

# Online Dictionaries and the Involvement Load Hypothesis: An Empirical Study

John B. Collins<sup>1</sup>

## Abstract

This paper explores the relatively under-researched area of online dictionary usage by EFL learners. Despite the inconvenient truth that students are increasingly choosing to utilize online dictionaries over their traditional paper and handheld electronic counterparts, this shift has not been sufficiently reflected in EFL literature – particularly in terms of how such technology could be affecting incidental vocabulary acquisition and reading comprehension. This paper sought to explore this line of inquiry within the frameworks of the involvement load (Laufer and Hulstijn 2001) and consultation trigger point hypotheses (Aust, Kelly, and Roby, 1993). Thirty undergraduate students studying English as a foreign language at a Japanese university completed a reading task followed by a reading comprehension test and a delayed vocabulary recall test. During the reading task, students were assigned a dictionary tool to use, either Weblio or Google Translate, with varying involvement load indexes. A battery of Mann-Whitney tests did not identify significantly different performances under the two dictionary conditions. A small to medium effect size was identified ( $r = -0.25$ ), indicating that dictionary type was impacting vocabulary recall performance, albeit modestly.

**Key Terms:** Online dictionaries, Google Translate, Japanese EFL students, incidental vocabulary acquisition, reading comprehension, involvement load hypotheses, consultation trigger point

## 1. Introduction

The arrival of digital technology and the Internet has led to what has been described as a revolution in the field of lexicography and the attitudes that learners have towards dictionaries (Lew & de Schryver, 2014). No longer are students required to carry about with them a heavy paper (PD) or handheld electronic dictionary (ED). Indeed all a student requires in Japan, where the present study took place, is a Wi-Fi connection for a limitless amount of language data to be at their fingertips, either on a computer, smartphone, or other electronic device. This shift has made the task of defining the term “electronic dictionary” increasingly problematic. Nesi (2000) defined electronic dictionaries as “...any reference material stored in electronic form that gives information about the spelling, meaning, or use of words.” (p.839). This definition includes such tools as a spell-checker in a word processing program, devices that scan and translate printed words, and electronic versions of traditional paper dictionaries. This definition, therefore, can also be applied to online versions of printed dictionaries such as the Merriam-Webster, Oxford, and Longman dictionaries. As Lorentzen and Theilgaard (2012) have shown, not only are a large number of students now accessing such online dictionaries via a one-word entry into Internet search engines, but also that their choice of search engine is overwhelmingly Google (p.654). However, rather than serving purely as a search engine, Google has itself now become a provider of definitions through the Google Translate tool, which as the author's previous study (Collins, 2016) has shown, has become an extremely popular

<sup>1</sup>Lecturer, Ritsumeikan Asia Pacific University (APU), Beppu City, Oita, Japan

e-mail:crm11427@apu.ac.jp

tool among Japanese EFL college students. It could be argued, therefore, that Google Translate now fulfills the role of a dictionary for a great number of students. Google Translate, therefore, is one of the tools on which the present study focusses.

The present study was carried out in order to build on the author's previous research (Collins, 2016), which investigated the increasing use of online dictionaries among Japanese college students. The study noted two important implications relating to vocabulary acquisition as a result of a potentially lowered "involvement load" (Laufer & Hulstijn 2001) and also a potentially lowered "consultation trigger point" (Aust, Kelly, & Roby, 1993). The purpose of the present study is to explore these two implications and to begin the process of establishing an empirical basis for encouraging, or discouraging, students from using certain types of online dictionary tools, if indeed such a need exists. While the present study did not produce any statistically significant results on which to base any assertions, it did identify a small to medium effect size ( $r = -0.25$ ) in favor of Condition #2 (Weblio) over Condition #1 (Google Translate) in terms of vocabulary recall performance.

## **1.1 Literature review**

The aim of the following literature review is threefold: first, to highlight the gap that currently exists in dictionary research in regard to online dictionaries and their potential impact on incidental vocabulary acquisition and reading comprehension; second, to frame the present study within the existing literature dealing with the involvement load hypothesis; and third, to position the present study within the framework of the consultation trigger point.

### **1.1.1 Framing the current study**

There is a large body of research into dictionary use by second language (L2) learners, and its impact on L2 reading comprehension and incidental vocabulary acquisition. Studies comparing the relative benefits of EDs and traditional PDs in regard to spelling, reading comprehension, rates of dictionary consultation, and vocabulary acquisition have, however, painted only a blurry picture. A number of studies have concluded in favor of EDs over traditional PDs in terms allowing students to look up a greater number of words within a similar or shorter period of time (Koyama & Takeuchi, 2004(b), 2007; Shizuka, 2003). Any benefits to reading comprehension, however, have yet to be clearly established. Studies by Koyama & Takeuchi (2004(b), 2007), for example, did not identify any significant improvement in reading comprehension. A further study by Koyama and Takeuchi (2004a) did, however, conclude that users of PDs displayed significantly higher rates of vocabulary retention. As previously argued (Collins, 2016), while such studies have been vital in establishing a conceptual framework within which to explore the relative benefits of dictionary types (see Koyama, 2004, and Töpel, 2014, for comprehensive reviews), the recent and potentially irreversible shift amongst students towards favoring free online dictionary tools over both EDs and PDs has yet to be adequately taken into account, indeed research into online dictionaries is very much still in its infancy. In fact the first complete book focusing on the use of online dictionaries was only very recently published (Müller-Spitzer, 2014) and was based, for the most part, on

studies involving participants with a professional involvement in language and linguistics such as lexicographers, linguists and translators – not L2 learners. It is this gap in the research that the present study aims to begin addressing.

### 1.1.2 The involvement load hypothesis

The primary aim of this study is to investigate implications relating to the involvement load hypothesis (Laufer and Hulstijn, 2001) and its impact on incidental vocabulary acquisition. Building on the “depth of processing” construct ( Craik & Lockhart, 1972), researchers have argued that the meaning and form of a word is more likely to be successfully retained by the learner if such information is acquired through a process that requires high mental effort (Hulstijn, 1992). Based on these assumptions, the involvement load hypothesis was first proposed by Laufer and Hulstijn (2001) and postulates that the effectiveness of vocabulary learning, that is to say the chances of successful vocabulary retention, depends on the level of cognitive effort or *involvement* that a task requires from the learner. Described as a motivational-cognitive construct, *involvement* is made up of the sum of three components: *need*, *search* and *evaluation* (Hulstijn & Laufer, 2001, p.543). The first component, *need*, describes the level and source of the motivation to search for a word. Need is *moderate* when it is imposed by the teacher or by the nature of the task, such as a reading task followed by comprehension questions. Need is strong when motivation is self-imposed by the learner (p. 543), such as requiring a word in order to write an original piece of text. *Search* refers to the processes involved when searching for the required word in a dictionary or other source. Here the distinction between a *moderate* and *strong* search is whether the search is being conducted to find the meaning of a given word (moderate), or to find a suitable word for a known meaning (strong) (Laufer, 2000, p. 53). The final component, evaluation, relates to how the student reaches a decision about which definition or word is the most suitable for a particular context. For example, if a student consults a dictionary that presents multiple definitions, the student must decide which is the most suitable by comparing the sentence in their book with the definitions offered in the dictionary. Such an evaluation carries a *moderate* evaluation load. If, however, the student were writing an original piece of text and consulted the dictionary for a word they required for a known meaning, they would need to evaluate how the new word, and its meaning, combines with the original sentence or text. Such a process would carry a *strong* evaluation load. By assigning a numerical value (0, 1 or 2) to each of these three components, an *involvement index* (maximum index value = 6) can then be calculated and language learning tasks can be compared in terms of the respective involvement loads that each task carries. Since the present study is primarily concerned with comparing dictionaries and the respective consultation and evaluation processes that each induce, the *search* and *evaluation* components are of most interest and relevance.

There is a limited base of research which supports the involvement load hypothesis. Hulstijn (1992) concluded that retention is higher when the meaning of an unknown word is inferred by the student (high cognitive load) than when the meaning is given to them (low cognitive load) (Hulstijn, 1992, p. 122). Hulstijn & Laufer (2001) compared the short- and long-term retention of

words learned through three tasks of differing involvement loads. Task one (involvement index=1) involved a reading comprehension task with marginal glosses provided. Task two (involvement index=2) involved a reading comprehension task combined with a vocabulary gap-fill task. Task three (involvement index=3) involved a writing composition task in which students had to use ten provided words. As their hypothesis predicted, the higher the involvement load, the better the vocabulary was retained (p.552). Whereas the Hulstijn and Laufer (2001) study focused on advanced students, Keating (2008) conducted a similar study involving three tasks of similarly varying involvement loads. While the results mirrored Hulstijn and Laufer (2001), Keating (2008) noted that when time on task is taken into consideration, however, retention gains did not hold (p.379). Knight (1994) identified significantly higher reading comprehension scores and vocabulary retention rates among learners who used dictionaries during a reading task than those who did not (p.291). Keating (2008) subsequently suggested that these gains could be attributed to the higher involvement load incurred by the search aspect (p.369). Knight's (1994) findings are not universal, however, and contradictory findings have been made (Bensoussan, Sim & Weiss, 1984; Nesi & Mera, 1991). It is a primary aim of the present study, therefore, not only to continue the process of exploring the validity of the involvement load hypothesis, but also to do so in a way that better reflects the dictionary preferences of increasingly technology-savvy EFL students.

### **1.1.3 The consultation trigger point**

The secondary focus of the present study deals with the *consultation trigger point*, which is described by Aust, Kelly and Roby (1993) as the point at which a student decides to consult a dictionary to confirm the meaning of an unknown word or phrase. Previous studies have shown that students using EDs not only have a positive attitude toward them, but also that they looked up more words than students using PDs (Aust, Kelly & Roby, 1993; Koyama & Takeuchi, 2004, 2007; Shizuka, 2003). Conversely, studies have also shown that students using PDs spend more time and effort on guessing or remembering words before consulting their dictionary (Kobayashi, 2007; Koyama & Takeuchi, 2007). Aust, Kelly and Roby (1993) concluded that because of a more efficient access to the word reference information, in their case a hyper-referenced glossary, students were more inclined to consult it (p.70). In other words, it could be said that ED technology has lowered the consultation trigger point and therefore encourages students to look up more words. As previously argued (Collins, 2016), the streamlined search processes of online dictionary technology has potentially lowered the consultation trigger point further still, possibly to a point where students are looking up too many words. Indeed caution has been voiced that, in such cases, dictionary use should be discouraged (Tang, 1997) or at the very least students should be encouraged to be more selective when consulting their dictionary (Prichard, 2008). A lowered consultation point is potentially exacerbated further by the use of online tools such as Google Translate, which, as described in detail in section 2.3, have essentially eliminated the evaluation stage of the consultation process and therefore streamlined the process to such a degree that students may be encouraged to consult such dictionary tools even more frequently.

### 1.1.4 The relationship between dictionary usage and reading comprehension

The relationship between vocabulary and reading comprehension has been well established (Nation, 2001; Grabe, 2009) and there is a considerable body of research that supports the position that vocabulary knowledge is a strong predictor of successful reading (see Grabe, 2009, for review). While the present study is concerned primarily with vocabulary retention in relation to the involvement load hypothesis, an additional aim is to tentatively explore the possibility that the use of Google Translate could be negatively affecting reading comprehension (as a result of the absence of example sentences and other information with which to make an informed decision about a definition's suitability, as described in detail in section 2.3). Assuming that the use of dictionaries during a reading task helps the reader to raise their lexical coverage of the text, and therefore improve their chances of better text comprehension (Laufer, 1989; Hu & Nation, 2000), it stands to reason that Google Translate may potentially be lowering the chances of successful comprehension as a result of the reader's adoption of an unsuitable definition. If this were the case, it would be reflected in lower reading comprehension scores for students under the Google Translate condition. Furthermore, there are a number of comparative studies that investigate the impact of dictionary type on reading comprehension (Koyama & Takeuchi, 2004(b), 2007; Aust, Kelly & Roby, 1993). While these studies have not produced any significant findings in support of certain dictionary types, and in light of the gap which currently exists in the literature surrounding online dictionaries and reading comprehension, the author felt that it was worthwhile to compare reading comprehension performance under the two conditions (described in 2.3) and to compare the results with previous studies.

The goal of the present study, therefore, was to continue the process of investigating online dictionaries, and dictionary-like tools, particularly in regard to their impact on vocabulary retention and the number of words being looked up by students. The secondary aim was to investigate to what extent, if any, reading comprehension performance is affected by online dictionaries, and dictionary-like tools. These aims are reflected in the three research hypotheses listed below.

## 2. Research methodology

### 2.1 Research hypotheses

The following research was carried out over a two week period and consisted of a reading task with the aid of a designated online dictionary, a reading comprehension test, and a delayed vocabulary recall test. In this way, the following three hypotheses were examined:

- 1) Assuming a *lower involvement load*, participants using Google Translate would show lower retention rates on the delayed vocabulary test than students using a typical online bilingual dictionary,
- 2) Assuming a lower consultation trigger point, participants using Google Translate would look up more words, and,
- 3) Participants using Google Translate would display lower reading comprehension scores than students using a typical online bilingual dictionary.

## 2.2 Participants and methodology

Participants comprised of 30 upper-intermediate Japanese-language basis students studying general English as part of an undergraduate degree. The 30 students came from four different class groups taught by two different teachers (the researcher and a colleague teaching the same level). While the participants were all at the upper-intermediate level, each class group was divided equally into the two task conditions in order to minimize any possible impact that varying student proficiency could have on the experiment. During the first stage, students were asked to read a passage taken from the textbook that is used in the students' reading skills and vocabulary development focused B-course (see appendix for full text). Students were allowed to consult the online dictionary that was designated to them, namely, Google Translate or Weblio (a description of each condition and basis for the selection of Google Translate and Weblio is described in section 2.3, below). The students were given 25 minutes to read the text and were instructed to underline all the words they looked up. Once the students had read the text, they were instructed to answer 14 comprehension questions which were taken from the textbook. Having used the textbook throughout the semester, participants were familiar with both the style of text and type of comprehension questions. One week later, participants were presented with an unannounced vocabulary recall test which comprised of all the words they had underlined. Each test, therefore, was individualized and had a different number of items. One point was awarded for each correct definition. Participants were not permitted to use a dictionary during the recall test (see Figure 1 for procedure). In order to individualize each recall test, and in light of the narrow time frame, students were required only to provide a correct Japanese translation for each word they had looked up (the limitations of this approach are described in section 4.0) The research was completed outside of regular class time and participants were compensated with a book voucher.

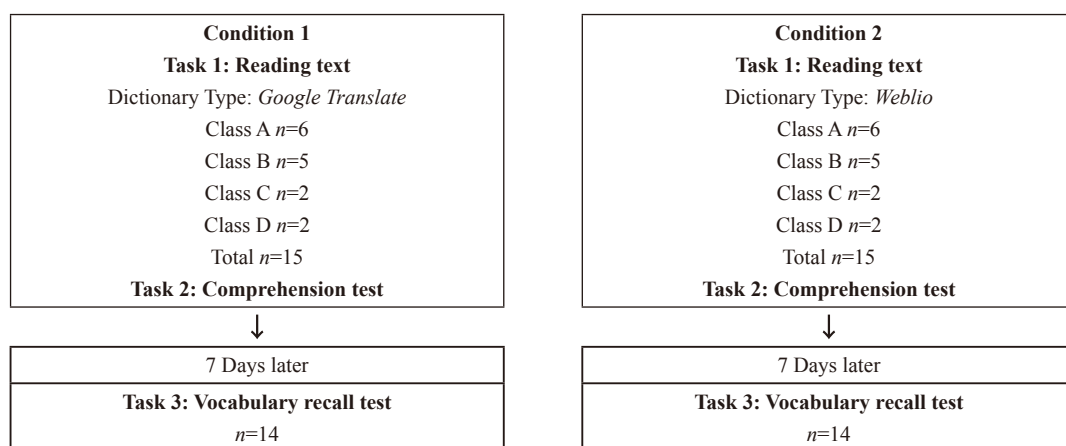


Figure 1. Experiment procedure

### 2.3 Google Translate, Weblio and the consultation process

The dictionaries used in the current study were selected based on the results of Collins (2016) which indicated that bilingual online dictionaries and online translation software were the most frequently used dictionary tools among 498 Japanese-language-basis undergraduate students (p.42). The first dictionary, Weblio (Figure 2), was chosen based on the author's previous in-class observations which indicated that Weblio was widely used by students at the university where this research took place. Based on this assumption, a screenshot of the Weblio website was used in the Collins (2016) survey as the example of an online dictionary – the results of which indicated that online bilingual dictionaries were the most popular choice for students accessing such tools via their PC or smartphone. Weblio is a free online English-Japanese/Japanese-English dictionary service. According to the company homepage, Weblio provides a comprehensive online dictionary service with word entries and other related information sourced from a range of specialist and general dictionaries and glossaries that have been registered with the website. When users search for a word, Weblio displays a list of entries based on an inter-disciplinary search (Weblio FAQ, 2016). Google Translate (Figure 3) is also a free online dictionary-like tool which, when used in its simplest form, provides a word or full sentence translation based on a statistical algorithm, rather than authentic dictionary entries and texts. Google Translate was selected on the basis that it was, according to the results of Collins (2016), the second most popular choice of PC/smart-phone based dictionary tool, and also on the grounds that the results of Lorentzen and Theilgaard (2012), as described above, indicated that the Google search engine is an extremely popular entry point for students seeking the definition of a word online.



Figure 2. Weblio – a typical free online bilingual dictionary



Figure 3. Google Translate

Nation (2001, p.419) described four steps learners go through when consulting a dictionary for unknown words:

1. Get information from the context where the word occurred
2. Find the dictionary entry
3. Choose the most suitable sub-entry
4. Relate the meaning to the context and decide if it fits.

As shown in Figure 3, Google Translate, as presented in its simplest form, does not display sub-entries; a translation is simply provided. Therefore there is no way of comparing other possible sub-entries, (or in this case, different translations) with the source text. It is clear, then, that Nation's third and fourth steps may be missed by the learner. Consequently, there are two potentially adverse outcomes for the learner in terms of vocabulary acquisition. Firstly, since the learner does not have access to a range of sub-entries from which to choose, they may simply assume that the definition presented to them is suitable, regardless of whether it is correct in that particular context or not. Secondly, as a result of not being able to compare and contrast sub-entries and example sentences with the original text, the depth of cognitive engagement with which the students engaged the information presented to them may potentially be reduced. In other words, it could be argued that the user-friendly design of Google Translate reduces the *involvement load* of the consultation process and therefore adversely affects the chances of successful incidental vocabulary acquisition. Compared with popular bilingual dictionaries such as Weblio (Figure 2), metalanguage and other important vocabulary-related information is also not present. The consultation process is extremely straightforward: the learner simply types in the word and a translation is provided. It could therefore also be argued that the simple design and ease of use of Google Translate encourages learners to look up more words. In other words, it could be lowering the *consultation trigger point*. Weblio, on the other hand, does provide a range of sub-entries, grammatical metalanguage, and



other important information, such as example sentences, that the learner could refer to. Whether or not they actually do, however, is another matter, as Tseng (2009) has indicated. In a study of the Yahoo! online dictionary search behavior of 38 EFL high school students, Tseng (2009) concluded that a considerable number of errors, in which students had provided incorrect definitions, could be attributed to students not making full use of the example sentences available in the dictionary (p.102)

#### **2.4 Input loads of each dictionary condition**

Following the examples of both Laufer and Hulstijn (2001) and Keating (2008), an involvement index was assigned to each task condition. Condition #1 (Google Translate) was calculated thus: *need* = 1, *search* = 1, and *evaluation* = 0 (total involvement index = 2). As described above, because Google Translate in its simplest form does not offer a range of sub-entries, the evaluation process cannot be carried out. Therefore a zero score was assigned for *evaluation*. Likewise, Condition #2 (Weblio) was calculated thus: *need* = 1, *search* = 1, and *evaluation* = 1 (total involvement index = 3). While the difference in involvement load is admittedly modest, the absence of an evaluation stage under the Google Translate condition does offer a window for reading comprehension and vocabulary retention variance. Furthermore, while a case could be made for the evaluation aspect to carry a heavier respective weighting due to its arguably more cognitively demanding nature, it is beyond the scope of the present study.

### **3. Results**

Descriptive statistics of the results are provided in Table 1. “Class code” represents the class group of which each participant was a member. The task conditions (Google Translate (GT) and Weblio (WEB)) were equally represented across the four class groups. “Net Words” indicates the number of words each participant looked up during the initial reading stage. “Comprehension” represents each participants' score on the post-reading comprehension questions. “Vocab Recall” represents each participants' score on the follow-up vocabulary recall test.

Table 1:  
*Descriptive Statistics for Google Translate (GT) and Weblio (WEB) Groups*

Subject #	Class Code	Net Words	Comprehension (%)	Vocab Recall (%)
GT 1	A	0	100	n/a*
GT 2	A	14	92.86	7.14
GT 3	B	6	100.00	33.33
GT 4	A	6	92.86	0.00
GT 5	B	13	78.57	38.46
GT 6	C	6	100.00	50.00
GT 7	A	2	85.71	50.00
GT 8	A	17	71.43	23.53
GT 9	B	5	78.57	60.00
GT 10	B	8	92.86	12.50
GT 11	D	19	92.86	36.84
GT 12	C	9	100.00	11.11
GT 13	D	10	85.71	10.00
GT 14	B	1	85.71	0.00
GT 15	A	12	78.57	41.67
N=15	AVE:	8.5	89.05	26.76
Subject #	Class Code	Net Words	Comprehension (%)	Vocab Recall (%)
WEB 1	A	3	92.86	66.67
WEB 2	A	6	85.71	100.00
WEB 3	B	8	100.00	25.00
WEB 4	A	0	92.86	n/a
WEB 5	B	1	100.00	0.00
WEB 6	B	19	92.86	63.16
WEB 7	D	5	78.57	25.00
WEB 8	A	17	78.57	52.94
WEB 9	A	39	78.57	20.51
WEB 10	B	6	92.86	50.00
WEB 11	A	8	100.00	50.00
WEB 12	C	6	100.00	66.67
WEB 13	B	13	92.86	23.08
WEB 14	D	40	78.57	15.00
WEB 15	C	5	100.00	0.00
N=15	AVE:	11.7	90.95	39.86

\*As subjects GT 1 and WEB 4 did not look up any words, they could not complete the vocabulary recall test.

**3.1 Results of statistical analysis**

In order to identify any relationship between students' rate of successful vocabulary recall under the two task conditions, a Mann-Whitney U test was conducted that indicated that vocabulary recall scores for participants using Google Translate (Mdn = 28.43%) were not significantly lower than for participants using Weblio (Mdn = 37.50%),  $U = 69$ ,  $p = 0.181$ , and that the effect size was small to medium ( $r = -0.25$ ). Results are presented in Table 2.

Table 2:  
*Results of Mann-Whitney U test for vocabulary recall scores*

Ranks					Test Statistics	
	Group	N	Median (%)	Sum of Ranks	Vocab. Recall Scores	
Vocab. Recall Scores	Weblio	14	37.50	174.00	Mann-Whitney U	69.00
	Google Translate	14	28.43	232.00	Asymp. Sig. (2-tailed)	0.181*
	Total	28				

(\*Correlation significant at the <0.05 level)

With regard to Hypothesis 2, the results of a Mann-Whitney U test indicated that the number of words looked up by participants using Google Translate (Mdn = 8) was not significantly higher than for participants using Weblio (Mdn = 6),  $U = 112.00$ ,  $p = 0.983$ . Results are presented in Table 3.

Table 3:  
*Results of Mann-Whitney U test for net words looked up*

Ranks					Test Statistics	
	Group	N	Median (%)	Sum of Ranks	Net Words Looked Up	
Net Words Looked Up	Weblio	15	8	232.00	Mann-Whitney U	112.00
	Google Translate	15	6	233.00	Asymp. Sig. (2-tailed)	0.983*
	Total	30				

(\*Correlation significant at the <0.05 level)

With regard to Hypothesis 3, an additional Mann-Whitney U test indicated that reading comprehension scores for participants using Google Translate (Mdn = 92.86%) were not significantly lower than for participants using Weblio (Mdn = 92.86%),  $U = 99.50$ ,  $p = 0.576$ . In fact, the respective median scores were exactly the same. Results are presented in Table 4.

Table 4:  
Results of Mann-Whitney U test for reading comprehension

Ranks					Test Statistics	
	Group	N	Median (%)	Sum of Ranks	Reading Comp. Scores	
Reading Comprehension Score	Weblio	15	92.86	219.50	Mann-Whitney U	99.50
	Google Translate	15	92.86	245.50	Asymp. Sig. (2-tailed)	0.576*
	Total	30				

(\*Correlation significant at the <0.05 level)

#### 4. Discussion, implications and conclusion

The vocabulary recall rates for students under both conditions were not significantly different, therefore Hypothesis 1 was not supported, although the initial data did show a large difference in mean retention score (Google Translate: AVE = 26.76%, Weblio: AVE = 39.86%). A small to medium effect size ( $r = -0.25$ ) was, however, identified indicating that Google Translate may be detrimental to incidental vocabulary acquisition. Considering this effect size, future studies could benefit from having a larger data sample. Hypothesis #2 was also not supported, that is to say, students using Google Translate were not looking up significantly more words than students using Weblio. It appears, then, that contrary to the author's expectations, when compared with Weblio, the streamlined search process of Google Translate has not led to a lowering of the consultation trigger point. Hypothesis #3 predicted that participants using Google Translate would produce lower reading comprehension scores. This was not supported by the current data; indeed under both conditions students produced extremely high comprehension scores. This result is consistent with previous studies (Koyama & Takeuchi, 2004(b), 2007; Aust, Kelly & Roby, 1993). As described in more detail below, this could be due to the reading text employed.

As described above, the involvement load index variance between the two conditions was modest (Weblio = 3, Google Translate = 2). Such a small difference in involvement load no doubt makes the task of identifying a significant variance in task performance problematic, if indeed such a variance exists. Furthermore, while the sample size was comparable with similar studies, a sample of 30 students is also quite modest and larger-scale studies should be carried out to replicate the current study before any concrete conclusions can be made. In addition to the small sample size, there are a number of further limitations which should be described here. Firstly, the nature of the vocabulary recall tests. As described above, students were required to provide a correct translation for each word they had looked up. While there are a number of ways to measure vocabulary knowledge, such as multiple-choice questions, matching words with synonyms, supplying a correct translation, and yes/no self-report word recognition tests (see Read, 2000, p87), a translation test was chosen because of the researcher's desire to only assign points if a student displayed a correct understanding of a word as it appeared *within the context of the text* – something which, given the time constraints, could not easily be achieved with another type of test question. However, as Read (2000) points out, a potential drawback of this approach is the necessity for the researcher to make

a subjective decision on the correctness of each translation. While the researcher in this case is a fluent speaker of the students' L1 (Japanese) and has over five years' experience as a Japanese-English translator, the subjectivity of the recall test grading must be acknowledged as a potential limitation. Future studies would benefit from a multi-faceted testing approach. Perhaps the most important limitation of the current study, however, relates to the reading comprehension text itself. Since the text was taken from a textbook that the participants were familiar with, it could be argued that the text and accompanying comprehension questions were simply not challenging enough for the students, indeed the average reading comprehension test scores were 89% (Weblio) and 91% (Google Translate) respectively. This is despite the considerable range in number of words looked up under both conditions (min = 0, max = 40). This raises questions about participants' reasons for consulting a dictionary in the first place. Rather than to find the definition of a completely unknown word, many participants may have simply wished to *confirm* what they already felt they knew. While it is beyond the scope of the present study, this could be a fruitful line of inquiry for future studies. Although the range of comprehension scores was very narrow, the results of an additional Spearman's Rank correlation test identified a statistically medium negative correlation between the number of words looked up and comprehension scores ( $r = -0.369$ ,  $p < 0.5$ ). It could be said, therefore, that students who looked up the lowest number of words tended to produce higher reading comprehension scores. This is consistent with previous studies (Bensoussan, Sim & Weiss, 1984, Knight, 1994). However, it appears that access to dictionaries was potentially allowing less proficient students to achieve high, and arguably inflated and temporary, comprehension scores. While this is beyond the scope of the current study, it is a potentially concerning implication in terms of how it undermines the validity of reading comprehension performance.

This study aimed to begin the process of addressing the gap in applied linguistics research which has arisen due to the advance of online dictionary tools into territory hitherto dominated by traditional paper and electronic dictionaries. While the results of the current study did not produce any ground-breaking or concrete findings, it went some way towards laying the groundwork for further comparative studies and illustrated how the involvement load index can be applied to online dictionary tools.

## References

- Aust, R., Kelly, M. J., & Roby, W. (1993). The use of hyper-reference and conventional dictionaries. *Educational Technology Research and Development*, 41 (4), 63-73.
- Bensoussan, M., Sim, D. & Weiss, R. (1984). The effect of dictionary usage on EFL test performance compared with student and teacher attitudes and expectations. *Reading in a Foreign Language*, 2, 262-276.
- Collins, J. B. (2016). Changes in electronic dictionary usage patterns in the age of free online dictionaries: Implications for vocabulary acquisition, *APU Journal of Language Research*, 1, 36-49.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research, *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Grabe, W. (2009). *Reading in a second language: moving from theory to practice*. New York, Cambridge University Press.
- Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13, 403 – 430.
- Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. Arnaud & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics*, 113–125. London: Macmillan.
- Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51 (3), 539-558.
- Keating, G. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research*, 12 (3), 365-386.
- Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities, *The Modern Language Journal*, 78 (3), 285–299.
- Kobayashi, C. (2007). Comparing electronic and printed dictionaries: Their effects on lexical processing strategy use, word retention, and reading comprehension. In K. Bradford-Watts (Ed.), *JALT2006 Conference Proceedings*. Tokyo: JALT.
- Koyama, T. (2004). Dictionary use in L2 learning: A critical review, *Bulletin of Osaka Ohtani University*, 48, 53-66.
- Koyama, T., & Takeuchi, O. (2004a). Comparing electronic and printed dictionaries: How the difference affected EFL learning. *JACET Bulletin*, 38, 33-46.
- Koyama, T., & Takeuchi, O. (2004b). How (sic) look-up frequency affects EFL learning?: An empirical study on the use of handheld-electronic dictionaries. In W.M. Chan, K. N. Chin, P. Martin-Lau, and T. Suthiwan (Eds.), *Proceedings of the CLaSIC 2004 Conference*.
- Koyama, T., & Takeuchi, O. (2007). Does look-up frequency help reading comprehension of EFL learners? Two empirical studies of electronic dictionaries, *CALICO Journal*, 25, 110-125.
- Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.): *Special language: From humans thinking to thinking machines*, 316-323, Clevedon: Multilingual Matters.
- Laufer, B. (2000). Task effect on instructed vocabulary learning: the hypothesis of 'involvement'. *Selected Papers from AILA '99 Tokyo*. Waseda University Press, Tokyo, 47-62.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22, 1-26.
- Lew, R. & De Schryver, G.-M. (2014). Dictionary Users in the Digital Revolution. *International Journal of Lexicography*, 27 (4), 341-359.
- Lorentzen, H., & Theilgaard, L. (2012). Online dictionaries: how do users find them and what do they do once they have? In R.

- Vatvedt Fjeld, & J. M. Torjusen (Eds.): *Proceedings of the 15th EURALEX International Congress*, 654-660. Oslo.
- Müller-Spitzer, C. (Ed.) (2014). *Using online dictionaries*. Berlin, Boston: De Gruyter Mouton.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nesi, H. (2000). Electronic dictionaries in second language vocabulary comprehension and acquisition: the state of the art. In U. Heid et al. (eds.): *IX EURALEX International Conference*, 839-847. Stuttgart.
- Nesi, H., & Meara, P. (1991). How using dictionaries affects performance in multiple-choice EFL tests. *Reading in a Foreign Language*, 8 (1), 631-643.
- Prichard, C. (2008). Evaluating L2 readers' vocabulary strategies and dictionary use. *Reading in a Foreign Language*, 20 (2), 216-231.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Shizuka, T. (2003). Efficiency of information retrieval from the electronic and the printed versions of a bilingual dictionary. *Language Education and Technology* 40, 15-33.
- Tang, G.M (1997). Pocket electronic dictionaries for second language learning: Help or hindrance? *TESL Canada Journal / Revue TESL du Canada*, 15 (1), 39-57.
- Töpel, A. (2014). Review of research into the use of electronic dictionaries. In Müller-Spitzer, C. (Ed.): *Using online dictionaries*. pp. 13-54, Berlin: de Gruyter.
- Tseng, F. P. (2009). EFL students' Yahoo! online bilingual dictionary use behavior. *English Language Teaching Journal*, 2 (3), 98-108.
- Weblio FAQ. (2016) Retrieved from <http://www.weblio.jp/info/faq.jsp>

## Appendix A Reading Text

### *The Science of Smiles*

Why do we smile? Many people automatically assume that there is a simple answer to that question – we smile because we are happy. That answer is correct, but it doesn't tell the whole story. Social scientists who study smiles say there's a lot more to smiling than just showing happiness. Smiling can actually have a great impact on a person's quality of life.

Marianne LaFrance is a social scientist who is particularly interested in smiles. She has studied smiles for over 20 years. LaFrance says that we use smiles to make and maintain relationships. We need to do this because we are social animals. As social animals, we need strong relationships in order to survive and thrive<sup>1</sup>. According to LaFrance, smiling is one of the most important tools to maintain social relationships. For example, smiling makes it easier to make new friends. One reason for this is that we are attracted to people who smile. Smiling can put people at ease. Smiling also helps people make the best of unexpected conditions and adjust to difficult situations. A smile can help reduce conflict and ease embarrassment. In many languages, there are sayings that express the social importance of smiling. For example, in English, people say, "Smile and the whole world smiles with you. Cry and you cry alone."

Smiling does more than just help us make and maintain relationships, however. It seems that the amount we smile and the quality of our smiles may have some connection to our quality of life. Two studies show the relationship between smiling and the quality and length of people's lives. One study is the "Yearbook Study." In 2010, LeeAnne Harker and Dacher Keltner, two social scientists from the University of California, Berkeley, compared the lives of women they found in a thirty-year-old year-book<sup>2</sup>. They rated the women's smiles by measuring the amount of muscle movement around the mouth and eyes. Then they asked the women to answer some questions about their lives. The results of their analysis showed that the women with the highest rated smiles in the pictures reported happier lives and happier and longer marriages.

*"Social scientists who study smiles say there's a lot more to smiling than just showing happiness."*

#### MORE FACTS ABOUT SMILES

- Babies smile inside their mothers before they are born.
- Not only do women smile more than men, but women have larger smile muscles than men.
- People who are guilty of some wrongdoing such as cheating on an exam are more likely not to be punished if they smile when they are caught.
- People who smile shortly after the death of a loved one, such as a husband or a wife, adjust more quickly to the loss than those who didn't smile.

Another study is the "Baseball Card Study" from 2010. Ernest Able and Michael Kruger from Wayne State University in Detroit, Michigan, found that the quality of the smile in pictures of baseball players could actually predict how long they would live. Able and Kruger also rated the players' smiles. The rating system had three levels: no smile, partial smile, or full smile. They found that the players with full smiles lived approximately<sup>3</sup> seven years longer than the players pictured



with partial smiles or with no smiles.

Research shows that smiling has many positive effects on our health. This might explain why the people in the studies with bigger smiles had longer lives. Studies show that smiling reduces stress and stress-related hormones<sup>4</sup>. It also lowers blood pressure. Smiling can affect the brain in the same way as exercise. For example, it increases the amount of feel-good hormones such as serotonin and endorphins. Endorphins not only make us feel better, but reduce pain as well. Furthermore, recent brain research shows that just the act of smiling can actually make us happier. In other words, we smile because something happens that makes us happy. But then, our smiles send a message back to the brain that makes us feel even happier.

Smiling is clearly good for us. We can even get the benefits of smiling just by making ourselves smile. One way to do this is to look at a picture of other people smiling. This is because smiling is contagious<sup>5</sup>. It is very difficult to look at others smiling and not smile back. Even thinking about people smiling can make you smile. It is easy to see that smiling is much more than just an expression of happiness. It's a powerful tool for maintaining both emotional and physical health.

---

<sup>1</sup>*thrive*: do well and be successful, healthy and strong

<sup>2</sup>*yearbook*: a book containing photos and other information that celebrates the previous year at a school or college

<sup>3</sup>*approximately*: about the same as

<sup>4</sup>*hormones*: chemicals naturally occurring in your body

<sup>5</sup>*contagious*: able to spread to other people, such as a disease

Source: “The Science of Smiles” by Gordon, D. & Blass, L., from *Focus Reading and Vocabulary 2*, National Geographic Learning/Cengage Learning