2) = .1.31, p = .522, \text{partial } \eta^2 = 0.026$. It was clear from the accuracy rates that both groups (with marginal annotations and without marginal annotations) were to some extent equal in learning to utilize of novel unaccusative verbs (see Table 2).

Table 2. Accuracy Rates

| Group | Test | N | Grammatical | Ungrammatical | Target items | Novel items |
|------------------------------|---------------------|----|-------------|---------------|--------------|-------------|
| With marginal annotations | Pre-test | 52 | 1.28 | 0.92 | 1.02 | 1.28 |
| | Immediate post-test | 52 | 1.4 | 1.04 | 1.18 | 1.28 |
| | Delayed post-test | 52 | 1.48 | 1.22 | 1.32 | 1.38 |
| Without marginal annotations | Pre-test | 52 | 1.3 | 0.94 | 1.04 | 1.28 |
| | Immediate post-test | 52 | 1.36 | 1.06 | 1.14 | 1.36 |
| | Delayed post-test | 52 | 1.4 | 1.1 | 1.16 | 1.42 |

Note. Maximum score = 2

Part (C) FL vocabulary learning after using marginal annotations

The findings revealed that the mean scores on the meaning recognition test were generally lower than those on the form recognition test, and this can be seen in Table 3, which shows the descriptive statistics for the vocabulary recognition scores by group. In addition, the mean meaning recognition scores on the immediate post-test were higher than those on the delayed post-test, while the mean form recognition scores from the immediate post-test were lower than those from the delayed post-test. In both immediate and delayed post-tests, the same prompts were

used, so the higher scores on the form recognition test may be due to practice efficacy because the given test was repeated. In the second step, a mixed-model ANOVA was computed with marginal annotations as the between-time and entrant element as the within-partaker factor on the vocabulary form recognition test.

The major significant impact for time was $F(4,2) = 612.096, p < .001$, partial $\eta^2 = 1.72$, as well as interaction impact for time and marginal annotations, $F(4, 2) = 8.332, p = .018$, partial $\eta^2 = 0.154$. The results revealed that marginal annotations had a positive effect in the immediate post-test in terms of remembering the target pseudo-word forms, while in the delayed post-test, it didn't have the same effect. This was clear after follow-up two-way ANOVAs, which showed that there was no significant influence of marginal annotations on the vocabulary form recognition scores on the delayed post-test, $F(3) = 3.838, p = .172$, partial $\eta^2 = 0.074$, whereas it revealed that there were significant impacts on the immediate post-test, $F(3) = 16.452, p = .006$, partial $\eta^2 = 0.282$. Once more, another mixed-model ANOVA was calculated with marginal annotations as the between-participant element and time as the within-participant element for the sake of the vocabulary meaning recognition test items.

The major significant impact for time was $F(4, 2) = 223.932, p < .001$, partial $\eta^2 = 1.382$, as well as the interaction influence of time and marginal annotations, $F(4,2) = 9.834, p = .009$, partial $\eta^2 = 0.18$. In terms of mixed-model ANOVAs the follow-up in two-way to investigate the influences of marginal annotations on the scores for the vocabulary meaning recognition test showed that there were significant impacts of marginal annotations on the scores for vocabulary meaning recognition test in both the immediate

post-test $F(3) = 15.678$, $p = .007$, partial $\eta^2 = 2.72$, and in the delayed post-test, $F(3) = 14.862$, $p = .009$, partial $\eta^2 = 0.258$. Based on this, the findings indicated that marginal annotations had a positive effect on learning the meanings of the target pseudo-vocabulary (see Table 3).

Table 3. Illustrative Statistics for the Vocabulary Recognition Tests

| Group | Test | N | Form | | Meaning | |
|------------------------------|---------------------|----|-------|------|---------|------|
| | | | Mean | SD | Mean | SD |
| With marginal annotations | Immediate post-test | 52 | 12.46 | 4.3 | 7.46 | 2.98 |
| | Delayed post-test | 52 | 13.16 | 3.92 | 6.46 | 2.68 |
| Without marginal annotations | Immediate post-test | 52 | 9.3 | 3.58 | 4.92 | 3.54 |
| | Delayed post-test | 52 | 11.62 | 4.08 | 4.16 | 3.4 |

Note: Maximum score for form recognition = 20, meaning recognition = 20.

Part (D) Stability of the acquired knowledge

Binary confidence ratings and reaction times were further examined to investigate the stability of the learned knowledge.

Firstly, reaction times:

Table 4 shows that using mixed-model ANOVA uncovered that reaction times diminished notably from one test time to the next; Grammatical: $F(4.2) = 48.95$, $p < .001$, partial $\eta^2 = 0.658$; Ungrammatical: $F(4.2) = 64.622$, $p < .001$, partial $\eta^2 = 0.786$. Marginal annotations were used to make considerable progress in this way for the ungrammatical sentences, $F(4.2) = 6.178$, $p = .050$, partial $\eta^2 = 0.116$, but not for the grammatical sentences, $F(4.2) = 0.156$, $p = .925$, partial $\eta^2 = 0.004$. It was obvious that especially when using marginal annotations that shorter reaction times correspond to higher grammaticality judgment test scores

for the ungrammatical sentences in the immediate post-test. Additionally, paired-samples t-tests revealed that the participants took longer to react to the ungrammatical sentences than the grammatical sentences, although this barely missed significance in the post-tests (the pre-test, $t(102) = -4.812$, $p = .020$, $t(102) = -2.004$, $p = .051$, and in the delayed post-test, $t(102) = -3.876$, $p = .058$) (see Table 5).

Table 4. Median of Reaction Times for The Grammaticality Judgment Test (milliseconds).

| Group | Test | N | Grammatical | | Ungrammatical | |
|------------------------------|---------------------|----|-------------|------|---------------|------|
| | | | Mean | SD | Mean | SD |
| With marginal annotations | Pre-test | 52 | 11474 | 3308 | 13576 | 6054 |
| | Immediate post-test | 52 | 10580 | 3140 | 11210 | 3580 |
| | Delayed post-test | 52 | 8506 | 2362 | 9332 | 2446 |
| Without marginal annotations | Pre-test | 52 | 12396 | 3756 | 13406 | 4462 |
| | Immediate post-test | 52 | 9686 | 2534 | 10354 | 3552 |
| | Delayed post-test | 52 | 8828 | 2882 | 8996 | 2446 |

Table 5. Linkages between Grammaticality Judgment Test Scores and Reaction Times.

| Group | Test | N | Grammatical | Ungrammatical |
|------------------------------|---------------------|----|-------------|---------------|
| With marginal annotations | Pre-test | 52 | -0.234 | -0.212 |
| | Immediate post-test | 52 | -0.618 | -0.91* |
| | Delayed post-test | 52 | 0.026 | -0.562 |
| Without marginal annotations | Pre-test | 52 | -0.176 | -0.502 |
| | Immediate post-test | 52 | -0.474 | -0.786* |
| | Delayed post-test | 52 | 0.174 | -0.412 |

Note: Significance level: * $p < .05$.

For both word form and meaning recognition, reaction times appeared to be reduced. In addition, a mixed-model ANOVA exhibited no significant difference in reaction times to the form recognition between the immediate and delayed post-tests, $F(3) = 0.228$, $p = .737$, partial $\eta^2 = 0.004$. Furthermore, while reaction times to the meaning recognition test diminished significantly by time, $F(2.78) = 62.572$, $p < .001$, partial $\eta^2 = 0.89$, marginal annotations were shown to have no significant effect on this trend, $F(2.78) = 3.112$, $p = .220$, partial $\eta^2 = 0.076$ (See Table 6). A contrasting model of the correlation between reaction times and form/meaning recognition has been presented in Table 7, which shows that while form recognition correlated negatively with reaction times, meaning recognition was shown to have a positive connection with reaction times overall. Moreover, paired-samples t-tests disclosed that meaning recognition took significantly longer than form recognition in both post-tests; the immediate $t(90) = -19.622$, $p < .001$, and in the delayed $t(102) = -5.592$, $p = .008$ (see Table 7).

Table 6. Median of Reaction Times for the Vocabulary Recognition Test (milliseconds).

| Group | Test | N | Form | | N | Meaning | |
|------------------------------|---------------------|----|-------|------|----|---------|------|
| | | | Mean | SD | | Mean | SD |
| With marginal annotations | Immediate post-test | 52 | 10580 | 3140 | 52 | 11210 | 3580 |
| | Delayed post-test | 52 | 8506 | 2362 | 50 | 9332 | 2446 |
| Without marginal annotations | Immediate post-test | 52 | 12396 | 3756 | 40 | 13406 | 4462 |
| | Delayed post-test | 52 | 8828 | 2882 | 38 | 8996 | 2446 |

Note: Absent values were kept out analysis by analysis

Table 7. Illustrative Statistics for the Vocabulary Recognition Tests

| Group | Test | Form | Meaning |
|------------------------------|---------------------|--------|---------|
| With marginal annotations | Immediate post-test | -0.98* | -0.288 |
| | Delayed post-test | -0.524 | 0.264 |
| Without marginal annotations | Immediate post-test | -0.622 | 0.246 |
| | Delayed post-test | -0.42 | 1.078** |

Note: Significance level: * $p < .05$, ** $p < .001$.

Confidence ratings:

A one-sample t-test has been used to count profit outcomes via deducting pre-test scores from immediate/delayed post-test scores were tested versus zero; this step occurred before inquiring into participants' confidence ratings for the grammaticality judgment test. In condition to marginal annotations, the significance was noticed for the immediate target items, $t(102) = 8.054$, $p < .001$, and for the delayed target items, $t(102) = 14.632$, $p < .001$. Also, in cases without marginal annotations, the significance was observed as well for the immediate target items, $t(102) = 4.478$, $p = .034$, delayed target items, $t(102) = 6.44$, $p = .004$, and the delayed novel items, $t(102) = 5.668$, $p = .009$ (see Table 8).

Table 8. Significance of Profits in the Grammaticality Judgment Test and d' Values

| Group | | Mean | d' |
|------------------------------|------------------|--------|--------|
| With marginal annotations | Immediate_target | 5.24** | 0.14 |
| | Delayed_target | 9.92** | 1.04** |
| Without marginal annotations | Immediate_novel | -0.08 | |
| | Delayed_novel | 1.7 | |
| | Immediate_target | 3.16* | -0.04 |
| | Delayed_target | 3.84** | 0.26 |
| | Immediate_novel | 1.16 | |
| | Delayed_novel | 2.16** | 0.98* |

Note: Significance from zero: * $p < .05$, ** $p < .001$.

Based on the significant profits observed for each of the grammaticality judgment tests, sensitivity index d_s were computed in the current study with the same technique developed by Kunimoto et al. (2001) and used also in Jund's study (2016). The value of d' or below zero is supposed to refer to a propensity to report high confidence for wrong responses and low confidence for correct responses. In contrast, a negative d' indicates low confidence for incorrect responses and high confidence for correct responses. It was obvious that with marginal annotations, d_s were shown to be significantly higher than zero for the delayed target items, $t(102) = 7.998$, $p < .001$, and for the delayed novel items without marginal annotations, $t(102) = 4.386$, $p = .038$, while d_s were shown to be not significantly different from zero in all other cases (see Table 8).

Participants tended to report raised levels of confidence in their correct responses to the target items in the delayed post-test with marginal annotations. Furthermore, all empirical cases findings in the vocabulary recognition tests were significantly bigger than zero. It was clear that all d' values were significantly greater than zero in form recognition items while d' values were shown to be not significantly different from zero or significantly less than zero in meaning recognition items. Participants generally reported low confidence in their right responses to the meaning recognition items while reporting high confidence in their correct responses to the form recognition items (see Table 9).

Table 9. Indication of profits for the d' values and vocabulary recognition test

| Group | Test | Form | | Meaning | |
|------------------------------|---------------------|---------|--------|---------|-------|
| | | Mean | d' | Mean | d' |
| With marginal annotations | Immediate post-test | 12.46** | 2.26** | 7.46** | -0.3 |
| | Delayed post-test | 13.16** | 1.88** | 6.46** | 0.52 |
| Without marginal annotations | Immediate post-test | 9.3** | 2.3** | 4.92** | -1.06 |
| | Delayed post-test | 11.62** | 2.98** | 4.16** | 0.3 |

Note: Significance from zero: *p < .05, **p < .001.

Part (E) FL reading comprehension after using marginal annotations

According to Table 10, the scores of reading comprehension were higher in marginal annotation cases than without marginal annotations. Furthermore, the scores of reading comprehension on text two were on average higher than those for text one. On the other hand, in order to check the impacts of marginal annotations on reading comprehension, a one-way ANOVA was used, and the findings showed that there were no significant influences on marginal annotations on reading comprehension scores for either text one: $F(3) = 0.77$, $p = .538$ or text two: $F(3) = 4.402$, $p = .144$ (see Table 10).

Table 10. Descriptive Statistics for Reading Comprehension Scores

| Group | N | Text one | | | Text two | | |
|------------------------------|-----|----------|------|------|----------|------|------|
| | | Mean | SD | SE | Mean | SD | SE |
| With marginal annotations | 52 | 22.46 | 3.98 | 0.78 | 26.3 | 2.82 | 0.56 |
| Without marginal annotations | 52 | 21.7 | 4.92 | 0.96 | 25.08 | 3.16 | 0.62 |
| Total | 104 | 22.8 | 4.44 | 0.62 | 25.7 | 3.02 | 0.42 |

Note: Maximum score = 30.

Discussion and conclusions

The current study's findings were in line with Jung's (2016) study, which clarified that marginal annotations have simplified learning of a group of English unaccusative verbs, as was clear in the scores on the delayed post-test. On the other hand, the results of the present study contradicted the findings of Guidi's (2009) and Martinez-Fernandez's (2010) studies, which demonstrated that marginal annotations had a small impact on learning grammatical features only (i.e., present perfect and Spanish subjunctive). Based on Zyzik's (2009) study, one possible reason behind the influences found in the current study is that English unaccusativity could be viewed as a lexico-grammatical feature that requires item-based learning, and is thus more volatile to marginal annotations, whereas present perfect and Spanish subjunctive involve complex conjugations. Zyzik (2009) confirmed that exposure to language-specific morphological and syntactic manifestations are the only way for the unaccusative/unergative distinction of a certain language to be learned.

Some other studies found significant effects of marginal annotations; i.e., in Nagata's (1999) study, where the target Japanese grammatical structures were also lexico-grammatical features. It was noticed that the target verbs in the current study could have been learned with an item-based method because learning that did not convey novel unaccusative verbs adds more ratification. On the other hand, the results of the current study agreed with those of some previous studies (i.e., Lee et al., 2008) in that exposure to input is an effective element in learning of English unaccusative verbs. Furthermore, some studies (i.e., Levin & Rappaport Hovav, 1995; Sorace & Shomura, 2001) detected that syntactic diagnostics

of unaccusative verbs tend to recognize semantically coherent subsets of verbs. This point was clarified by Sorace (2000, 2004) as learning unaccusative verbs may require both a data-driven process triggered by exposure to positive proof of individual unaccusative verbs and top-down implementation of the syntactic restrictions to new unaccusative verbs that subsume similar lexical-semantic features. The present study in fact seeks to highlight the necessity of taking into consideration the nature of the target constructions alongside the possible utility of marginal annotations for facilitating learning of FL grammatical features.

It was proposed by Guidi (2009) that the influences of marginal annotations might interact with the type of target linguistic item concerning its complexity, abstractness, and appropriateness. In addition, it was revealed that the impacts of marginal annotations occurred for both grammatical and ungrammatical grammaticality judgment tests, which means that the entrants learned that they cannot use a certain unaccusative verb in the passive voice; however, it could be utilized in the active voice. Thus, this area needs more research into how marginal annotations affect learning of various types of FL grammatical features, which may lead to fruitful outcomes regarding the combined impacts of marginal annotations and the nature constructions of the target language.

Depending on the findings of the confidence ratings and the data of the reaction time, it can be said that marginal annotations played a role faster recognition of ungrammatical stimulates as well as powerful confidence in the right answers to the target items in the post-tests, which reinforces the advantages of marginal annotations for learning English unaccusative verbs. Ungrammatical sentences tend to take longer to make

a judgment because there is no structural impersonation in the learner's internal grammar, so it was not surprising in this study to find that it overall took longer to judge ungrammatical sentences than grammatical sentences. Based on this finding, Juffs (2001) suggested that the participants might try various analyses before considering the sentence ungrammatical and then giving up.

Results showed that marginal annotations made it easy to learn the target pseudo-words. Furthermore, the complex cognitive processes involved in meaning recognition seem to be reflected through some elements; relatively lower confidence ratings compared with form recognition, along with longer reaction times and long-term influences of marginal annotations on learning word meanings. In other words, word meaning recognition demands producing phonemic recoding from a registered series of letters and finding its semantic qualities in the lexical inventory. As well as identifying visual patterns in the perceived letters which is enough for form recognition. Thus, the reason behind more vigorous retention might be deeper processing of the meaning provided in marginal annotations than the word form.

Although there was no significant difference found between the experimental group and control group in terms of reading comprehension scores, the scores were higher with the group who used marginal annotations. At the same time, the performance of the participants in general was better on reading comprehension tests; thus, the maximum impact may hide between-group differences, as shown in the comparatively high mean scores and small SDs. Furthermore, more difficult reading comprehension tests, as expected by Horiba (2000) and

Yoshimura (2006) may lead to a need to search for the marginal annotations, as well as affecting the reliability of the reading comprehension tests and increase the capabilities of SDs. Horiba (2013) added that increasing the level of reading comprehension difficulty may lead to discover whether FL learning from engaged in reading tasks actually happens without interrupting comprehension.

It can be concluded that former studies have rarely proved the positive impacts of marginal annotations on FL grammar learning, so the current study has attempted to spotlight these effects. Besides, the reinforced solidity of the knowledge gained has been proven through the ratings of twofold confidence and the response latency data. At the same time, it should take into consideration that the scores on the grammaticality judgment test in all conditions (with and without marginal annotations and novel and target items) were increased, and that the benefits were not considerable, but still significant. Consequently, to uncover the clearer function of marginal annotations in enhancing gaining FL grammatical features, more experimental examinations are needed. The findings of the present study revealed significant impacts of marginal annotations on vocabulary acquisition, besides demonstrating that marginal annotations may present various supports to learning forms and meanings of words. The current research suggests using online data (e.g., think-aloud protocols) to be able to conduct closer investigations of the differential effectiveness of marginal annotations in gaining FL lexical features. In conclusion, to investigate whether marginal annotations can act as a catalyst of FL learning and FL reading comprehension at the same time,

the discriminability of reading comprehension measures needs to be carefully considered and modified.

References

- Al-Seghayer, K. (2001). The effects of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, 5, 202–232.
- Bell, F., & LeBlanc, L. (2000). The language of glosses in L2 reading on computer: Learners' preferences. *Hispania*, 83, 274–285.
- Boers, F. (2000). Enhancing metaphoric awareness in specialized reading. *English for Specific Purposes*, 19, 137–147.
- Bowles, M. (2004). L2 glossing: To CALL or not CALL. *Hispania*, 87, 541–552.
- Cambridge Proficiency Test: Cambridge English Language Assessment. (2013). *Cambridge English: Proficiency (CPE) practice test*. Retrieved from <http://www.cambridgeenglish.org/exams/proficiency/how-to-prepare/>
- Cobb, T. M. (2011). *Compleat Lexical Tutor*. Retrieved from <http://www.lex tutor.ca/>
- Coleman, M., & Liau, T. L. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60, 283–284.
- Croft, W. (1995). Modern syntactic typology. In M. Shibatani & T. Bynon (Eds.), *Approaches to language typology* (pp. 85–144). New York: Clarendon Press.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297- 334.
- Davis, J. N., & Lyman Hager, M. (1997). Computers and L2 learning: Student performance, student attitudes. *Foreign Language Annals*, 30, 58–72.
- Davies, M. (2008–). *The Corpus of Contemporary American English (COCA): 400+ million words, 1990-present*. Retrieved from <http://www.americancorpus.org/>
- Educational Testing Services. (2012). *Official TOEFL iBT Tests*. New York: McGraw-Hill.
- Eskey, D. E. (2005). Reading in a second language. In E. Hinkel (Ed.), *Handbook of research in Second language teaching and learning* (pp. 563–579). Mahwah, NJ: Lawrence Erlbaum.
- Gettys, S., Imhof, L., & Kautz, J. O. (2001). Computer-assisted reading: The effect of glossing format on comprehension and vocabulary retention. *Foreign Language Journals*, 34, 91–99.
- Godfroid, A., & Uggem, M. S. (2013). Attention to irregular verbs by beginning learners of German: an eye-movement study. *SSLA*, 35, 291–322.
- Grabe, W. (2009). *Reading in a second language: moving from theory to practice*. Cambridge: Cambridge University Press.

- Guidi, C. (2009). Glossing for meaning and glossing for form: A computerized study of the effects of glossing and type of linguistic item on reading comprehension, noticing, and L2 learning. Unpublished PhD dissertation, Georgetown University, Washington, DC, USA.
- Horiba, Y. (2000). Reader control in reading: Effects of language competence, text type, and task. *Discourse Processes*, 29, 223–267.
- Horiba, Y. (2013). Task-induced strategic processing in L2 text comprehension. *Reading in a Foreign Language*, 25(2), 98–125.
- Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51, 539–558.
- Hwang, J. B. (2001). Focus on form and the L2 learning of English unaccusative verbs. *English Teaching*, 56, 111–133.
- IBM Corp. (2011). *IBM SPSS Statistics for Windows (Version 20.0)*. Armonk, NY: IBM Corp.
- Juffs, A. (2001). Psycholinguistically oriented second language research. *Annual Review of Applied Linguistics*, 21, 207–221.
- Jung, J. (2016). Effects of glosses on learning of L2 grammar and vocabulary. *Language Teaching Research*, 20(1), 92–112.
- Ko, H. M. (2005). Glosses, comprehension, and strategy use. *Reading in a Foreign Language*, 17, 125–143.
- Ko, H. M. (2012). Glossing and second language vocabulary learning. *TESOL Quarterly*, 46, 56–79.
- Koda, K. (2012). Development of second language reading skills: Cross-linguistic perspectives. In S.M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 303–318). London: Routledge.
- Krashen, S. (1982). *The input hypothesis*. Oxford: Pergamon Press.
- Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition*, 10, 294–340.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1–26.
- Lee, S.-K., Miyata, M., & Ortega, L. (2008). A usage-based approach to overpassivization: The role of input and conceptualization biases. Paper presented at the 26th Second Language Research Forum, Honolulu, HI, USA, October 17–19.
- Leow, R.P. (2009). Modifying the L2 reading text for improved comprehension and acquisition: Does it work? In Z.-H. Han & N.J. Anderson (Eds.), *Second language reading research and instruction: Crossing the boundaries* (pp. 83–100). The University of Michigan Press.
- Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity: At the syntax–lexical semantics interface*. Cambridge, MA: MIT Press.

- Martinez-Fernandez, A. M. (2010). Experiences of remembering and knowing in SLA, L2 development, and text comprehension: A study of levels of awareness, type of glossing, and type of linguistic item. Unpublished PhD dissertation, Georgetown University, Washington, DC, USA.
- Nagata, N. (1999). The effectiveness of computer-assisted interactive glosses. *Foreign Language Annals*, 32, 469–479.
- No, G., & Chung, T. (2006). Multiple effects and the learnability of English unaccusatives. *English Teaching*, 61, 19–39.
- Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. *Applied Linguistics*, 28, 66–86.
- Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, 33, 829–856.
- Robinson, P., Mackey, A., Gass, S. M., & Schmidt, R. (2012). Attention and awareness in second language acquisition. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 247–267). London: Routledge.
- Rott, S. (2005). Processing glosses: A qualitative exploration of how form–meaning connections are established and strengthened. *Reading in a Foreign Language*, 17, 95–124.
- Rott, S., & Williams, J. (2003). Making form–meaning connections while reading: A qualitative analysis of word processing. *Reading in a Foreign Language*, 15, 45–75.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools, Inc.
- Sharwood, M. S. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, 15, 165–179.
- Shin, J. A. (2011). Overpassivization errors in Korean college students' English writings. *Korean Journal of Applied Linguistics*, 27, 255–273.
- Sorace, A. (2000). Gradients in auxiliary selection with intransitive verbs. *Language*, 76, 859–890.
- Sorace, A. (2004). Gradience at the lexicon–syntax interface: Evidence from auxiliary selection and implications for unaccusativity. In A. Alexiadou, E. Anagnostopoulou, & M. Everaert (Eds.), *The unaccusativity puzzle: Explorations of the syntax–lexicon interface* (pp. 243–268). Oxford: Oxford University Press.
- Sorace, A., & Shomura, Y. (2001). Lexical constraints on the acquisition of split intransitivity. *Studies in Second Language Acquisition*, 23, 247–278.
- VanPatten, B. (2012). Input processing. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 268–281). London: Routledge.
- Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *SSLA*, 19, 289–307.

Yoshimura, F. (2006). Does manipulating foreknowledge of output tasks lead to differences in reading behavior, text comprehension and noticing of language form? *Language Teaching Research*, 10(4), 419–434.

Zobl, H. (1989). Canonical typological structures and ergativity in English L2 acquisition. In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language* (pp. 203–221). Cambridge: Cambridge University Press.

Zyzik, E. (2009). The role of input revisited: Nativist versus usage-based models. *L2 Journal*, 1, 42–61.

Appendix A

A-Target English unaccusative verbs in text 1 & text 2.

| Text one | | | Text two | | |
|--------------------|-----------------|---------------------------|--------------------|-----------------|---------------------------|
| Unaccusative Verbs | Alternating | Frequency per 450 million | Unaccusative Verbs | Alternating | Frequency per 450 million |
| Putrefy | Alternating | 624 | excavate | Alternating | 22 |
| Abate | Non-Alternating | 1136 | date to | Alternating | 1486 |
| Ascend | Alternating | 1518 | emanate | Alternating | 2.044 |
| Compile | Alternating | 3.628 | composed of | Non-Alternating | 4.28 |
| Abruption | Alternating | 5.108 | persevere | Non-Alternating | 5.368 |
| Minify | Alternating | 5.402 | develop | Alternating | 6.368 |
| Eroding | Non-Alternating | 8.954 | Vanish | Non-Alternating | 15.162 |
| Gather | Alternating | 21.05 | appear | Non-Alternating | 18.232 |
| Stabilize | Alternating | 21.746 | | | |

B-Target pseudo-words

| Text one | | Text two | |
|----------------|--------------|----------------|--------------|
| original words | pseudo-words | original words | pseudo-words |
| faucets | golands | Absence | Fration |
| undermost | tarragon | Fluctuations | Cabrons |
| shores | stovons | Predator | Tralion |
| gardens | klaners | Successors | Morbis |
| discoveries | phosens | Evidences | Zenters |

C- Text features

| | Text one | Text two |
|---------------------|-----------------------|----------------------|
| Title | Natural gas resources | The Ediacaran Period |
| Number of words | 682 | 699 |
| Average readability | 23.2 | 26.8 |