Chapter 2

Abstract

The Effects of Input Enhancement and Involvement Load on L2 Readers' Incidental Vocabulary Learning With a Pop-up Dictionary
(インプット強化と関与負荷がポップアップ注釈つきの L2 読解を通した付随的語彙学習に与える影響)

By VOSLAR MATEJ

This study investigated the effect of pop-up dictionaries on vocabulary acquisition and strategies employed by its users. The pop-up dictionary is a newly emerging tool that helps learners translate words in an electronic text. Previous studies have examined the effects of pop-up dictionary compared with other dictionaries and found convincingly positive effects of pop-up dictionaries on vocabulary acquisition without drawbacks to reading comprehension. This study had two primary goals, each examined by an individual experiment. In Experiment 1, reading with a pop-up dictionary was compared to reading a glossed text. The results suggested that reading with glosses is only effective if the glossed words are subjectively relevant to the reader. This result provides evidence for the importance of pop-up dictionaries because they allow learners to select words relevant to them. In Experiment 2, the effects of input enhancement and involvement load on the use of pop-up dictionary were analyzed. The results of Experiment 2 confirmed the limited effect of input enhancement, further suggesting the importance of word relevance perceived by the reader. Simultaneously, a more complex form of pop-up dictionary did not lead to a facilitative cognitive load. Readers' attention to new words was also recorded in Experiment 2. The results suggest that readers can pay

attention to unknown words under the right conditions, even without input enhancement. Prolonged engagement with target words (and their glosses) led to better meaning recognition, but not form recognition or meaning recall. Another important finding of Experiment 2 is the importance of the distinction between intentional and incidental learning. Although the vocabulary test was not announced, a portion of participants attempted to memorize new words. Such choice proved to have implications on how input enhancement and gloss complexity influence vocabulary acquisition. For both experiments, higher frequency words were more likely to be learned. In line with previous research, no effect of pop-up dictionary on reading comprehension was found. The study concludes that pop-up dictionaries are effective for vocabulary learning.

Table of Contents

Table of Tables

Table of Figures

Table of Appendices

Acknowledgment

I owe a debt of gratitude to everyone who has assisted and supported me throughout my study at the Master's program. First and foremost, I cannot express enough gratitude to my chief advisor, Professor Yuji Ushiro, for his guidance, insightful comments, and encouragement during my undergraduate and graduate academic years. I am immensely grateful for his advice and his concern about my health during my studies. I would also like to express deep gratitude to my two deputy advisors. Professor Yuko Hijikata gave me indispensable insight into statistical analysis and theoretical background and provided tremendous support in writing my thesis. Professor Hirosada Iwasaki guided me on the theory of second language acquisition and the formality of thesis writing. I express my sincere gratitude to them and many other professors who supported me during my studies.

I would also like to offer special thanks to my senpai and kohai from Professor Ushiro's seminars. Mr. Kozo Kamimura provided many valuable insights into vocabulary acquisition studies, and Ms. Tomoko Ogiso was very supportive in seminars and other academic responsibilities. Mr. Yamato Sasaki offered guidance in research, and Mr. Takeru Nishi reviewed the draft of my thesis. I want to thank my senpai, who graduated one year before me, for always making the laboratory a brighter place with brilliant humor.

I would like to extend my gratitude to my mates in Professor Ushiro's seminar. Mr. Shigeru Aoki and Mr. Ryuhei Okada offered emotional support when I was under heavy stress due to academic responsibilities. Mr. Ryuya Komuro offered support by engaging in joint research for the AsiaTEFL conference while also

sharing his in-depth theoretical knowledge on many occasions. Ms. Yu-Wei Cheng offered emotional support, helping me pursue my studies of Korean.

I would like to thank everyone who has participated in my experiments, as their contribution was invaluable to the completion of this study. I would like to extend my gratitude to my part-time job employers who were supportive when I was occupied with writing this thesis. Their support and understanding allowed me to finish my thesis on time.

I feel indebted to the Japanese government, specifically the Ministry of Education, Culture, Sports, Science and Technology, for inviting me to study in Japan and offering me a livelihood scholarship, and paying my tuition throughout my stay. Without the government's support, I would not be able to achieve this. I promise to build on the experiences I gained in Japan and work towards preserving good relationships between Czechia and Japan.

Finally, I would like to express my gratitude to my partner and my friends. They made my time in Japan indescribably enjoyable. I am also grateful to my family for always taking good care of me.

Matej VOSLAR

November 8, 2020

in Tsukuba

1. Introduction

1.1. Context of Present Study

Education is quickly getting digitalized. With the spread of Covid-19 across the world, many public schools or private teaching organizations moved their classes online. Even before its impact, the internet's role in education has been growing. In classes, smartphones, computers, or interactive blackboards have an emerging trend in the past few years. In 2016, the Japanese government announced that they would be considering the distribution of tablets in school to help further digitalize education (Ministry of Education, Culture, Sports, Science and Technology, 2016). This move will increase the role of internet resources and other digitalized content in classes, including the English language. It is expected that in the near future, students will learn with a digitalized interactive piece of software and not with a physical book.

The internet also has an impact on the way vocabulary is learned. It is now possible to look up the meaning of a new word fairly easily in numerous online dictionaries and find sentence examples. With internet access, a new form of looking up new words has emerged, which simply would not have been possible in a pen-and-paper environment. That method is achieved by pop-up dictionaries, a remarkable piece of software that allows the reader to look up the meaning of whichever word they come across in digitalized text. These pop-up dictionaries have been popular among second language (L2) learners of Japanese. However, they are gradually making their way into English as a foreign language (EFL) and English as a second language (ESL) methodology, mostly advertised as commercial software

designed to improve English reading. It is expected that they might also be employed as part of public English education in the future.

This study contributes to the limited amount of research that has been conducted on pop-up dictionaries so far. The author of this study has direct experience with Japanese-to-English pop-up dictionary. Their extensive use of pop-up dictionaries led them to believe that its convenience might hinder vocabulary acquisition. This study started with an experiment designed to evaluate whether pop-up dictionaries could be more effective than glosses (Experiment 1). It also revealed intriguing information about participants' behavior, as they often abandoned the use of the pop-up dictionary. In reaction to these findings, a Experiment 2 was conducted to analyze whether participants pay attention to the new words they came across and whether their abandonment of the pop-up dictionary was intentional or due to a lack of attentional capacity.

1.2. Organization of Present Study

The contents of this thesis are organized as follows. First, Section 2 defines incidental vocabulary learning. The overview addresses previous research of its three principal methods: lexical inference, dictionary use, and glosses (Mitarai & Aizawa, 1999). The noticing hypothesis (Schmidt, 1990) and an overview of research on noticing in second language acquisition (SLA) are also presented, as noticing is a central notion for one of the research questions. Section 3 provides a summary of the goals of this study. Section 4 reports the design and findings of Experiment 1, where the effect of single gloss and pop-up dictionary on vocabulary acquisition have been compared. Section 5 informs about Experiment 2, which compared the effects of input enhancement and gloss with the effectiveness of a pop-up dictionary while

analyzing participants' attention toward target words. The findings of both experiments are synthesized in Section 6. Finally, Section 7 provides the theoretical and pedagogical implications of this study. It also lists the aspects which could not have been addressed in this study but should be addressed in future studies.

2. Literature Review

2.1. Vocabulary Knowledge

It is estimated that vocabulary knowledge makes up two thirds of language knowledge needed to understand a text written in the English language (Qian, 2002). It is estimated that learners need to know at least 98% of the words in a text to read it fluently (Hu & Nation, 2000). To achieve this level of coverage, learners should know around 40,000 English lemmas, and to get as close as 95% coverage level, they would need to know around 10,000 lemmas (Grabe, 2009). Cultivating vocabulary knowledge is, therefore, an essential part of the learning process.

There is a distinction between receptive and productive vocabulary knowledge, which means that a learner might recognize a word and/or its meaning in a text; however, they might not know how to use it in a speech or writing (Nation, 2013). Learners supposedly acquire the receptive knowledge before the productive knowledge of a word because only the knowledge of distinctive features of a word is needed to identify it, but a more complete knowledge of the word is needed to produce it. Based on this assumption, a set of tests can determine the learner's level of vocabulary acquisition. These suggested tests are in ascending order of difficulty (Laufer & Goldstein, 2004): recognition, form recognition, meaning recall, form recall (Schmitt, 2010).

For example, Wesche and Paribakht (1996) developed the Vocabulary Knowledge Scale (VKS) as a unified test to measure incremental vocabulary knowledge based on the examinees' self-report and their ability to provide meanings or meaningful example sentences to target words.

Kadota, Noro, Shiki, and Hase (2014) define the development of vocabulary knowledge as a shift of vocabulary knowledge from episodic (or context-dependent) to procedural (context-independent). Some studies (Webb, 2007) have employed a pair of tests where words are first presented without context and then with context to measure vocabulary knowledge from this aspect.

The modes of acquiring new vocabulary can broadly be divided into *intentional* and *incidental* vocabulary learning (Fichtner & Barcroft, 2019; Jiang, 2000). In comparison with native speakers of English, learners who engage with EFL acquire a large portion of their vocabulary knowledge through intentional vocabulary learning. This mode of learning includes intentionally memorizing words acquired from textbooks or flashcards. However, textbooks on their own are considered to be an insufficient source of vocabulary knowledge needed for communication (Horst, 2005). The use of both intentional and incidental vocabulary learning is supported by Nation and Webb (2011). They suggest that learning a vocabulary intentionally and then being exposed to it in various contexts is the most effective way to learn vocabulary. The importance of learning vocabulary in context has been emphasized by many researchers (e.g. Hamada, 2014; Hasegawa, 2012, 2013) and learning from context is considered superior to learning from definitions (McKeown, 1985). Oxford and Scarcella (1994) claimed that words learned outside context are not likely to stay in the learners' long-term memory.

2.2. Incidental Vocabulary Learning

In the incidental mode of vocabulary learning, new words are learned organically as a side product of reading books, watching movies, or other communication forms (Schmitt, 2010). It was estimated that even EFL learners

acquire most of their vocabulary incidentally (Hulstijn, 2003). While an attractive prospect, incidental vocabulary knowledge is not without its disadvantages. The most commonly listed disadvantage is that learners need to encounter the word at least eight times before it can be committed to memory (Waring & Takaki, 2003).

In language acquisition research, incidental vocabulary learning is usually defined as (a) occurring during meaning-oriented activities (Brown, Waring, & Donkaewbua, 2008), (b) limited to cases where learners do not have the intention to learn the vocabulary items (Barcroft, 2004). Researchers mostly attempt to satisfy the latter condition by experimenting without announcing a vocabulary test to the participants before they read. The former condition is usually satisfied by including the target words in a text and instructing the participants to focus on the meaning of the text. However, some studies intentionally shift some of the participants' attention towards vocabulary or other feature by using input enhancement (see Section 2.13), while still referring to the incidental learning mode.

The term *semi-incidental* vocabulary learning was coined for situations where learners are thought to be focusing primarily on the context, although some saliency-increasing features of the text (such as input enhancement or input flooding) are included to increase the focus on vocabulary (Pellicer□Sánchez, Conklin, & Vilkaitė□Lozdienė, 2020).

However, acknowledging the possibility of increased focus toward vocabulary from the learners is not limited to the cases where the vocabulary salience is increased through external factors. Bruton, López, and Mesa (2011) claim that incidental vocabulary learning, defined as learning vocabulary without the intention to learn, is, for multiple reasons, an “impracticable term.” For example, even when a

learner engages with a word in a way such as looking it up in a dictionary or inferring it from context, it is impossible to assess whether they intend to memorize that word. They also cite Hulstijn's (2001) book, which reveals that participants' vocabulary acquisition was not affected by whether or not a post-test was announced if vocabulary saliency was increased through input enhancement. Therefore, Bruton et al. (2011) claim that deciding not to announce a vocabulary test does not guarantee that participants will not intentionally try to memorize some vocabulary items.

However, as is apparent from a meta-study conducted by Uchihara, Webb, and Yanagisawa (2019), the term incidental vocabulary learning is still popular among researchers to describe a mode of vocabulary acquisition in a meaning-oriented task without the announcement of a vocabulary post-test. It is acknowledged that incidental vocabulary acquisition can be considered purely non-intentional neither in the studies reported in this literature review nor in the scope of this study.

The respective ways vocabulary can be learned incidentally are described by Mitarai and Aizawa (1999) as (a) inferring (or guessing) the words' meaning, (b) looking up the word in a dictionary, or (c) using glosses.

2.3. Involvement Load Hypothesis

For the reasons stated above, incidental vocabulary learning is less time-effective than intentional vocabulary learning. Researchers, therefore, needed a way to assess the effectiveness of different meaning-oriented activities for vocabulary acquisition.

The involvement load hypothesis (ILH; Laufer & Hulstijn, 2001) was devised based on the hypothesis that deeper levels of processing lead to a higher probability of such information being learned (Craik & Lockhart, 1972). Similarly, ILH states

that the more a learner engages with a new word, the greater the chance they will successfully learn that word. The involvement load is described in three components: need (whether the learner needs to know the meaning of the word), search (looking up the meaning), and evaluation (checking if the meaning fits the context). Therefore, the involvement load is widely used in research related to dictionaries and glosses (the details are described in Sections 2.6 and 2.7). Each component can be awarded zero to two points, as described in Table 1. Therefore, ILH is constructed as a scale of 0 to 6, where tasks with higher involvement load are predicted to lead to higher vocabulary gains.

Table 1

Scoring of Component in Involvement Load Hypothesis, Adapted from Fatalaki (2014)

Component	Score
Need	
The learner does not feel the need to learn the word.	0
The learner is required to learn the word.	1
The learner decides to learn the word.	2
Search	
They do not need to learn the meanings or forms of the word.	0
The meaning of the word is found.	1
The form of the word is found.	2
Evaluation	
The word is not compared with other words.	0
The word is compared with other words in a provided context.	1
The word is compared with other words in a self-provided context.	2
Total Score	0~6

Tokuda (2006) expanded on this model by adding the *confirmation* component and claiming that incidental vocabulary acquisition can be further

promoted by providing feedback. A study by Frishkoff, Collins-Thompson, Hodges, and Crossley (2016) proved the positive effect of feedback on word learning and also reported gains in accuracy and confidence when feedback was provided. There is a possibility that confirmation, whether through external corrective feedback or the learner's confidence in their evaluation, is an essential part of the evaluation, and therefore, without confirmation, the involvement load of evaluation is not enough to promote vocabulary acquisition (Komuro & Voslar, in writing).

The involvement load hypothesis awards task points for each of the three components (need, search, evaluation), and the overall load is calculated as the sum of those points.

2.4. Technique Feature Analysis

Some researchers challenged ILH and found that each of the three components contributes to vocabulary acquisition differently (Laufer, 2003; as cited in Hu & Nassaji (2016). To address this issue, Nation and Webb (2011) devised a more detailed scale of 0 to 18 points, which can be used to predict the effect of a task on vocabulary acquisition. They dubbed this scale the *technique feature analysis*.

This method of analysis views each task from more aspects than ILH. These aspects are motivation, noticing, negotiation, retrieval, creative use, and retention. These categories contain a total of 18 sub-aspects, as shown in Table 2, and the task with a higher number of aspects present is predicted to be more beneficial to vocabulary acquisition. The higher validity of the technique feature analysis, compared to ILH, has been demonstrated by Hu and Nassaji (2016).

Table 2

Technique Feature Analysis, Table Adapted From Nation & Webb (2011)

Criteria	Score
Motivation	
Is there a clear vocabulary learning goal?	0~1
Does the activity motivate learning?	0~1
Do the learners select the words?	0~1
Noticing	
Does the activity focus attention on target words?	0~1
Does the activity raise awareness of new vocabulary learning?	0~1
Does the activity involve negotiation?	0~1
Retrieval	
Does the activity involve the retrieval of the word?	0~1
Is it a productive retrieval?	0~1
Is it a recall?	0~1
Are there multiple retrievals of each word?	0~1
Is there spacing between retrievals?	0~1
Generation	
Does the activity involve generative use?	0~1
Is it productive?	0~1
Is there a marked change that involves the use of other words?	0~1
Retention	
Does the activity ensure successful linking of form and meaning?	0~1
Does the activity involve instantiation?	0~1
Does the activity involve imaging?	0~1
Does the activity avoid interference?	0~1
Total Score	0~18

2.5. Lexical Inference

Learners can acquire new vocabulary by guessing the meaning of an unknown word (Huckin & Coady, 1999). According to Huckin and Bloch (1993),

learners make assumptions about an unknown word's meaning and then try to test their assumption with their knowledge or the context (Nassaji, 2003). Lexical inference is supported by the breadth of vocabulary knowledge (i.e., how many words the learner knows, Mochizuki & Aizawa, 2000), explicit knowledge of morphemes (Zhang & Koda, 2012), or the learners' English proficiency (Nakagawa, 2006).

It has been pointed out that inference as a method is prone to failure, and learners might ignore words which they are not able to infer the meaning of (Fraser, 1999; Hulstijn, 1993). The ability of learners to infer the meaning of a new word partially depends on the quality of the context as described by the degree of semantic relationships (Hamada, 2015). The model of acquiring new vocabulary from lexical inference has been described by Jiang (2000), based on Levelt's (1993) model of vocabulary knowledge. Bengeleil and Paribakht (2004) have identified that readers can use clues from inside or outside of the same sentence as the inferred word to help guess its meaning. However, lower-proficiency readers might need to draw inferences from multiple sources. Nevertheless, Bensoussan and Laufer (1984) found that context helped learners infer meanings of new words in only 13% of the cases.

Mondria (2003) found that there is a close effect on vocabulary acquisition between providing meanings and asking learners to infer their meanings, in the intentional learning condition. However, lexical guessing was more time-consuming and, therefore, less effective.

Huckin, Haynes and Coady (1993) pointed out that learners sometimes fail at lexical inference because they assign an incorrect meaning to the new word. Sometimes learners hold onto the new word for a few following sentences until they

can figure out the meaning of the word, and sometimes they abandon it, unable to guess its meaning.

2.6. Glosses

Glosses are an effective way to support incidental vocabulary learning (Jacobs, Dufon, & Hong, 1994; Yanagisawa, Webb, & Uchihara, 2020; Watanabe, 1997). The term *gloss* refers to a word, phrase, or hint to the meaning of an English word in a text.

There is a large variety of gloss types, the taxonomy, described in Roby (1999). The distinction can be determined by the gloss's position (whether it is listed right after the glossed word or found in a glossary outside of the text). A gloss in a text intended for the speakers of a particular language may show the meaning of the glossed word in their native language. However, a gloss could also be a synonym in the target language, a definition (such as those found in English to English dictionaries), a picture, a video, or the combination of the above. Some glosses also present multiple meanings (either possible translations of a multisemous word or unrelated distractors) from which the learner has to select the correct translation. Reading in digital contexts allows for *pop-up dictionaries* (sometimes called *CALL gloss* or *hypertext gloss*), making it possible for the readers to see the gloss for any word of their choosing, thus removing the need to consult an external dictionary (pop-up dictionaries are further discussed in Section 2.8).

A considerable amount of attention has been given to glosses in incidental vocabulary acquisition research. Multiple studies compared the effects of gloss on learning new words to other methods of incidental learning. Previous studies have demonstrated that using glosses is more advantageous to vocabulary learning than

guessing the words' meaning (Yanagisawa et al., 2020) or using a dictionary (Hulstijn, Hollander, & Greidanus, 1996).

When learners are presented with a single gloss (SG) — a gloss which only displays the correct meaning — there is no need for *search* or *evaluation* in the sense of ILH. On the other side, the learner with only a dictionary will have to look up the word's meaning and often contrast multiple translations of a polysemous word with the context, creating an opportunity for search and evaluation (Hulstijn & Laufer, 2001).

Multiple-choice gloss (MCG), a gloss with one correct answer and one or more distractors/incorrect answers, is often proposed to increase the learner's involvement load (Beal, 2007). However, MCG did not always yield higher vocabulary gains in previous research. For example, Watanabe (1997) found no difference between SG and MCG conditions. However, it has been argued that MCG was possibly not effective because the distractors were too difficult for the learners to dismiss. This conclusion is also supported by Nagayama and Mori (2003), who found that high-proficiency learners are more likely to benefit from MCG, which was also supported by Kasahara (2004), who found that learners need to know about 5,000 lemmas to benefit from MCG.

Researchers have raised concerns about the *fossilization* of lexical knowledge when learners are presented with a single first language (L1) gloss (Jiang, 2000). If learners acquire a single meaning of a polysemous word (such as *bark*, the sound made by a dog), it might be difficult for them to learn a new meaning (such as *bark,* the outer covering of a tree) as pointed out by Ushiro, Hasegawa, and Nahatame

(2013). In this regard, it might be more beneficial for learners to look up new words in a dictionary.

Glosses also have the advantage of focusing the learner's attention on the glossed words. However, as only words selected by the author of the learning material will be glossed, learners might come across a non-glossed word with an unknown meaning.

2.7. The Use of Dictionary in Incidental Vocabulary Learning

Based on ILH, learning new words with a dictionary should be superior to single glosses. However, it has been found that learners will often abandon the use of the dictionary. Learner's use of dictionaries is usually limited to words closely related to the topic or words used multiple times in the text (Hulstijn et al., 1996). Peters, Hulstijn, and Sercu (2009) discovered that announcing a vocabulary test will promote dictionary use. However, whether a word is relevant to comprehension has a more significant effect on dictionary use. Knight (1994) found that dictionary contributes to vocabulary acquisition and reading comprehension more than inferencing words from context and that frequently using dictionaries does not hinder comprehension.

2.8. Pop-up Dictionaries

One method, often regarded as *extensive reading,* has been proposed to increase the vocabulary learners can acquire incidentally. Extensive reading is defined as a solitary learning activity that learners should seek with pleasure. They should be able to read about a topic which interests them, and it is advised to try reading as fluently as possible, without searching for words' meanings in a

dictionary. As has been fore-mentioned, the learner has to know at least 95-98% of words in the text to read fluently (Grabe, 2009). Therefore, reading materials must be carefully curated to the learner, according to their taste and English proficiency, to achieve this. Low vocabulary knowledge might discourage learners from reading a text they were initially interested in. The hurdle of insufficient vocabulary knowledge can be partially overcome by adding glosses to the text. However, the author's or teacher's static glosses might not reflect which words the learner knows and which they do not.

With the increased use of technology and internet communication for language learning, a new type of gloss has emerged. The *pop-up dictionary* (sometimes called *hypertext gloss*) is a piece of software, which will display a dictionary entry for a foreign word when the user clicks on it or hovers their mouse over it. Some notable examples include Rikai-kun (a browser extension for English speaking learners of Japanese) or the built-in pop-up dictionary on MacOS systems, which can be activated by selecting a word and pressing the Command, Control, and D keys. Visuals examples of these pop-up dictionaries can be seen in Figure 1 and 2.

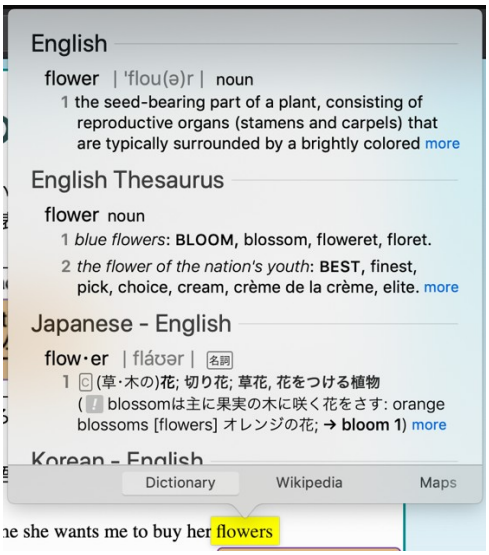


Figure 1. Pop-up dictionary embedded in MacOS.

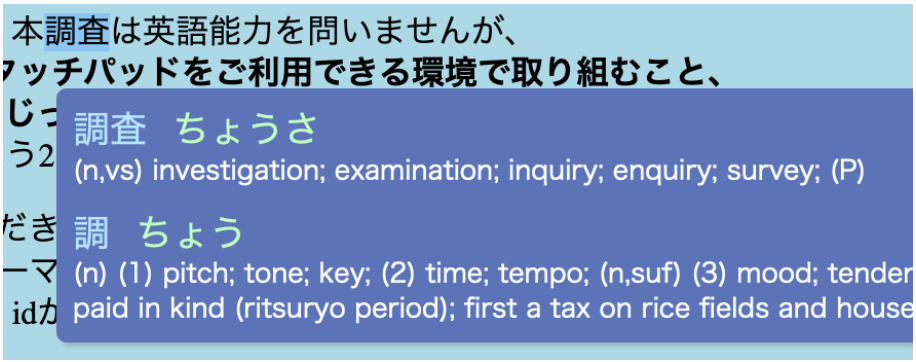


Figure 2. Rikai-kun, a widely used pop-up dictionary for L2 learners of Japanese.

The pop-up dictionary seemingly combines the strengths of both reading with glosses and with a dictionary. Learners can use it to look up whichever word they like, and at the same time, looking up a word takes just one click. Therefore, it does not distract the learner from reading, and the learner should be more likely to use it. Since pop-up dictionaries can be used to look up any word in a text, it could encourage both intensive and extensive reading.

As the gloss is only displayed after the reader clicks or hovers over a word, it may lead to *retrieval*, also called *testing effect* (Nation, 2013). Retrieval means that a

form-meaning connection which is in the reader's memory – although it has not been strengthened enough yet – may become stronger through repeated attempts to recall the word's meaning before looking it up in a dictionary (Webb & Nation, 2017)

Few studies have examined the effectiveness of pop-up dictionaries. Movahedi and Shourkaee (2018) compared the use of pop-up dictionaries and electronic dictionaries and found no difference in vocabulary acquisition, although the pop-up dictionary had a better effect on the learners' attitude towards learning. Marefat, Rezaee, and Naseriah (2016) conducted an experiment where a gloss would only appear after a word was clicked and found that displaying the gloss next to the target word (as opposed to the margins) is more beneficial for vocabulary learning. However, this cannot be considered a study about pop-up dictionaries, as participants could only look up the meaning of visually enhanced target words. Alessi and Dwyer (2008) found a positive effect of the pop-up dictionary (which they call hypertext gloss) on reading comprehension.

There have been several studies where the translation was provided after typing the word in a box on the same screen (*type-in dictionary*) and compared against pop-up dictionary (Alharbi, 2016; Liu & Lin, 2011; Mekheimer, 2018). These studies often show the same results for type-in and pop-up dictionaries or a slight advantage of the latter. In the research conducted by Liu and Lin (2011), the pop-up dictionary produced twice as many lookups, but in Mekheimer (2018), the type-in dictionary was the mostly used one. A type-in dictionary is also a considerably non-intrusive method of providing translations to whichever word the reader pleases. However, it involves engagement with the word's form and produces separation from the text being read for a longer period than a pop-up dictionary. As

Liu and Lin conclude, the load produced by holding the word's form in memory might not be beneficial to vocabulary acquisition, providing more evidence for the advantage of pop-up dictionaries. Pop-up dictionaries also proved to be the most time-efficient and most favored by the participants (Mekheimer, 2008).

An L2 Japanese study conducted by Tabata-Sandom (2016) found that high proficiency students used pop-up dictionaries more efficiently and in combination with top-down processing, while low-proficiency students relied on them overtly.

It is expected that Japanese schools will adopt electronic textbooks in the near future. Each student will view the textbook on a tablet device, and many of the textbook elements will be interactive. The author assumes that pop-up dictionaries will be one of the interactive elements in this learning process. There are currently advertisements for reading practice software in the private English teaching sector, which uses the pop-up dictionary.

So far, the research has shown the positive effects of pop-up dictionaries on vocabulary acquisition and learners' attitudes. However, the amount of research on pop-up dictionaries is still scarce. Furthermore, the above-reported findings contradict the author's direct experience with the Rikai-browser extension, which learners used in learning Japanese. Due to its high accessibility, the author repeatedly used it to look up the meaning of words they had looked up previously. Since it was very easy to look up a word, the author did not feel the need to remember the meaning of new words. This realization led to the willful removal of this extension from the author's browser software. This experience became the incentive for this study.

2.9. Noticing

Closely related to the topic of glosses is the noticing hypothesis proposed by Schmidt (1990). According to Schmidt, when a learner pays attention to a specific linguistic element, that element enters their awareness. At a certain level of awareness, this element is considered 'noticed' by the learner. Schmidt claimed that noticing is a necessary condition for learning to occur and denied the possibility of implicit learning (learning without awareness).

This view has been challenged by Tomlin and Villa (2004), who claimed that detection is enough for learning to take place. While both views agree that detection (or focal attention) is a necessary condition, Tomlin and Villa argued for learning without conscious processing. In reaction, Schmidt adjusted the original claim that awareness was necessary for learning and insisted that a positive correlation must exist between noticing and learning, mostly due to the difficulty of proving the complete absence of learners' consciousness (Schmidt, 1994).

Additionally, multiple studies using artificial English grammar (a grammatical distinction between animate and inanimate objects) have proven that learners can learn the grammatical rule from the input and apply it above chance, including participants who were not able to explain what the rule was (Leung & Williams, 2012). Smith and Yu (2008) have shown that infants can learn form-meaning connections implicitly based on cross-situational statistics. These results suggest that learning can take place without the learner's awareness.

Nevertheless, the link between noticing and learning is supported by empirical research (Godfroid & Schmidtke, 2013), who used offline and online measures to confirm the connection between attention, awareness, and learning.

Similarly, Izumi and Bigelow (2000) suggest that noticing-the-gap (noticing one's language ability deficit) is necessary for learning new linguistic forms. In an eye-tracking study by Brusnighan and Folk (2012), the words that were viewed longer by participants were more likely to be memorized, suggesting that attention is essential for vocabulary acquisition.

2.10. Attention and Awareness

Words *attention* and *awareness* are used to describe noticing, although the exact definition may vary from researcher to researcher. Schmidt (1995) states that awareness is "a conscious registration of the occurrence of some event" and defines noticing as a low-level form of awareness. In Schmidt's view, an increased level of attention and awareness are basically the same, which means that when learners focus on a particular element, we can assume the element goes into the learner's consciousness. In this study, the definitions used in Godfroid and Schmidtke (2013) will be employed. In their study, attention is measured as the time participants spent focusing on each word in an eye-tracking reading task. Awareness is then measured as the ability to recall the word read. Therefore, attention is measured online, while awareness is measured offline (in a post-test). Consequently, it suffers from participants' inability to remember everything they paid attention to in the reading task.

2.11. The Role of Noticing in Incidental Vocabulary Learning

A possible reason for the lack of effectiveness of incidental vocabulary learning compared to intentional vocabulary learning is the lack of attention to each linguistic element. It can be assumed that glosses support learners' attention and,

therefore, facilitate vocabulary learning because, as Nation (2013) points out, reading while focusing on vocabulary leads to more significant vocabulary gains, and glosses make target words stand out in the texts.

Another important factor is the frequency (further discussed in Section 2.14) a certain linguistic element appears in the input (Rott, 2005). A word appearing multiple times is more likely to be noticed by the learner, and words with higher frequency in the text are also likely to be closely related to the topic of the text; therefore, the learner might give them more attention.

2.12. Noticing in Learners with ADHD

Attention deficit hyperactivity disorder (ADHD) is the most common disorder affecting learning (American Academy of Pediatrics, 2000). Since this disorder manifests itself without hyperactivity in many cases, it is estimated that not everyone with this disorder will be diagnosed. As boys are four times more likely to be diagnosed than girls, it is expected that selection bias plays a role in who is diagnosed with ADHD (Gershon, 2002). Previous studies have confirmed that students diagnosed with ADHD have trouble learning vocabulary (Sabet, Farhoumand, Mahdavi, & Naseh, 2015), which is not surprising given the importance of attention in vocabulary learning as suggested by previous research (Nation, 2013; Schmidt, 1994; Tomlin & Villa, 1994).

2.13. Input Enhancement

Input enhancement (IE; Sharwood Smith, 1993; as cited in Izumi & Bigelow, 2000) has been proposed to increase the learner's attention to new linguistic elements. Input enhancement can be understood as using visual aids (such as bold or

colored font) to make certain elements of a text more salient. Sometimes the term *input enhancement* is used to refer to other techniques of increasing the input's salience.

One such technique is *input flooding*, which involves artificially increasing the frequency of target elements in a text. Previous studies have shown a positive effect of input flooding on grammar (Arani & Yazdanimoghaddam, 2016) or vocabulary (Namaziandost, Rezvani, Polemikou, & Popescu, 2020). The term *elaboration* is used when the materials are modified to provide a more informative context for target elements. This technique has been confirmed to promote the retention of new word's meanings (Kim, 2003). For this study, the term *input enhancement* will only be used for visual aids, which are sometimes also described as a *typographical enhancement*.

Winke (2013) has shown that input enhancement leads to more attention to new grammatical forms but found no effect on grammar learning or reading comprehension. Loewen and Inceoglu (2013) found that input enhancement promoted awareness of new forms but did not lead to higher acquisition rates. Conversely, Park, Choi, and Lee (2012) found that input enhancement promoted grammar learning but hindered text comprehension. Gass, Svetics, and Lemeli (2003) have also confirmed the effect of input enhancement on acquiring grammar. In Cho (2010), input enhancement encouraged receptive but not productive grammar knowledge.

Similarly, in incidental vocabulary acquisition studies, input enhancement did not reliably help participants learn new words. Sánchez Gutiérrez, Serrano, and Garcia (2019) found that IE positively affected form recognition but not meaning

recognition. LaBrozzi (2016) found that input enhancement can lead to higher vocabulary retention, but different types of enhancement affect vocabulary learning differently. Petchko (2011) found no effect of input enhancement on form recognition, meaning recognition, or meaning recall when reading texts at 98% known-word coverage. Meganathan, Thai, Paramasivam, and Jalaluddin (2019) found positive effects of IE on vocabulary acquisition during extensive reading activities. Peters (2012) found that participants were more likely to take notes of words and phrases with input enhancement. Corbetta and Schulman (2002) suggest that input enhancement has a bottom-up effect on attention, which is only effective when combined with attention driven by the learner's motivation. Such an explanation could mean that input enhancement might be effective in intentional vocabulary learning tasks, but not when participants read a text for comprehension.

In a study employing a pop-up dictionary, de Ridder (2002) found that input enhancement increased the number of words looked up by participants. However, the increased use of glosses did not influence comprehension or vocabulary acquisition. Instead, vocabulary learning was affected by the type of task (e.g., vocabulary-oriented or contents-oriented).

2.14. The Effect of Frequency on Vocabulary Acquisition

By the term frequency, we can understand two different properties of a lexical phrase. One is the global frequency of a word, which narrates how often this word was used in actual conversations and/or media. The global frequency of a word is measured by how commonly the phrase appears in a certain corpus, and the standard unit is the number of occurrences per million words (pmw). Researchers have identified what is called the *word frequency effect*. This effect states that more

frequent words will be processed more quickly than less frequent words (see e.g., Jescheniak & Levelt, 1994). Popov & Reder (2020) have discovered that Dutch learners of English were sensitive to the frequency effect for words they have previously known.

However, in this study, frequency will refer to the local occurrence of a word, meaning how many times it was used through the reading material. Items with increased frequency are more likely to be memorized (Rott, 2007). Previous research has shown a positive effect of higher word exposure on spelling and meaning recognition (Sánchez Gutiérrez et al., 2019). As was previously described, artificially increasing the local frequency of an element is called input flooding, and previous research has shown that it is effective for vocabulary acquisition (Arani & Yazdanimoghaddam, 2016; Namaziandost et al., 2020). Input flooding can be effectively used in listening or reading, although the effect is shown to be more significant when reading (Rashtchi & Yousefi, 2017).

Although meeting the same word several times increases its salience (Zhu, 2015), it can also provide new contexts for the word. The instance-based theoretical framework (Bolger, Balass, Landen, & Perfetti, 2008) is based on the proposition that learning new words is incremental (Fukkink, Blok, & de Glopper, 2001) and the knowledge of a word can be strengthened by presenting it in different contexts (van Daalen-Kapteijns, Elshout-Mohr, & de Glopper, 2001). Based on this framework, Bolger et al. conducted two experiments. Presenting a word in four different contexts significantly improved its acquisition (measured by meaning recall) compared with presented the same word four times in the same context.

2.15. The Relation Between Working Memory and Noticing

The working memory (WM) is a short-term memory model that incorporates both storing and processing of information (Baddeley & Hitch, 1974). This model is integral to reading because reading involves intaking new information, processing that information, and integrating it with previously read information or previous knowledge of the reader (Cowan, 2010). The capacity of working memory is limited, and the limit has been defined by Miller (1956) as 7±2 items, also known as the *magic number*. The integrated nature of storing and processing in this model also means a possible trade-off relationship between these two (Harrington & Sawyer, 1992).

The WM model consists of smaller elements. Important to the process of reading are the audio loop, episodic buffer, and executive control. The audio loop is considered integral to reading because the read information is repeated inside the audio loop to make it accessible for processing. It has been estimated that the maximum length of audio information that can be rehearsed in the loop is around 2 seconds, and theorized that the maximum length of an audio loop is the defining feature of WM capacity, as all information has to be stored there to be processed, and people who can pronounce more in 2 seconds are more likely to remember more (Timarova, 2008).

Previous research has shown an effect of WM capacity on reading, as individuals with higher WM capacity score higher on reading comprehension tests (Harrington & Sawyer, 1992). Additionally, research has found the influence of a reader's ability to update information (Palladino, Cornoldi, De Beni, & Pazzaglia, 2001) or stay focused on reading without giving in to intrusive thoughts (McVay &

Kane, 2012). These influences may depend on individual differences in the executive function of WM. Executive control is considered responsible for the shifting of attention and is thought to manage which information is preserved (or renewed) in the working memory and which information is to be deleted (or faded) from memory. The switching between different tasks is also considered the responsibility of executive control. Numerous tests, such as the Wisconsin Card Sorting Test (Grant & Berg, 1948), have been designed to measure individual differences in task switching.

Based on Schmidt's (1995) noticing hypothesis, Robinson (2003) proposed a model of noticing, where the learner's ability to notice linguistic elements depends on the amount of attention given by the learner and the rehearsal in the learner's working memory. According to Robinson, if an element enters short term memory but is not rehearsed in working memory, it can only be subject to *detection*, but not *noticing*. Detection then may lead to automatic processing of input, but never to learning processes, which are only available to elements that have been rehearsed in working memory and, therefore, *noticed*.

Because attention is closely related to working memory (which is a limited resource), there is an ongoing discussion on whether the attention at a given time is also limited. Some studies analyzed the effect of working memory capacity on the ability to notice and uptake information from recasts, but with varying results (Chen, 2013; Sagarra & Abbuhl, 2013). In a study conducted by Bergleithner (2007), a connection was found between participant's working memory and the number of linguistic forms they noticed. Rott and Williams (2003) found that grammatical

processing in output might leave little room for acquiring new words from the input, but the amount of attention to new words can be increased by using glosses.

Based on Schmidt's (2012) assumption that the ability to notice linguistic elements varies between learners, Simard and Foucambert (2013) measured the participants' attentional capacity using the Trail Making Test and found that participants with lower attentional capacity devoted more attention to elements with visual input enhancement when reading in L2. Chun and Payne (2004) found that participants with lower phonological memory capacity looked up more glossed words.

Gu and Johnson (1996) identified selective attention (top-down strategy where learners choose which words to focus on) to be the key to successful vocabulary learning. Prichard (2008) also identified a pattern where advanced learners are more selective when considering whether to look up a word. Robinson (2013) suggests that these results contrast Neumann's research (Neumann, 1996; as cited in Robinson, 2003), who considered learners choosing target forms to be the result of limited attentional capacity.

Although the direct effect of working memory on noticing is still disputable, it is believed that learners are not able to notice all that is presented to them, as illustrated by the following citations:

> *"Learners are not free to notice anything and everything they wish to notice."*
>
> Izumi (2013)

> *"It is highly possible that during reading, the readers fail to notice unknown words and vocabulary learning will not occur."*
>
> Azari (2012)

These views are also in accordance with Jiang, Costello, Fang, Huang, and He (2006), that found that the shifts of attention may occur without awareness.

The answer to this problem is closely related to the topic of pop-up dictionaries. Perhaps learners are not always purposefully abandoning external dictionaries (such as paper or website dictionaries), rather they fail to notice an unknown word. In such cases, the pop-up dictionary would presumably have no advantage over an external dictionary.

2.16. Measuring Noticing

For measuring noticing, both offline and online methods have been employed in previous research. Offline measurement might include asking participants questions about the target elements or other ways of eliciting a reaction after the task (e.g., reading a text.). This method is based on the belief that "reportability is the key property of awareness" (Baars & Franklin, 2007). However, because this type of measurement might be influenced by memory loss, online measurement methods are considered to be more suitable (S. Song, 2007, p.5).

One of the commonly employed online measures is the think-aloud protocol. In this method, participants verbalize what they are noticing or sometimes answer questions from the experimenter. Because online measurement might influence how participants interact with the experiment material (Chaudron, 1985), eye-tracking has been proposed as an accurate way of measuring noticing with the least intrusion possible (van der Schoot, Vasbinder, Tako, & van Lieshout, 2008).

Many variables can be acquired from eye-tracking. Suppose we define areas of interest (AOI), such as individual words or sentences. In that case, we can then measure the length of individual fixations inside that area or the order of movements between areas of interest. Rayner (1998, 2009) provides a more detailed list of important variables:

- *fixation duration*: the time for which the eye stops in AOI before it moves either outside of that area or to another point inside that area,
- *gaze duration*: the sum of all fixation times before the gaze moves to another AOI,
- *regressions*: the number of times participant's gaze returns to an AOI,
- *total reading time*: the sum of all gaze durations on a single AOI

According to Rayner, the first fixation duration is significant in analyzing experiments with single words as AOI. If the AOI encompasses more than a word (a phrase or a sentence), gaze durations are often dubbed *first pass reading time, (second pass reading time*,..), and in research with sentences or multi-word phrases, the second pass reading time plays a crucial role.

2.17. Limitations of Previous Research

The pop-up dictionary is still an emerging area of SLA research. So far, most studies have examined the benefits of using a pop-up dictionary compared to other types of dictionaries or reading without a dictionary (Liu & Lin, 2011; Mekheimer, 2018). However, there is still room for examining the differences between pop-up dictionaries and glossed texts, as the latter takes away the choice of the words the reader looks up. The question is if choosing which words to focus on could be more effective than the instructor's selection of such words in a material.

Another limitation would be the difference in the amount of information shown to the reader. A book dictionary (or an online dictionary) will most likely show multiple meanings, sometimes with syntactic information, collocations, or use examples. However, pop-up dictionaries or type-in dictionaries in previous research have mostly shown short translations according to the context. Therefore, there is a

possibility for research examining the effect of pop-up dictionaries based on the amount of information they show.

As mentioned previously, the author had mixed experiences with the pop-up dictionary. It was an essential companion when the author was trying to read online articles in Japanese. However, the author found themself looking up the same word over and over again. The author assumed this was an effect of the pop-up dictionary being too easy to access. This view was also shared by a few of the author's acquaintances. Previous studies have only reported participants' favorable opinions of the pop-up dictionary; perhaps it would be interesting to see what learners believe about the influence of pop-up dictionaries on learning new words.

Lastly, researchers sometimes assume the limits of learners' attentional capacity without sound empirical evidence or based solely on offline measures, proving what the participants paid attention to only to a limited degree. As will be discussed in Section 5.1, attention is critical to pop-up dictionaries because participants must first pay attention to a word before looking it up. Therefore, participants' attention must be examined using an online method to better understand how it influences the readers' use of the pop-up dictionary.

3. The Present Study

This study is concerned with a pop-up dictionary as a supporting tool for incidental vocabulary learning. It mainly focuses on the learner's strategy when reading a text with a pop-up dictionary and its effect on vocabulary acquisition.

This study has been conducted in two phases: Experiment 1 and Experiment 2. Experiment 1 was intended to assess the effectiveness of a pop-up dictionary in comparison with a single gloss. Although this study offered a valuable insight into the learner's pop-up dictionary strategy, the tested sample was limited in number, and the results were possibly affected by the experiment's flawed design.

Experiment 2 was conducted on a larger and more homogenous sample. It addressed some of the flaws in the design of Experiment 1, and rather than comparing reading with pop-up dictionaries to reading with other supporting tools (such as a single gloss or an external dictionary), it was intended to investigate the factors that affect the use and effectivity of learning new words with pop-up dictionaries. Experiment 1 failed to address the effect of input enhancement on the use of pop-up dictionaries. Experiment 2 was designed to provide data on how input enhancement affects participants' attention to the enhanced words (target words) and drew conclusions based on the noticing hypothesis. The author also acknowledges that the pop-up dictionary in Experiment 1 did not reflect the reality of the majority of pop-up dictionary software since it only displayed a single meaning of each word a meaning preset to fit the context of the word. Software programs like Rikai-kun will display several possible meanings for the selected word, meaning learners sometimes have to choose the meaning which fits the context the most. In Experiment 2, the two conditions (single context-fitting meaning vs. multiple

meanings pulled from a dictionary) were also compared. The results contrasted with the involvement load hypothesis and discussed in this study.

4. Experiment 1

4.1. Objectives

This experiment was designed to compare the effect of two gloss types. The two types were single gloss (G) vs. pop-up dictionary (PD) on incidental vocabulary acquisition. It also analyzed the effect of the frequency of a word in the text. Some target words were manufactured only to appear once (F1), while others appeared three times (F3). It was intended to answer the following three research questions:

RQ1.1 Is the number of successfully memorized target words affected by the number of non-target words looked up by participants in the P group?

RQ1.2 Does the pop-up dictionary affect reading comprehension?

RQ1.3 Does the number of times a target word appears in the text affect vocabulary acquisition?

4.2. Method

4.2.1. Participants

Data were collected from 24 participants who speak Japanese as their native language. However, after data screening, data from only 11 participants were used in the analysis. The reason for this was mainly that only 12 participants completed the task. One participant was removed because they just skimmed through the experiment without serious engagement. The experiment was conducted in two parts with at least a week (but no more than two weeks) between both parts. However, only five of the selected 11 participants completed the second part, and therefore the data from the delayed test was discarded.

Since it was difficult for the author to gather native Japanese EFL learners at the university, participants were acquired through social media, which caused a

sizable disparity between the participants' ages (from 15 to 31 years) or their occupations. All participants who completed both parts of the experiment were sent an 800-yen-worth Amazon gift card.

4.2.2. Materials

The entirety of the experiment was conducted using an online software developed by the author of this study. In this study, Early People in the Central American Land Bridge (Appendix A), a text intended for young native English speakers, was adapted and used. The length was 821 words and the FKGL reading ease level was 7.1.

From the text, 16 words were chosen as target words. These were compared with the JACET8000 Level Marker word list using the Word Level Checker (http://someya-net.com/wlc/index_J.html), and words in levels 5000 and lower were replaced with less common synonyms. For the full list, see Appendix C.

In measuring participants' comprehension of the text, eight questions were prepared with four answer options. Questions and answers were all presented in English. The list of comprehension questions can be found in Appendix B.

In measuring vocabulary acquisition, two meaning recall tests were conducted: in the first test, the target words were presented without context; in the second, they were displayed in the context in which they appeared in the text. For each word, participants were also asked if they had previous knowledge of the word.

Each participant was assigned randomly to either gloss (G) or pop-up dictionary (PD) groups. For this study, the functionality was almost the same for both groups — the meaning of a word was displayed after clicking on it. Both groups were also asked not to use any external dictionary to look up the meaning of

unknown words. In this study, the phrases *click on a word* and *look up the meaning of a word* will be used interchangeably. The two differences between the groups were that the G group could only look up the meanings of target words, and these target words were enhanced (underline + blue font) for this group. The PD group could look up any word they wanted, but there was no input enhancement present. A translation for each word in the text was prepared, according to the word's context. Figure 3 provides an example of the pop-up dictionary in use.

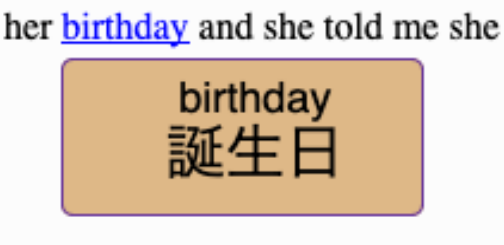


Figure 3. The pop-up dictionary used in Experiment 1.

At the end of the experiment, participants were asked to fill out a survey on questions relating to their thoughts about the gloss. The participants also indicated their previous knowledge of the text.

4.2.3. Procedure

The entirety of the experiment was conducted online. Potential participants were asked to access the experiment's link, register with their email address, and fill out a few questions regarding their age and English proficiency. They were randomly assigned to one of two groups (G for gloss / PD for pop-up dictionary), and a brief explanation of the experiment was shown. This page contained a sample sentence on which participants could experience how the gloss works. After proceeding to the next page, the material text was shown in its entirety, and participants read it at their own pace. The words which they clicked and the time they took to complete reading

the text was recorded. On the next page, eight comprehension questions were displayed. On the following pages were the non-context meaning recall test and the context meaning test. Participants chose which words they had previously known during the non-context recall test. On the last page was a questionnaire asking participants about whether they were familiar with the topic of the text material and whether they thought the gloss was beneficial to comprehension or vocabulary learning.

4.3. Scoring and Analysis

The data from this experiment was analyzed as a set of data points, where one data point measured the interaction of a single participant with a single target word. The data also contained information about whether the participant clicked the word. The answers to vocabulary tests (without and with context) were each recorded separately and evaluated with 2 points for an entirely correct answer and 1 point for a partly correct answer. Data points where the word was known to the participants (and the participant could correctly answer its meaning) were removed from the analysis.

Data for each participant were also analyzed to understand the effect that the gloss type had on their understanding of the text, reading speed, or the strategy with which they used the gloss.

ANOVA regression analysis was performed for all effects, and p value was calculated together with Cohen's *d* effect size.

4.4. Results

4.4.1. Post-test Analysis

Two separate ANOVAs were performed for the non-context and the context vocabulary test. The variables of group (G × PD), previous knowledge of the text (yes × no), frequency (F1 × F3), and whether a word was clicked or not was used. The significance of each factor is shown in Tables 3 and 4.

Table 3

ANOVA Factors of Non-Context Test Scores

Factor	*Df*	*Sum Sq*	*Mean Sq*	*F*	*p*		η^2
Group	1	.07	.067	0.100	.752		.039
Looked Up	1	17.83	17.828	26.533	.000	*	.164
Topic Familiarity	1	0.84	0.842	1.337	.250		.008
Word Frequency	1	7.61	7.605	12.070	.001	*	.069
Residuals	122	77.25	.630				

Table 4

ANOVA Factors of Context Test Scores

Factor	*Df*	*Sum Sq*	*Mean Sq*	*F*	*p*		η^2
Group	1	4.34	4.336	4.978	.028	*	.021
Looked Up	1	.29	.291	.326	.569		.008
Word Frequency	1	2.85	2.853	3.276	.073		.008
Group x Looked Up	1	9.38	9.377	10.493	.002	*	.070
Residuals	121	105.38	.894				

In the non-context test, the word frequency (d = -.735) and whether it was looked up (d = -.851) by the participant were significant factors. As Figure 4 and 5 show, more frequent words and words that were clicked were more likely to be answered correctly in the non-context recall test.

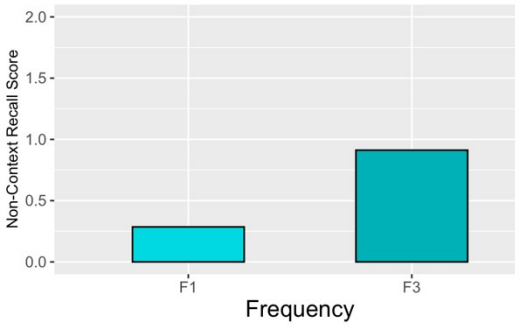


Figure 4. Effect of frequency on non-context recall.

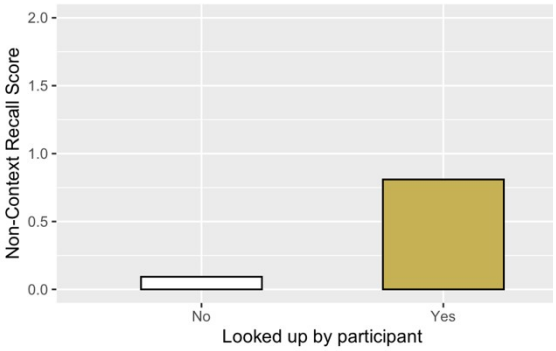


Figure 5. Effect of lookup on non-context recall.

For the context score, group was a significant factor, and there was also an interaction between group (d = .391) and whether the participant looked up the word. In this test, the PD group scored significantly lower (see Figure 6), but as Figure 7 shows, this was true for words that the participants did not look up. Simultaneously, participants in the G group scored considerably higher on words that they did not look up. Table 5 shows results of both tests.

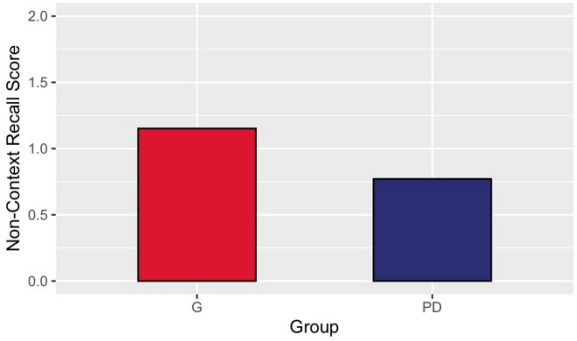


Figure 6. Effect of group on context recall.

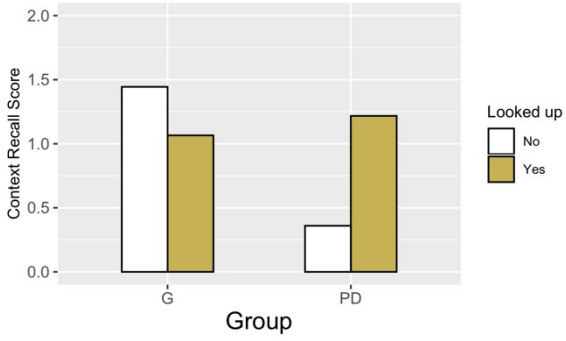


Figure 7. The effect of group and looking up a word on context recall.

Unfortunately, the data sample was too small to assess whether the number of lookups was a significant factor. Whether participants who looked up more words had trouble memorizing them could also not be determined. There was no difference between the two groups in total number of dictionary/gloss clicks (p = .519, d = .401) or the number of word types displayed (p =.932, d = -.054).

Table 5

Descriptive Statistics of Post-Test Scores

		Non-context Score		Context Score	
Group	*N*	*M*	*SD*	*M*	*SD*
Gloss Type					
PD	48	.63	.94	.77	.95
G	79	.53	.89	1.15	.99
Word Frequency					
F1	70	.29	.70	.86	.97
F3	57	.91	1.00	1.19	.99
Looked Up					
No	43	.09	.43	.82	.96
Yes	84	.81	.99	1.11	.99
Sum	181	.57	.90	1.01	.99

4.4.2. Analysis of Lookups

An ANOVA analysis was performed to determine factors that contribute to a word being clicked. As Table 6 shows, group, frequency, and topic familiarity were all significant factors, with additional interaction between group and frequency. Participants in the G group looked up more target words by a great deal (d = .644); this was, however, not true for F3 words, which were looked up by both groups to the same extent. This relationship is illustrated in Figure 8. More frequent words were more likely to be looked up (d = -.576). Participants who were more familiar with the topic looked up more words (d = -.927).

Table 6

ANOVA Factors of Lookups

Factor	*Df*	*Sum Sq*	*MeanSq*	*F*	*p*		η^2
Group	1	2.563	2.563	16.45	.000	*	.021
Word Frequency	1	2.11	2.120	13.54	.000		.008
Topic Familiarity	1	3.137	3.137	20.14	.000		.008
Group × Frequency	1	1.626	1.626	10.44	.002	*	.070
Residuals	122	19.006	.156				

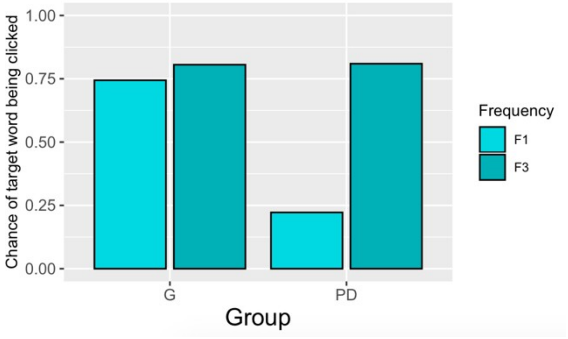


Figure 8. Interaction of group and frequency on lookups.

4.4.3. Participants Analysis

The number of times a participant used the dictionary to look up the meaning of a word was not affected by their familiarity with the topic (p = .142, d = -.771) or their English proficiency (p = .211, d = -.771). There was also no significant effect of the gloss type (p = .906, d = -.207). The same can be said about the number of unique word types each participant looked up. However, when the analysis is limited to target words only, we can see that the G group participants were more likely to look up target word's meanings (p < .001, d = -5.205). One more interesting result is that participants in the G group marked 82.29% of target words as previously known but looked up 97.71% of target words, meaning they were using the gloss to look up the meaning of previously known words.

The accuracy of the participants' answers to the reading comprehension questions were also not affected by their previous knowledge (p = .383, d = .146), proficiency (p = .086, d = -.394), and group (p = .887, d = .233).

Table 7 and Figure 9 and 10 display the effect (or lack thereof) of group on these variables. Figure 11 displays difference in target words looked up by each group.

Table 7

Descriptive Statistics of User Analysis

		Comprehension Score		WPM		Looked-up types	
Group	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Proficiency							
Low	7	.66	.25	73.21	23.76	17.71	11.42
High	4	.78	.28	131.95	41.04	12.75	3.60
Topic Familiarity							
Low	6	.63	.27	105.10	50.69	18.33	12.03
High	5	.80	.19	81.95	27.**71**	13.00	4.53
Gloss Type							
PD	6	.78	.26	104.65	59.69	16.20	14.84
G	5	.64	.26	86.17	20.21	15.67	.52
Sum	11	.70	.25	94.57	41.51	15.91	9.40

Note. WPM stands for words per minute; it shows the speed at participants to read the text.

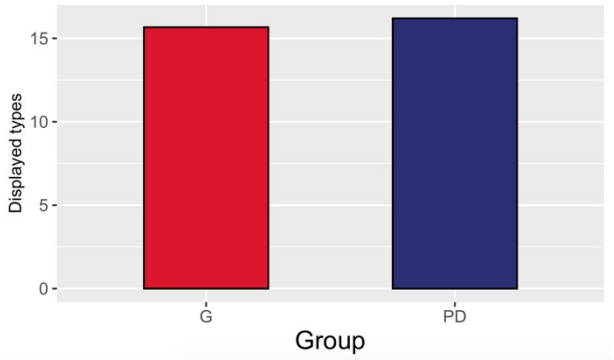


Figure 9. No significant difference in looked up types between groups.

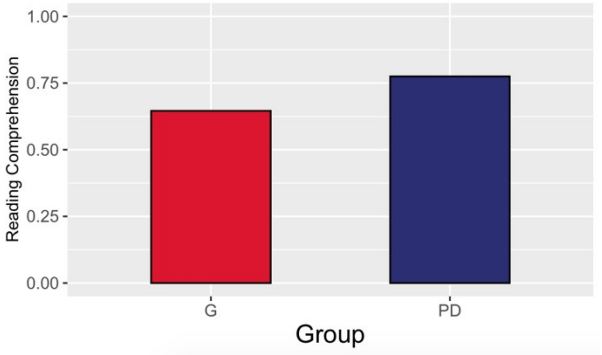


Figure 10. No significant difference in reading comprehension between groups.

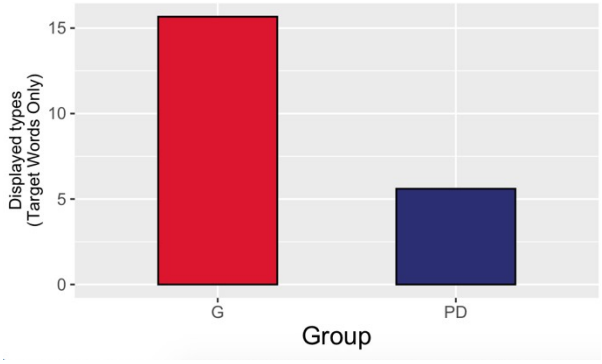


Figure 11. Effect of group on the number of target words looked up.

4.5. Discussion

RQ1.1 Is the number of successfully memorized target word affected by the number of non-target words looked up by participants in the PD group?

The initial hypothesis was that readers who use the pop-up dictionary might often use it to look up different words than those intended by the material's author. This hypothesis was confirmed, as participants in the G group looked up twice many target words as the PD group. However, when comparing the vocabulary test results, no main effect was found for non-context scores on both groups. For the non-context group, the main predictors were the word frequency and whether the word was clicked or not. Moreover, the mean of non-context vocabulary scores was around 0.5 on a 2-point scale, which corresponds to participants retaining an average of four new target words out of 16.

These results, therefore, show that although the G group looked up almost all target words, they could not retain more than the PD group. The reason for this might be due to memory constraints — after all, a 25% score on a meaning recall test can be considered relatively high. However, participants in the PD group used the pop-up dictionary just as frequently, because their use included non-target words. Still, they were able to achieve 25% accuracy on the non-context post-test.

It seems that participants in the PD group were selective in which words to look up. Participants chose words because of their relevance to the text, and therefore it might have been easier to retain them, although it is not clear whether this selectiveness was intentional or if they could not pay attention to novel words.

For the context post-test scores, there was an interaction of group and whether the word was clicked or not. In the G group, new target words not clicked seemed slightly easier to be retained. It is possible that participants focused on these words and tried to guess their meanings from the context. Although the cognitive load of guessing a meaning was not always enough to form a strong form-meaning connection, seeing the context again might have reactivated the word knowledge for meaning recall. Words that were not clicked in the PD group had a much lower recall rate, as participants could possibly not focus on them.

RQ1.2 Does the pop-up dictionary affect reading comprehension?

This experiment produced no significant group effect on reading comprehension. The number of types looked up also did not produce a statistically significant effect. These results are in line with Mekheimer (2018) or Liu and Lin (2011), who also did not find an effect of pop-up dictionary on reading comprehension. As this study did not employ intrusive methods such as a paper dictionary or online dictionary use, small differences in reading comprehension scores could be predicted. However, the interesting point is that although participants in different groups used the gloss to look up different words, their accuracy score was not affected. Since the reading comprehension data was aggregated after the reading comprehension test, it is impossible to see whether both groups differed in which questions they were likely to answer correctly. However, it could be a compelling aspect of future experiments.

RQ1.3 Does the number of times a target word appears in the text affect the results described in RQ1.1 and RQ1.2?

It was confirmed that more frequently appearing words were more likely to be recalled in a non-context test. This result might have been heavily influenced by the fact that frequency was a significant factor in deciding whether a word will be looked up or not. The results indicate that participants evaluated more frequent words as words closely related to the text's topic and were more compelled to look up their meaning. However, it is also possible that the participants did not perform such an evaluation since frequent target words made up a greater portion of the text, which caused the participants to click on them.

In either case, the results of this experiment advocate for the use of input flooding as an effective tool that ensures the reader looks up the words we choose.

Discussion Overview

The results of Experiment 1 show that despite being a powerful tool, the pop-up dictionary in the hands of a language learner might lead to some unfavorable results. It was expected that participants in Experiment 1 would use the pop-up dictionary to search for all unknown words to understand the text better. For most participants with a pop-up dictionary, this was, however, not the case. These participants looked up fewer target words than those in the G group. They would often skip over words they did not know, as the vocabulary test results show.

It is difficult to assess why participants in this experiment did not frequently use the pop-up dictionary. Perhaps they did not feel the need to use it, as they could grasp the meaning of the text even with a limited understanding of the words used in the text. There is also a possibility that the participants could not notice the gap in their knowledge and did not pay enough attention to the unknown words. However, since participants who were more familiar with the text looked up more words, there is an indication of a motivational factor. Since the topic of the text material was on pre-ancient people in Middle America, perhaps future experiments could compare a group of learners studying ethnology or archeology and compare their pop-up dictionary use to learners with different majors.

In the G group, input enhancement compelled participants to look up almost every target word, including previously known words. There is a chance that the visual salience of the target words led participants to realize the gap in their vocabulary knowledge and look up the word. Unfortunately, the effect of input

enhancement could not be isolated, as there was no group with the combination of pop-up dictionary and input enhancement. To better understand the strategy of learners reading with pop-up dictionaries, Experiment 2 addresses this limitation by comparing pop-up dictionary group with or without input enhancement.

5. Experiment 2

5.1. Objectives

The main objective of Experiment 2 was to investigate the strategy of a reader using a pop-up dictionary. This experiment was mostly motivated by the fact that participants in the previous study used pop-up dictionary only scarcely, which was not expected, as a pop-up dictionary is a relatively easy and accessible method of learning the meaning of words (compared to looking up the word in an external dictionary or guessing the word's meaning from context). The three following aspects were given special consideration in designing Experiment 2: text type, group design, and online measurements.

Text Type

One possible explanation for why participants did not give the expected attention to target words in Experiment 1's PD group is that they did not feel the need to know those words. As described in the involvement hypothesis (Laufer & Hulstijn, 2001), the desire to know the meaning of a word is crucial in vocabulary learning. The text used in Experiment 1 was an expository text aimed at children (albeit native speakers of English), meaning there were possibly redundancies in the language, which allowed the reader to understand the text without comprehending all of its elements. To reduce the possibility that participants will replicate this behavior in Experiment 2, another material candidate, a narrative text, *The Midas Touch*, was selected. In a small-scale pilot study (N = 2), the narrative text proved to contain a much higher rate of words, which the reader felt compelled to learn its meaning compared to the expository test used in Experiment 1.

Group Design

Experiment 1 had a severe limitation in that the two experimental groups were not minimal pairs. They employed different types of gloss and differed in the presence of input enhancement. To address this issue (and assess the effect input enhancement has on the way a pop-up dictionary is used), Experiment 2 divided participants into four groups, based on two factors (input enhancement and gloss complexity) with no other differences between them.

Online Measurements of Attention

An online measurement method was needed to better understand the participants' strategy (or strategies). First, the think-out-loud protocol was deliberated; however, considering that participants may not voice out everything they notice – even under a think-out-loud protocol – coupled with the possibility of reactivity (Chaudron, 1985), this idea was dismissed.

The eye movement measurement was also a method considered. There were also drawbacks to this method, as it would be difficult to execute such movement with the functionality of a pop-up dictionary. Coincidentally, in the shadow of the global pandemic, the university forbade any experiments where students would come into direct contact with each other.

In response to the situation, the author developed an online software that collects data about cursor movement on the experiment's website. All words except those in the vicinity of the cursor were masked with black squares, which allowed the collection of data on words each participant looked at and the duration, virtually simulating the functionality of an eye movement measurement.

In this study, an external test of attentional capacity was not performed due to time constraints (the author was afraid that long experiment time would demotivate participants when conducted remotely). Participants were also not asked about developmental or learning disorders because it is estimated that many people have these disorders without being diagnosed (Gerschon, 2002); therefore, it would not be a reliable factor.

Research Questions

The following research questions will be answered in this study, together with other findings from Experiment 2:

RQ2.1 Does input enhancement increase the chance a word will be looked-up by the participant?

RQ2.2 Is presenting a single context-fitting meaning in a gloss more effective for vocabulary acquisition than presenting multiple dictionary entries for each word?

RQ2.3 Are participants able to pay attention to all unknown target words?

5.2. Method

5.2.1. Participants

This study was limited to students at the University of Tsukuba graduate and undergraduate schools to obtain a more homogenous sample than in Experiment 1. The only other condition was that they were native speakers of Japanese. First, participants were collected through a notice board for several classes conducted by the author's professors. This method proved ineffective, and the author chose to distribute flyers advertising the experiment around the city of Tsukuba. All participants who finished the experiment were sent a 1500-yen Amazon gift card, addressed to their university email, as promised on the flyer.

At the time of the analysis, 57 students participated in the study, of which only 47 had completed the entire experiment. Two participants did not engage meaningfully with the experiment, and therefore their data was removed, resulting in data of 45 participants ready for analysis.

5.2.2. Materials

Text

As described in previous sections, a narrative text (*The Midas Touch*) about a Greek legend was selected for this experiment. A small-scale pilot study preceded this decision. In the study, participants were asked to select all unknown words in a text, and for each unknown selection, how much they think they need to know the word's meaning on a scale of 1–4. For the comparison, excerpts from the text used in Experiment 1 and *The Midas Touch* were used. The results showed that although the excerpts from both texts contained roughly the same percentage of words unknown to the participants, more of those words were marked by the participants as important in understanding the text in *The Midas Touch*.

The Midas Touch is 1,034 words long, and the FKGL grade of readability was calculated to be 5.1 (https://www.webfx.com/tools/read-able/check.php).

Reading Software

A software was developed to collect information about the participants' interaction with the text. The software analyzed the letter of the text closest to the cursor and only revealed the nine nearest letters. As Figure 12 shows, spaces between words were left unmasked so that participants could navigate better through the text, for example, when they wanted to reread the previous section of the text. The time spent hovering over each word, the delay before the user clicks a word, and the

number of times they clicked each word was recorded. All participants could confirm how this software works on a tutorial screen before reading the experimental text.

As scrolling could severely interfere with the accuracy of the collected data, the text had to be presented in chunks small enough to fit a standard-sized screen. After finishing a chunk, participants could move onto the next chunk. They could not, however, return to previous chunks of text.

There was a transition screen between each chunk of text, where the participants were asked to click a button to continue reading. As a result, participants could take a break when they needed one, but this primarily ensured that the cursor would return to an initial position, as the button was positioned at the top of the screen.

Figure 12. Example of reading software used in Experiment 2.

Gloss

Each word in the text was stored in a database. Two sets of translations were assigned to each word. One set would only display a single context-fitting gloss in Japanese (single gloss; SG). The second set contained up to five Japanese translations in the order they appear on the Japanese-English online dictionary jisho.org (multiple choice gloss; MCG). These five translations were not guaranteed to fit the context of the word. If a dictionary entry for an English word had less than five translations, less than five meanings would be presented to the participant. However, to prevent variance between target words, all target words were engineered to display five meanings. Some translations of target words were derived from the

target word synonyms when the original target word would not produce enough translations (translations for *slobber* were replaced with translations for *saliva*) or when the translation was a low-frequency word in Japanese (translations for *torrent* were replaced with translations for *stream*). Each participant was randomly assigned either SG or MCG condition, which applied to all of the words they clicked.

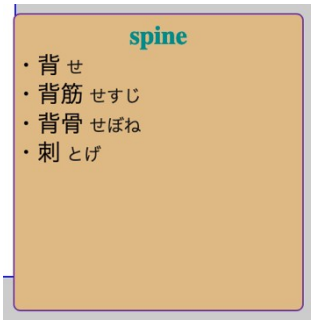


Figure 13. Example of gloss used in Experiment 2.

Figure 13 shows the MCG gloss in action. In the MCG condition, participants would sometimes have to choose the correction that fits the context the most. In this case, *spine* referred to the sharp things on the back of a porcupine's back, which is 刺 (とげ) in Japanese.

Comprehension Questions

A set of 10 true or false questions were created to measure each participant's level of comprehension of the text. All questions asked were about information explicitly stated in the text. The questions were written in English, and they were formed in a way that did not require the knowledge of any of the target words in answering them.

Vocabulary Tests

At first, the vocabulary knowledge scale (VKS) was considered as a method of assessing the number of target words retained by the participants. However, in a different study conducted by the author, the VKS provided very limited data. Participants could not recall the meaning of most of the words, and they were not rewarded for meaning recognition. Therefore, a set of three tests — form recognition, meaning recall, and meaning recognition — was used. These three tests were conducted in the order described here.

Form Recognition Test

In this test, the participants were presented with 40 low-frequency words. Among them, 20 were target words, and the other 20 were filler words never used in the text. Participants then had to select all words they saw in the text with a checkbox. The form recognition test also functioned as an offline method of measuring the participants' awareness of these words.

Meaning Recall Test

In the second vocabulary test, participants were only presented with the 20 target words, and they were prompted to answer the meaning of each word in Japanese.

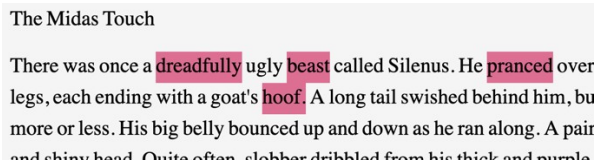
Meaning Recognition Test

The last part of the vocabulary test displayed the same 20 target words. However, this time, participants only had to choose the correct meaning from three options. One was the correct option, the second was a translation of an English word that sounds similar to the target word, and the third was a translation of a different word appearing in the text or related to its theme.

Unknown Words

After completing the three vocabulary tests, participants were presented with a screen containing the text in its full length. They were then asked to click and select all words which they had not known before the study. The participant's input was recorded for all words found in the text. While analyzing the data, it was discovered that participants would often forget to select a target word whose meanings they were not knowledgeable of (based on the meaning recall test). This discovery could be due to participants' fatigue or possibly the fact that they misinterpreted those target words as similar-sounding or similar-looking words. It is also possible that participants misunderstood the instructions, which asked to answer based on their knowledge before the test.

To address this oversight in the experiment's design and to obtain more accurate data about the participants' previous knowledge of the words, an email was sent to participants with a link to a follow-up study, where they were presented with each target word (and its translation) with the option to select the degree to which they knew the word before the study. Only a portion of participants completed this study, and for those participants, the data about previously known words were updated.


The Midas Touch

There was once a dreadfully ugly beast called Silenus. He pranced over
legs, each ending with a goat's hoof. A long tail swished behind him, bu
more or less. His big belly bounced up and down as he ran along. A pair

Figure 14. Participants had to choose all previously unknown words on this screen.

Figure 14 shows the unknown word selection screen. Once clicked, a word would turn red, indicating it is a previously unknown word. Participants were informed that clicking a word again would return it to a normal state.

Questionnaire

After completing the experiment, participants were asked to answer a questionnaire. This questionnaire mostly contained questions about the participant and their academic background in the first half and questions about how they felt about the pop-up dictionary as a device to promote reading comprehension and vocabulary acquisition. In determining the level to which vocabulary acquisition was incidental, participants were also asked whether they tried to remember the words they looked up intentionally. Most of this data was collected on a 5-grade Likert scale, while the two last questions of the questionnaire asked for comments about the pop-up dictionary.

5.3. Analysis

5.3.1. Eye Movement

The time each participant spent hovering over each word and the number of times each word was clicked by the participants was recorded, which resulted in over 100,000 data points. The data concerning the 20 target words were selected and aggregated to obtain the first gaze duration and total reading times for each occurrence of the target word. The number of times each occurrence of a target word was fixated on and the number of times it was clicked was also recorded. From the data obtained, the total reading times across all occurrences were calculated. All gaze durations shorter than 34ms were omitted from the analysis, based on the claim that fixations shorter than 50 ms are not long enough for vocabulary intake to happen, as

is acknowledged in Godfroid, Boers, and Housen (2013), and the fact that the error of time measurement in JavaScript is 16 ms ($\frac{1}{60}$ of a second).

Since the cursor's position and not the actual gaze was measured, it is important to acknowledge this method's limitation. All gaze duration times include — along with the time participants spent fixating on the target word — the time participants spent fixating on the gloss. These measures will be identified as *word and gloss* (WG) *duration*. Therefore, the time participants spent fixating on a word before clicking it and displaying a gloss was extracted from the data and used for analysis as *duration to click* (TC). TC gaze durations are important as they tell us the minimal time participants spent interacting with the English form. However, they are not very accurate when a participant clicks a word very quickly upon gaze. Comparing the original WG gaze duration values and the TC values would suggest the time it took the participant to decide to click the word.

The analysis was mainly performed with participant × target word interactions as the unit to account for the gaze duration data. This resulted in 900 data points (45 participants × 20 target words). After removing the words that were not reported by the participant as previously unknown, 799 data points were left for the analysis.

Since the two ways of measuring reading times (including or excluding time after clicking the word) and the number of interactions between reading time measures and other factors were far too complicated, it was decided to perform a separate ANOVA test for eye-tracking factors. Because two different ANOVA tests were performed on the same set of data, a possibility of a type I error is acknowledged. However, the author claims the validity of the statistics without

reading times, at least on the premise that other studies did not account for reading times by the design of their experiment. It is then left to the reader's discretion, whether they wish to evaluate the results according to the Bonferroni correction (Armstrong, 2014), where the null hypothesis can only be rejected if the p value is lower than $\frac{\alpha}{k}$, where α is the level of significance, and *k* is the number of performed tests. In the context of the current study, this would require a p value of .025 or lower for a significant result.

5.3.2. Scoring

In the Form Recognition test, a point was awarded to each selected target word, as seen in the text. The overall accuracy was roughly 50% (*M* =.47, *SD* = .50).

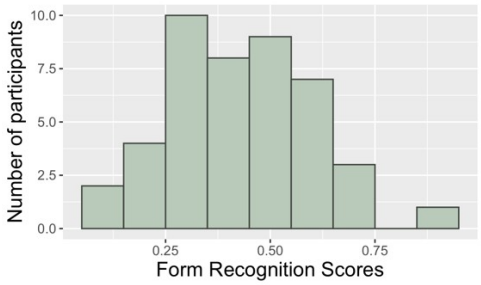


Figure 15. Distribution of form recognition scores.

In the Meaning Recall test, a point was given to each target word whose Japanese translation was provided. Translations provided in the gloss as well as unlisted synonyms were also considered correct. The average accuracy was over 25% (*M* = .26, *SD* = .44) and the highest achieved score was 65%.

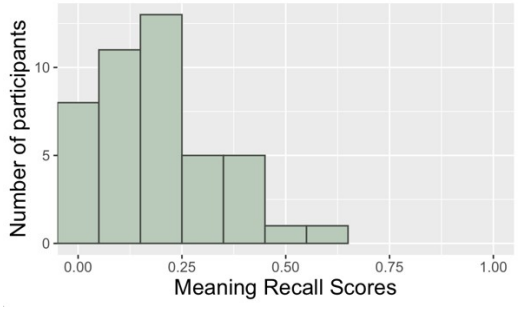


Figure 16. Distribution of meaning recall scores.

In the Meaning Recognition test, the number index of each selected answer was recorded, and the correct answer was awarded one point. This test's accuracy was the highest (M = .81, SD = .39), with the lowest score being around 50%. However, there were only two participants with a full score in the Meaning Recognition test, and only one word was guessed correctly by all participants. Therefore, the author maintains that although the mean scores for each participant were high, the ceiling effect was not confirmed. While the correct answer was selected 727 times, phonologically related distractors attracted the participants in 76 cases (8.4%), and thematic distractors were chosen in 92 cases (10.2%).

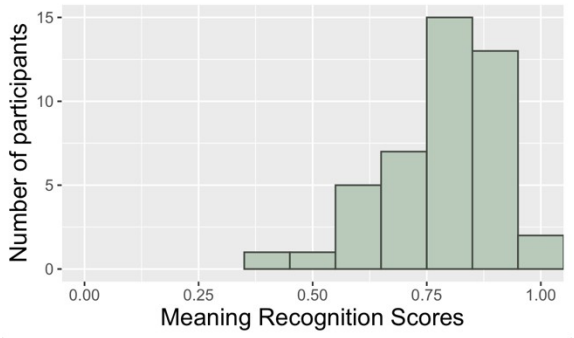


Figure 17. Distribution of meaning recognition scores.

5.3.3. Unknown Words

The number of unknown words in the text and the distinction between unknown and known target words were based on the participant's self-report.

However, participants failed to report all unknown words, as they could not answer the correct meaning of some target words in the post-test, which they did not list as previously unknown. All participants for which this was true were asked to answer whether they knew each target word in a subsequent study. Of 45 participants, only 22 have either reported their previous knowledge of target words correctly or have participated in this subsequent study. For participants who took part in the subsequent inquiry, their self-report data on unknown target words was updated. For all participants, target words that were not answered correctly in the Meaning Recall test were marked as previously unknown, regardless of the participants' report. It is assumed that some of the previously unknown words which were successfully acquired failed to be reported as unknown, which led to them being unnecessarily excluded from the analysis, shortening the sample size. The method for final judgement of previous word knowledge is illustrated by Table 8.

Table 8

Previously Known Words Judgement

	Self-reported as previously unknown	
Meaning Recall	*No*	*Yes*
Incorrect / No Answer	Unknown	Unknown
Correct Answer	Known	Unknown

Note. If a word was answered correctly in the meaning recall test, but the participant failed to report it as previously unknown, it was considered previously known. This might have led to the unnecessary trimming of useful data.

5.3.4. Method of Analysis

In determining the relationships between variables obtained in this study, a choice of analysis method had to be made. Although the linear mixed model was the

desired method of analysis, 799 points of data proved insufficient to construct a converging model with random slopes and interactions between fixed effects. Therefore, the effect of categorical variables was analyzed using the Analysis of variance (ANOVA). Due to the number of continuous factors obtained from eye-tracking, these were transformed into binary variables (low × high) and used in a separate analysis.

5.4. Results

5.4.1. User Analysis

First, the data from the 45 participants were analyzed. Their mean accuracy during the comprehension for each participant test was calculated (M = .69, SD = .15) and used as a dependent variable. The analysis showed no effect of input enhancement (p = .702) or gloss type (p = .463) on the accuracy of the answers. Whether participants tried to memorize new words (p = .538) or how many times they used the pop-up dictionary (p = .907) also had no effect on comprehension. Pearson's test of correlation showed that participants who scored better on the meaning recognition test also scored better on the comprehension questions (r = .554; p < .001). Figure 18 displays the distribution of comprehension scores.

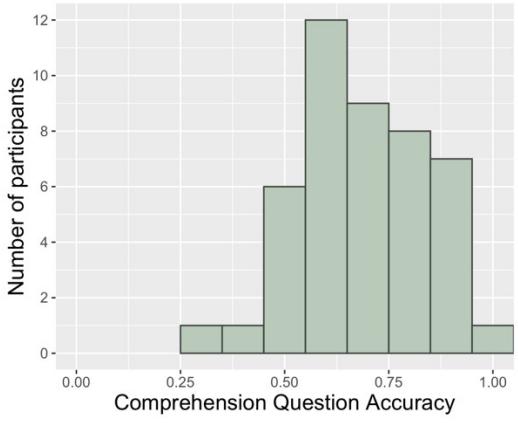


Figure 18. Distribution of comprehension scores.

5.4.2. Questionnaire Results

In a questionnaire issued after the study, participants were asked to answer questions such as “do you think this type of gloss promotes comprehension?” Participants reported their insight on a scale of 1-5, with 1 meaning “I do not think so at all” and 5 meaning “I fully agree.” Statistical analysis of participants' answers found a significant interference between the input enhancement and gloss type as illustrated by Figure 19 (p = .009). From Table 9, we can see that participants self-reported that MCG helped their comprehension, but only when combined with input enhancement. In the case of no input enhancement, MCG seems to have hindered comprehension.

This result is fascinating, because it also reflects how gloss type and input enhancement interact to predict vocabulary acquisition. (For more information, see results for RQ2.2.).

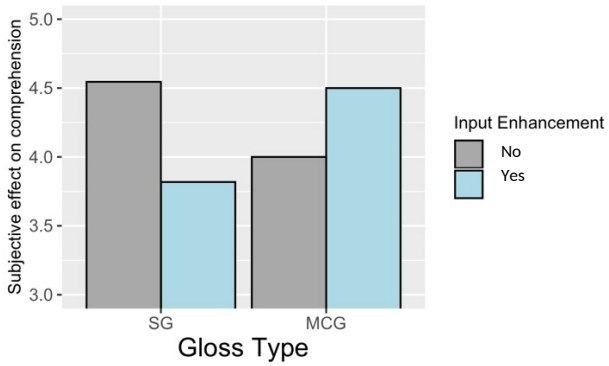


Figure 19. Experimental groups' view of the effect of pop-up dictionary on comprehension.

The answers to a certain question varied significantly between groups. The question was, “did you feel the words presented by this gloss were more memorable?” and participants with input enhancement replied more positively than

those without input enhancement (p = .041, Cohen's d = 1.790). Figure 20 and Table 9 describe this result in detail.

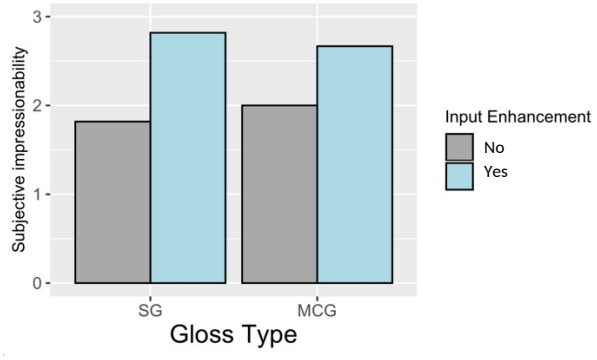


Figure 20. Experimental groups' view on the impressionability of words.

Table 9

Descriptive Statistics of Questionnaire Results

	Subjective measures			
	Helps Comprehension		Target Word Impressionability	
Group	*M*	*SD*	*M*	*SD*
SG IE-	4.55	.52	1.88	1.68
SG IE+	3.82	1.25	2.82	1.60
MCG IE-	4.00	.45	2.00	.89
MCG IE+	4.50	.52	2.67	1.50

It should be noted that there was no significant effect of input enhancement or gloss type on whether the participants tried to learn the words intentionally. Participants who were considered intentional learners (participants who answered 3 or above on this question) were equally distributed and constituted roughly one-third of each of the four experimental groups (input enhancement × gloss type). There was also no significant difference in how much participants from each group thought the

gloss makes them lose track of where they were reading or on their motivation to infer the words from context or to check them in a dictionary later.

Participants were also asked to report their English proficiency in the form of any English proficiency certificate they might have received. Among their answers, the Test of English for International Communication (TOEIC®; Educational Testing Service) was the most common and therefore was used as a point of reference. Some participants reported their proficiency with the Test of English as a Foreign Language (TOEFL®; Educational Testing Service) test scores or the Japanese English proficiency test (Eiken) levels. These were converted to the TOEIC score using two reference tables. (https://toiguru.jp/toeic-and-eiken; https://www.conversation.jp/faq/faq-english/TOEIC-TOEFL.html) Participants who reported their Eiken level as 3 or lower were considered to have taken this examination during their previous education. Therefore, their level of English proficiency was considered unknown. Therefore, proficiency data was collected only from 31 students.

Although the sample size did not allow for English proficiency to be included in the analysis, Figure 21 shows a histogram that provides a pictorial reflection of the participant's English level.

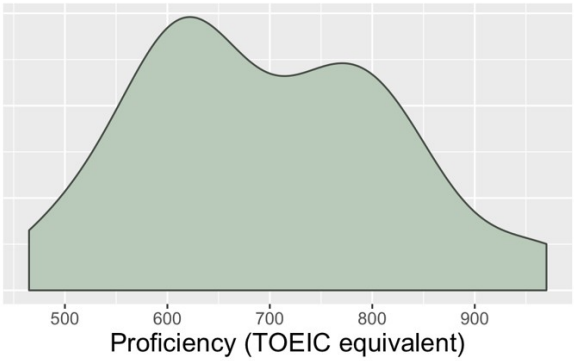


Figure 21. Histogram of participants' proficiency.

Two questions in the questionnaire were designed to grasp a better idea of how participants felt about the pop-up dictionary. The first question was, "did you feel this kind of dictionary is easy to use?" Among 47 students who participated in the study, 39 said the dictionary was easy to use. The main reasons were that the Japanese meanings were shown on the same page, and therefore did not disturb their reading. Four participants in the SG group stated that they appreciated that a context-fitting translation was shown, and therefore there was no need to seek different translations. One participant presumably compared this dictionary to static glosses when they stated that looking up the meaning of a word even when it appeared the second time was a nice feature. One participant compared this dictionary to its counterpart in Google Chrome and appreciated that the dictionary in this study was activated by clicking the word (instead of selecting it).

These results suggest that eight participants were not satisfied with the gloss. The main reason, provided by six participants, was the lack of information about collocations. Two participants in the MCG group were discontent because sometimes they could not apply any of the shown translations to the context. One participant wished that the MCG dictionary would be more structuralized, dividing translations into sections by the closeness of their meaning.

The second question was optional, asking participants to state any other thoughts about the dictionary, with sample comments such as “I did not try to guess the meaning of new words” to give participants an idea of what to write. Only 25 participants answered this question. Among them, nine participants admitted to relying on the dictionary. They expressed that the ease of use of the dictionary made them rely on lexical inference less, and they were conscious that this had a negative

impact on their ability to remember those words. Two of them stated that they would sometimes see an unfamiliar word and click it without paying attention to the spelling. One participant admitted to relying on the dictionary so much that they looked up the same word repeatedly.

Four more participants commented that they could not focus on the context when the correct translation for a word was provided. One of them stated that it was sometimes difficult to choose a context-fitting translation from MCG. One other participant also found it difficult (and time-consuming) to choose the correct translation from MCG.

Another set of four participants commented about using the pop-up dictionary to look up words they had previously known. Some stated it was due to a lack of confidence in their knowledge of that word. One participant expressed curiosity about what the dictionary entry looks like for a word they had previously known.

Only three participants reported trying to guess new meanings of a word. One more participant (MCG group) stated that they would have liked to see more information in the gloss.

5.4.3. RQ2.1 Does input enhancement increase the chance a word will be looked-up by the participant?

A single binary variable showing which words were clicked at least once by each participant was calculated from the data. When limited to unknown words, which the participant is fixated on, it was found that participants clicked most but not all target words (M = .88, SD = .32). An ANOVA test of variance was conducted to explore the effect of several variables on this binary variable. Table 10 shows the p values and effect sizes of each variable.

As is apparent from the table, there was a significant effect of input enhancement for unknown target words (p = .023, d = .935). There was an interaction between input enhancement and gloss type (p < .001), where participants with input enhancement seem to have looked up fewer words when their gloss type was MCG. Although it was not the initial intention to include the self-reported intentional learning (1-5 scale of how much the participants tried to memorize the words they looked up), there was a significant interaction between this variable and gloss type (p < .001). As Table 11 and Figure 23 show, MCG caused intentional vocabulary learners to focus on target words less than incidental vocabulary learners.

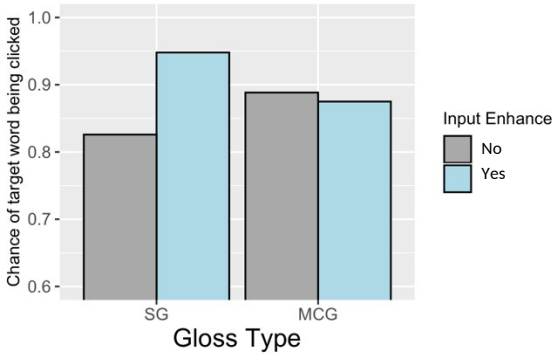


Figure 22. Differences between groups in number of target words clicked.

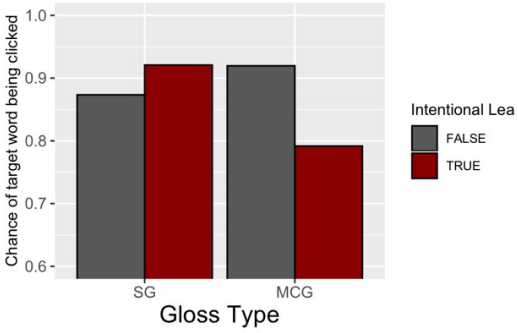


Figure 23. Effect of gloss type and number of target words clicked.

no
yes

Table 10

Factors of Chance a Target Word Is Looked Up

Figure 24. Chances of a target word being clicked based on accuracy.

Learning Mode	1	.33	.331	3.353	.067		.003
Gloss Type	1	0	.001	.008	.929		.000
Word Frequency	1	.84	.836	8.465	.003	*	.009
IE	1	.51	.511	5.167	.023	*	.005
Learning Mode×Gloss Type	1	1.12	1.123	11.37	.001	*	.016
Gloss Type×IE	1	1.13	1.134	11.477	.001	*	.014
Residuals	792	78.24	.099				

Table 11

Descriptive Statistics of Lookup Factors

	Chance a target word will be clicked		
	N	*M*	*SD*
Input Enhancement×Gloss Type			
SG IE-	201	.83	.38
SG IE+	192	.95	.22
MCG IE-	206	.89	.32
MCG IE+	200	.86	.33
Learning Mode×Gloss Type			
SG Inc.	292	.87	.33
SG Int.	101	.92	.27
MCG Inc.	286	.92	.27
MCG Int.	120	.79	.41
Frequency			
F1	362	.85	.36
F3	437	.91	.28

There was a significant effect of frequency (p = .004, Cohen's d = -.203). More frequently appearing words were more likely to be clicked. Overall, participants clicked 88.36% of target words, which they marked as previously unknown. Among cases where the target word was not clicked — only 13% (12 cases) — participants were able to answer the word's meaning correctly on a recall test. The effect of meaningful factors on clicking on target words is visualized in Figure 22 to 24.

Because participants were not instructed to focus on a particular set of words, such as the 20 target words employed in this study, it was essential to examine the participants' behavior in a way that is not limited to target words. For such purposes,

the following variables were calculated: click count (non-distinct number of times a participant used the gloss) and non-target types clicked (distinct number of non-target word types clicked by the participant).

Presented next is the participant analysis with the number of non-target words clicked as a dependent variable. As Table 12 shows, there was only a significant effect (p = .030) of intentional learning. As can be seen from Figure 25, intentional readers clicked much fewer words.

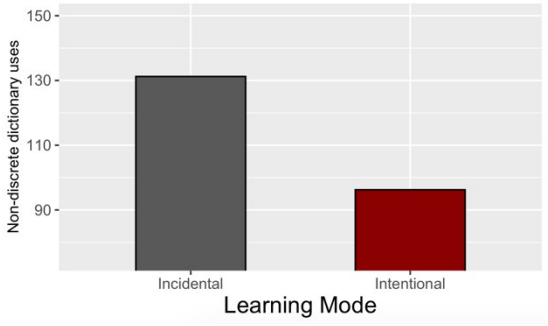


Figure 25. Effect of learning mode on the total number of dictionary uses.

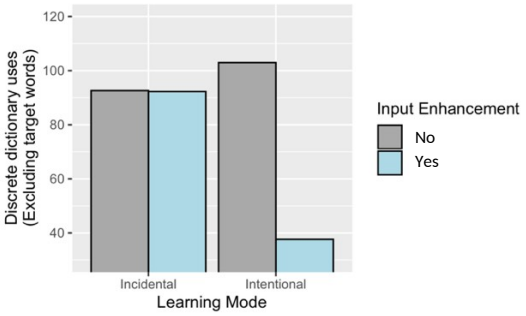


Figure 26. Effect of learning mode and IE on the number of non-target types looked up.

The analysis of non-target types clicked shows that intentional learning and input enhancement are significant (p = .028). The group of intentional learners with input enhancement clicked significantly fewer non-target words than any other group. These results can be seen in Figure 26 and Table 12.

Table 12

Effects of Learning Mode and IE on Looking Up of Non-Target Words

	Number of non-target words clicked		
	N	*M*	*SD*
Learning Mode×Input Enhancement			
Inc. IE-	15	92.67	37.29
Inc. IE+	17	92.29	49.56
Int. IE-	7	103.00	37.10
Int. IE+	6	37.67	16.98

5.4.4. RQ2.2 Is presenting a single context-fitting meaning in a gloss more effective for vocabulary acquisition than presenting multiple dictionary entries for each word?

Three dependent variables were used to measure vocabulary acquisition: form recognition, meaning recall, and meaning recognition. Independent variables, gloss type, and input enhancement, alongside other variables such as frequency or whether the participant was trying to memorize the words were used. The effect of these variables on form recognition can be seen in Table 13.

Participants who learned vocabulary intentionally were able to recognize more target words ($p = .023$, $d = .337$). More frequently appearing target words were also more likely to be recognized ($p < .001$, $d = -.577$). Gloss type alone did not influence form recognition accuracy significantly. There was, however, a significant interaction between intentional learning and gloss type ($p < .001$). As shown in Figure 28, form recognition was most successful for intentional learners who had access to MCG. Second, came incidental learners who did not have access to MCG.

Table 15 shows significant effects on the meaning recall test. Again, the frequency was a significant predictor ($p < .001$, $d = -.516$), as more frequent words were more likely to be answered correctly. Input enhancement proved beneficial for meaning recall ($p = .008$, $d = -.680$), and there was an interaction between input enhancement and whether the participant had clicked the word ($p = .006$). New words' meanings were recalled most successfully when displayed with input enhancement, but the participant did not click them. However, there was also an interaction between input enhancement and gloss type ($p = .009$). As is apparent from Figure 34, MCG was only beneficial when combined with input enhancement. Simultaneously, participants with MCG, but no input enhancement showed the lowest score in meaning recall. Furthermore, intentional vocabulary learning showed signs of interaction with gloss type to reflect their interaction for form recognition, although the results did not show statistical significance ($p = .064$).

In the case of meaning recognition, frequency ($p < .001$, $d = -.483$) and whether the participant clicked the word ($p < .001$, $d = .256$) were revealed to be the main effects. Like previous tests, the meanings of more frequent target words were more likely to be recognized, and so were words clicked by the participant. These results are shown in detail in Table 14.

Table 13

Factors of Form Recognition

Factor	*Df*	*Sum Sq*	*Mean Sq*	*F*	*p*		η^2
Learning Mode	1	1.15	1.154	5.208	.023	*	.006
Gloss Type	1	0	.001	.004	.951		.000
Word Frequency	1	14.79	14.786	8.465	.000	*	.072
IE	1	.01	.009	.040	.841		.000
Learning Mode x Gloss Type	1	4.53	4.535	20.372	.000	*	.023
Residuals	779	172.57	.222				

Table 14

Descriptive Statistic of Form Recognition Effects

	Form Recognition Score		
	N	*M*	*SD*
Learning Mode×Gloss Type			
SG Inc.	292	.46	.50
SG Int.	101	.37	.48
MCG Inc.	286	.36	.48
MCG Int.	120	.60	.49
Frequency			
F1	362	.29	.45
F3	437	.56	.50

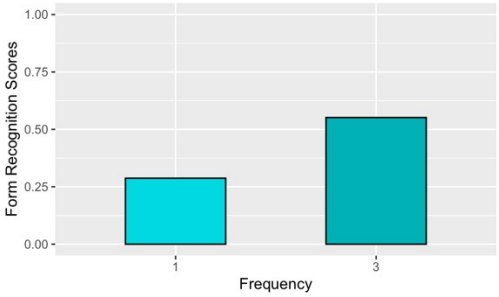


Figure 27. Effects of frequency on form recognition.

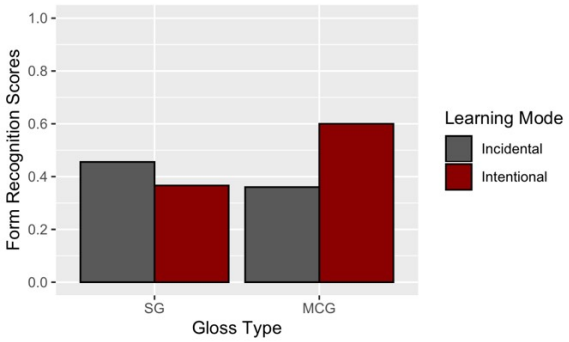


Figure 28. Effects of gloss type and learning mode on form recognition.

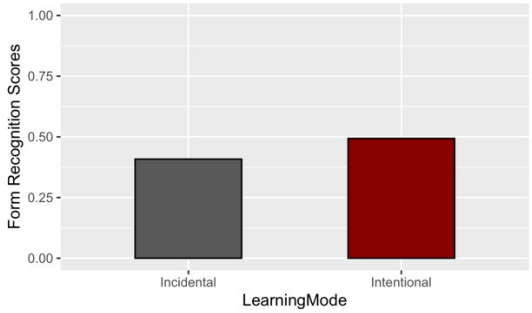


Figure 29. Main effect of learning mode on form recognition.

Table 15

ANOVA Factors of Meaning Recall

Factor	*Df*	*Sum Sq*	*Mean Sq*	*F*	*p*		η^2
Learning Mode	1	.00	.000	.001	.982		.000
Gloss Type	1	.07	.068	.500	.480		.001
Word Frequency	1	6.79	6.788	50.056	.000	*	.055
Looked Up	1	.08	.076	.561	.454		.002
IE	1	.83	.834	6.154	.013	*	.008
Awareness	1	7.85	7.850	62.841	.000	*	.056
Learning Mode × Gloss Type	1	1.23	1.235	9.106	.003	*	.013
Frequency × Looked Up	1	.87	.869	6.407	.012	*	.007
Looked Up × IE	1	.90	.904	6.666	.010	*	.007
Gloss Type x IE	1	.93	.925	6.284	.009	*	.008
Awareness × Frequency	1	.56	.560	4.481	.034	*	.016
Residuals	789	106.99	.136				

Table 16

Descriptive Statistic of Meaning Recall Factors

	Meaning Recall Score		
	N	*M*	*SD*
Input Enhancement×Gloss Type			
SG IE-	201	.19	.39
SG IE+	192	.19	.40
MCG IE-	206	.11	.32
MCG IE+	200	.24	.43
Looked-up×Input Enhancement			
Lookup- IE-	58	.03	.18
Lookup- IE+	35	.28	.46
Lookup+ IE-	349	.17	.38
Lookup- IE+	357	.21	.41
Learning Mode×Gloss Type			
SG Inc.	292	.21	.41
SG Int.	101	.13	.34
MCG Inc.	286	.15	.36
MCG Int.	120	.23	.42
Frequency			
F1	362	.08	.27
F3	437	.27	.44

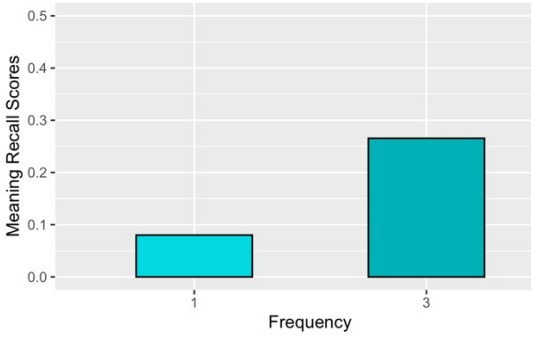


Figure 30. Main effect of frequency on meaning recall.

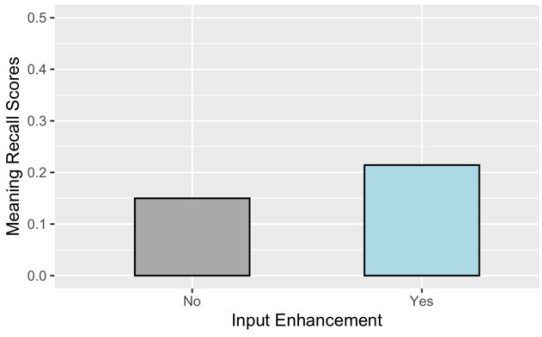


Figure 31. Main effect of IE on meaning recall.

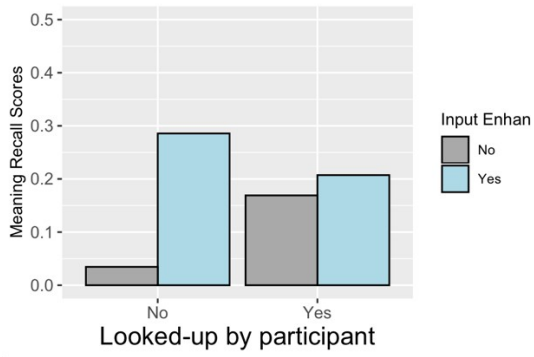
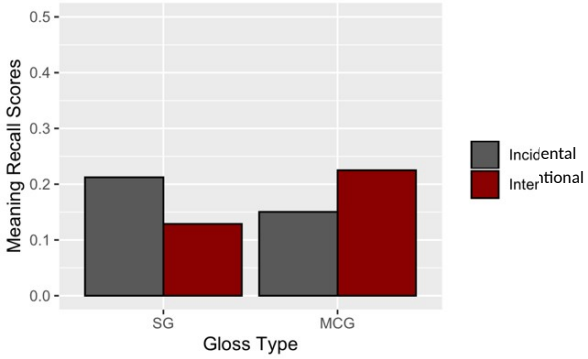


Figure 32. Effects of IE and lookups on meaning recall.



F*igure 33*. Effects of gloss type and learning mode on meaning recall.

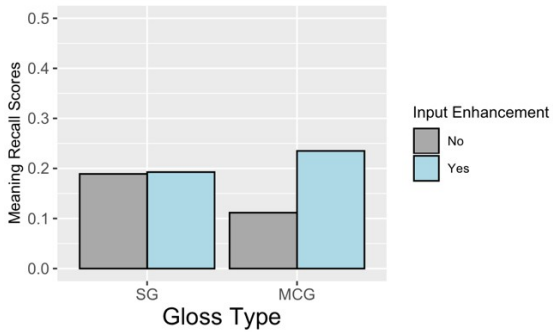


Figure 34. Effects of gloss type and IE on meaning recall.

Table 17

ANOVA of Meaning Recognition Factors

Factor	*Df*	*Sum Sq*	*Mean Sq*	*F*	*p*		η^2
Learning Mode	1	7.25	7.249	50.139	.000	*	.033
Gloss Type	1	.07	.068	.500	.480		.001
Frequency	1	6.79	6.788	50.056	.000	*	.055
Looked Up	1	.08	.076	.561	.454		.002
IE	1	.83	.834	6.154	.013	*	.008
Learning Mode × Gloss Type	1	1.23	1.235	9.106	.003	*	.013
Frequency × Looked Up	1	.87	.869	6.407	.012	*	.007
Looked Up × IE	1	.90	.904	6.666	.010	*	.007
Gloss Type × IE	1	.93	.925	6.284	.009	*	.008
Residuals	789	106.99	.136				

Table 18

Descriptive Statistics of Meaning Recognition Factors

	Meaning Recognition Score		
	N	*M*	*SD*
Input Enhancement×Gloss Type			
SG IE-	201	.82	.39
SG IE+	192	.80	.40
MCG IE-	206	.74	.44
MCG IE+	200	.81	.39
Looked-up×Input Enhancement			
Lookup- IE-	58	.45	.50
Lookup- IE+	35	.63	.49
Lookup+ IE-	349	.83	.38
Lookup- IE+	357	.82	.38
Frequency			
F1	362	.69	.47
F3	437	.88	.33

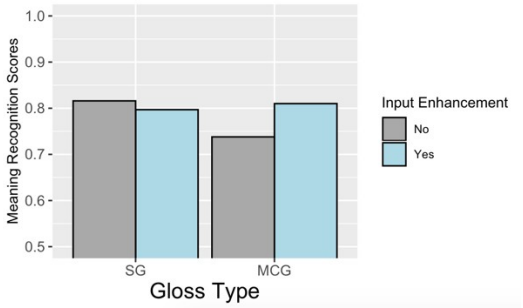


Figure 35. Effects of gloss type and IE on meaning recognition.

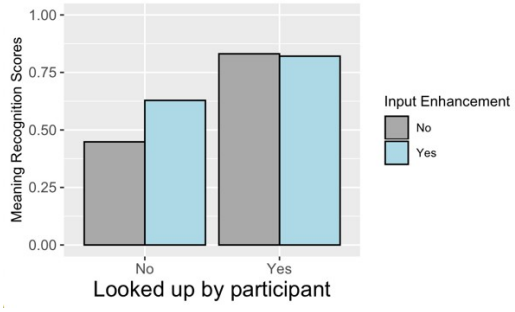


Figure 36. Effects of lookups and IE on meaning recognition.

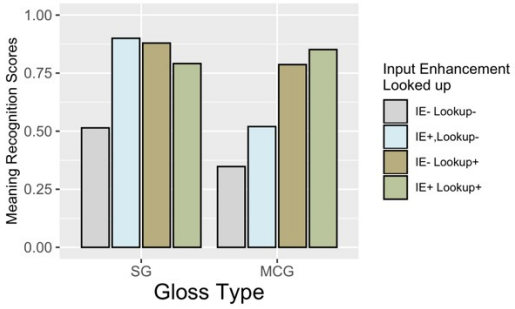


Figure 37. Interaction between experimental groups and learning modes.

5.4.5. RQ2.3 Are participants able to pay attention to all unknown target words?

The initial hypothesis of this study was that readers cannot focus on all unknown words because their cognitive resources are being spent elsewhere, such as on trying to understand the meaning of the text. Therefore, it was expected that participants would sometimes not fixate on unknown target words at all. However, only four such cases were found. Four participants skipped over one of two target words each. These words were both one syllable long (rite, clod), which might have contributed to this. For example, these words might have been read while the cursor was still resting on the previous word.

In an attempt to assess the quality of gaze durations, the WG longest gaze duration for each participant per each target word was calculated. This computation was required because participants would sometimes skim over an unknown word and then return to it later; therefore, first-pass reading was not a reliable measure of the quality of gazes.

The analysis based on the longest WG gaze duration for each target word showed that most participants fixated on most target words longer than their overall

average gaze duration. Only in less than 4% of cases was the longest gaze duration shorter. On the other hand, it was not unusual for participants to perpetually gaze at the word (or its gloss) for more than five times their average gaze duration. Figure 40 shows the distribution of the longest gaze durations. Parts of the histogram in red show gaze durations shorter than the participants' average. Figure 41 shows the breakdown of these below-average values and absolute total reading time values (ms) in the rectangles.

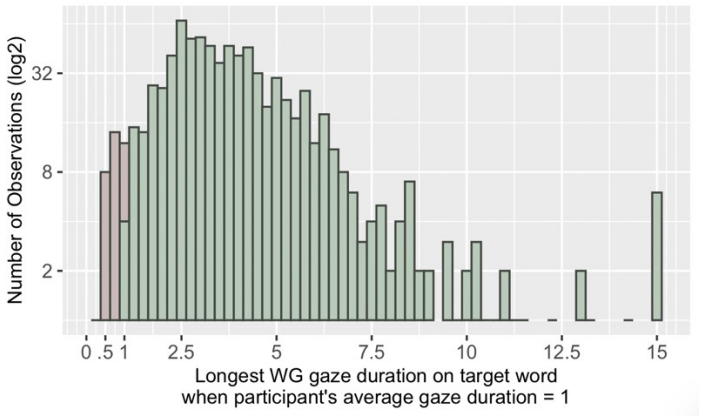


Figure 38. Distribution of relative values of the longest gaze duration.

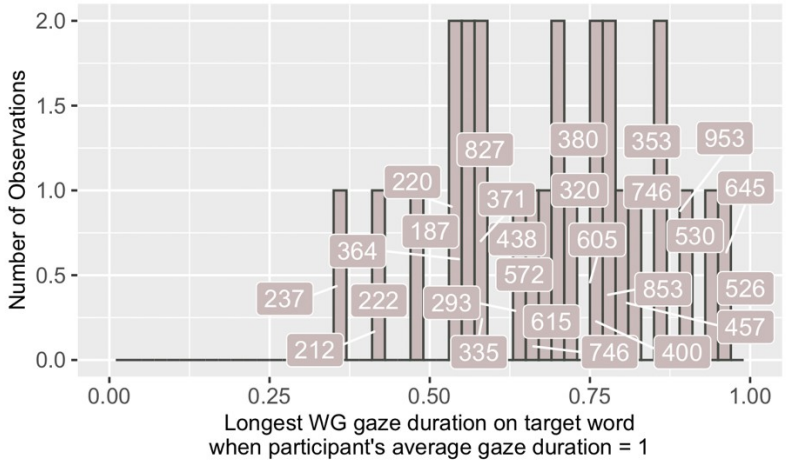


Figure 39. Breakdown of total reading times of words that were read faster than average.

From these results, we can conclude that participants did not have trouble fixating on new words in the majority of cases. On the contrary, these words attracted their attention more than regular words or target words they had previous knowledge of.

The effect of input enhancement on total reading times was also analyzed using a simple t-test. The analysis predicts no effect of input enhancement on the time participants spent gazing on words and their glosses (p = .979, d = 1.862). However, input enhancement had a shortening effect on the time participants spent looking at a word before clicking it (p = .002, d = 1.916).

5.4.6. Effects of attention and awareness on vocabulary acquisition

Although this was not one of the study's primary goals, the collection of gaze data and the form recognition test allowed the investigation of the relationship between these factors and the effects they had on vocabulary acquisition in a fashion similar to Godfroid and Schmidtke (2013). Since vocabulary acquisition was influenced by word frequency in a complicated way and analyzing the effect of gaze duration across three different occurrences would be very difficult, only the 10 F1 words were chosen for analysis.

A linear mixed effect model analysis was attempted to account for continuous variables such as the refixation count or reading time variables; however, every variable added or removed from the model changed the significance of other factors drastically. Due to the number of factors, including all of them was not possible in the analysis. Therefore, only descriptive results for refixation count are provided.

The number of refixations on a target word seems to have had a positive effect on form recognition. However, participants who refixed on a word more than

four times were less likely to recognize the word form in the post-test, as shown in Figure 42.

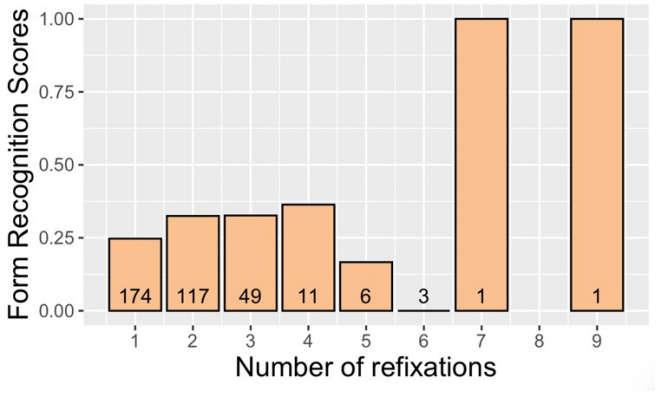


Figure 40. Relationship between refixation count of F1 words and form recognition scores (numbers inside bars note number of observations).

Awareness seemed to inform meaning acquisition in a major way. In the ANOVA analysis of meaning recall and recognition tests in Section 5.4.4, the factor considering whether the word was selected correctly in the form recognition test was also included. When a participant was aware of a word being used in a text, they were also four times more likely to recall its meaning ($p < .001$, $d = -.564$). For meaning recall, there was also an interaction between awareness and whether the word was clicked ($p = .005$). This relationship can be seen in Figure 43. In 727 lines of data, there were 391 instances of a word being clicked at least once but not being recognized in the form recognition test. Such words had less chance of having their meaning recalled than other words, including those not clicked by the participant. The frequency of the target word seemed to amplify the relationship between awareness and meaning recall, as shown in Figure 44 ($p < .001$).

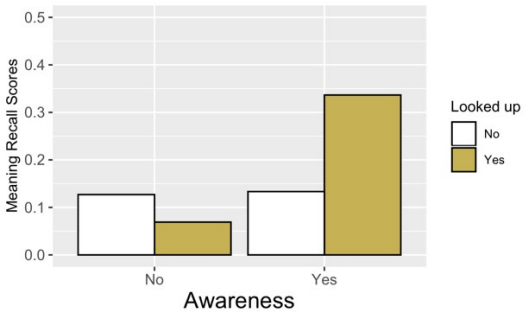


Figure 41. Effects of awareness and lookups on meaning recall.

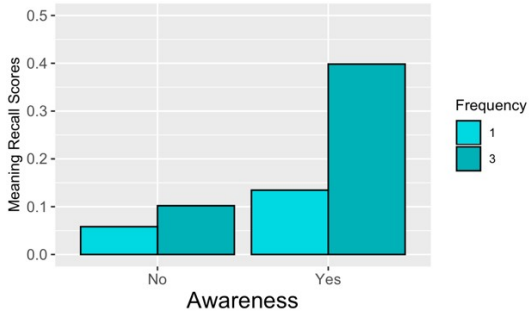


Figure 42. Effects of awareness and frequency on meaning recall.

Awareness was also predictive of meaning recognition (p = .007, d = .789) as words whose form was recognized were more likely to have their meaning known among the three presented options. According to the analysis, the effect of awareness was also slightly influenced by the gloss type (p = .003). However, the interaction of these two factors did not drastically influence the mean scores.

In the previous section, it was established that participants could pay attention to most unknown target words. This section also investigates how longer attention spans (reading times) correlate with vocabulary acquisition. Since large differences in reading times between looked-up and non-looked-up words were expected, only clicked words were included in this analysis. All reading time variables were divided into binary groups (low, high) and split by the median to perform an ANOVA with a maximum of two levels per factor. This method was chosen because it produced two equally sized groups for each factor. It also allowed for including outlying measurements without their value significantly influencing the statistics. The refixation count was also transformed into a binary variable (0 or 1). In the ANOVA, five attention factors (total reading time, total reading time adjusted, first gaze duration, first gaze duration adjusted, and refixation) and four other main factors

(learning mode, frequency, input enhancement, and gloss type) were employed without checking for interactions, to make the model simpler.

Analysis of form recognition scores showed the main effects of frequency and learning mode. The effect of frequency blocked any effect of attention as measured by reading times. This essentially means that frequency was a reliable predictor of form recognition, and differences in attention spans within words with the same frequency were negligible. The same result could be found for meaning recall, except that no main effect of learning mode was found. Only in the meaning recognition scores there was found, besides the effect of frequency, an effect of total reading time. There was no effect of adjusted reading times, which means that the longer participants spent looking at the gloss, the more they were likely to recognize the meaning in a post-test, and this effect was identified in both frequency conditions. Only significant variables are reported in tables 19 to 21. Figures 45 to 47 show the distribution of shorter and longer attention spans between F1 and F3 words and how they contribute (or do not contribute) to vocabulary acquisition. Note the ratio of observation counts (white numbers inside bars) drastically changes between F1 and F3, as the frequency was a strong predictor of attention ($p < .001$). However, analysis involving isolating F1 or F3 data and splitting attention spans by the median inside those groups did not produce the main effect of reading times for form recognition or meaning recall.

Table 19

ANOVA Effects of Frequency and Learning Mode on Form Recognition

Factor	*Df*	*Sum Sq*	*Mean Sq*	*F*	*p*		η^2
Frequency	1	12.72	12.719	4.671	.000	*	.029
Intentional Learning	1	1.07	1.069	4.671	.031	*	.007
Residuals	696	159.26	.229				

Table 20

ANOVA Effect of Frequency on Meaning Recall

Factor	*Df*	*Sum Sq*	*Mean Sq*	*F*	*p*		η^2
Frequency	1	7.40	7.401	51.780	.000	*	1.655
Residuals	696	99.48	.143				

Table 21

ANOVA Effects of Frequency and Attention on Meaning Recognition

Factor	*Df*	*Sum Sq*	*Mean Sq*	*F*	*p*		η^2
Frequency	1	6.09	6.093	44.851	.000	*	9.481
Total Reading Time	1	.55	.553	4.071	.044	*	.001
Residuals	696	159.26	.229				

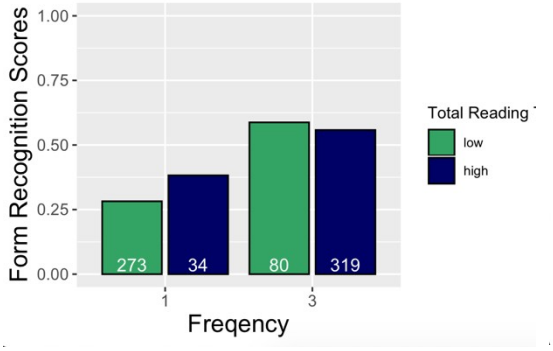


Figure 43. Effects of frequency on form recognition divided by attention span.

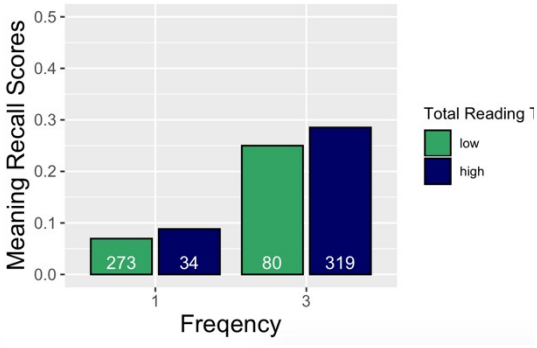


Figure 44. Effects of frequency on meaning recall divided by attention span.

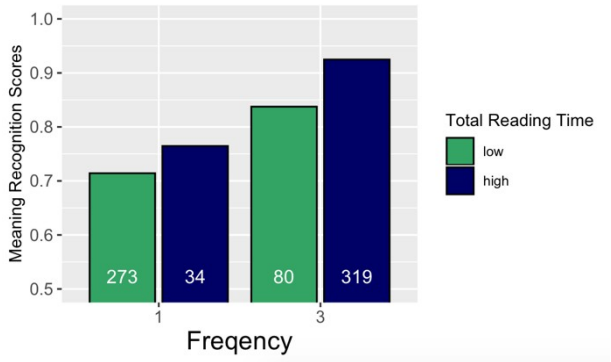


Figure 45. Effects of frequency on meaning recognition divided by attention span.

5.5. Discussion

5.5.1. The Effects of Input Enhancement

In the second experiment, participants' self-reports on their attempt or refusal to memorize words were added as one of the main factors. Different readers chose different strategies when reading a text. Some readers focus more on acquiring new vocabulary while some focus on understanding the message.

It is commonly accepted that announcing a vocabulary post-test will will trigger intentional vocabulary learning (Hulstijn, 2001). Since the initiative to learn words does not come from the reader, announcing a vocabulary test to the reader is considered an extrinsic factor.

In this study, none of the participants were told about a post-test, and it is assumed that they read the experiment material for comprehension. However, half of the participants read the text with the 20 target words highlighted. This form of input enhancement could be considered another extrinsic factor — since it brings the reader's attention to the target words — theoretically increasing the chance the reader will make an effort to memorize those words.

However, in Experiment 2, participants' decision to memorize words was not influenced by input enhancement. Therefore, it is likely that intrinsic factors, such as integrative motivation towards learning English (e.g., interest in English speaking countries' culture), or the relevance the reader attaches to those words might influence the reader's reading strategy more than input enhancement. The result of this analysis is in accordance with Peters et al. (2009), who found that the relevance of a target word to comprehension is a stronger predictor of vocabulary acquisition than the decision to announce or conceal a vocabulary test. The results also support the body of research that considers the effect of intrinsic motivation to be greater than the effect of extrinsic motivation (Lin, McKeachie, & Kim, 2003). For example, Masoudi (2017) found that students will learn new vocabulary more effectively if they can choose the words to learn themselves, instead of those words being chosen for them by input enhancement.

While input enhancement did not motivate all participants to memorize highlighted words, participants who had chosen to memorize words were influenced by the input enhancement. When input enhancement was present, the number of non-target words looked up by participants in the intentional learning group dropped significantly. It is unclear whether this group of participants anticipated a vocabulary test, but we can observe how input enhancement interacted with their selective strategy.

Participants in the SG groups clicked more target words when input enhancement was present. However, there was no difference in the number of clicked target words in the MCG groups. The pop-up dictionary was devised as a non-intrusive way to provide meanings for unknown words. Nevertheless, the involvement load of having to choose between multiple meanings might have obstructed the participants' smooth reading and might have led to abandoning the gloss in some instances. In the design experiment, the gloss disappeared when participants moved the cursor to another word (this was intended to prevent content overlap on smaller screens). Since participants could not see the context and the gloss simultaneously, searching and evaluating might have been difficult. However, only two participants who used MCG reported in the questionnaire that it was difficult for them to use the pop-up dictionary.

Input enhancement did not help participants identify the target words in the form recognition test. As Benati (2016) states, the effect of input enhancement on attention varies from study to study. Some studies confirmed the effect of input enhancement on noticing; however, these researches used different measures of noticing: choosing the correct spelling for a target word (Vu & Peters, 2020) or the

analysis of the participants' notetaking (Cho, 2010). LaBrozzi (2016) states that input enhancement had a positive effect on form recognition; however, the test conducted by LaBrozzi was actually a meaning recall test. Kim (2003) conducted a similar form recognition test to this study and found no effect of typographical input enhancement on noticing. Similarly, Leow (2001) found no effect of input enhancement on noticing or understanding Spanish imperative forms.

As input enhancement in this study was visual, it was expected that participants would pay greater attention to the form of target words and do better on a form recognition task. On the contrary, it seemed to bring the participant's awareness of the meaning of those words. In 12% of the cases, participants chose not to look up a word; they seemed to memorize the meaning of the words much better if it was visually enhanced. The results suggest that input enhancement might have motivated participants to guess the meaning from the context, if the word was highlighted. (However, there has not yet been an empirical study to the author's knowledge, which would tie input enhancement and increased rates of lexical inference.) Such choice on participants' part is also reflected in the meaning recall test, as participants who did not look up the visually enhanced words' meaning were more likely to recall the meaning than participants who looked up the meaning. Due to the cognitive load that guessing encompasses, it can be argued that the involvement load arising from meaning inference (guessing) is a strong factor for vocabulary acquisition.

Although input enhancement was also beneficial to recognize the meaning of words not looked up by the participant, words looked up still had a greater chance of being recognized by the participants. Since in the meaning recognition test, the target

word in English with three Japanese choices was presented, participants might have depended on them simply remembering one of the choices being shown in a gloss. To assess the extent to which the meaning of target words could be recognized, perhaps a meaning recognition test where participants have to choose the correct synonym (either in L1 or L2) for each target word would have been better.

The input enhancement did not influence how much attention participants paid to the words, as the total reading time was roughly the same for both groups. However, the adjusted total reading time (i.e., the time it took participants to decide to click the word) was lower for participants with input enhancement. This result suggests that input enhancement influenced the time it took participants to decide whether to click the word or not. Few participants with input enhancement reported that if a word appeared unfamiliar to them, they would click it without paying much attention to its form.

The words differed in their ability to attract a click from the participant. The factors of whether a word will be clicked can only be speculated here. For example, the word *hydrangea* did not play a key role in the story. However, it attracted 100% of participants who did not know this word, perhaps because of its form. Since the recall test was non-contextual, participants could only infer the meaning while reading. By the design of the experiment, it was not possible to determine whether participants attempted to infer the meaning of a word or if they simply decided to skip it.

5.5.2. The Effects of MCG

The gloss type alone did not seem to affect any of the post-test measures. Because the five translations shown in the MCG condition were mostly technically

correct (as they were often synonymous translations of the same word), this type of gloss was possibly insufficient to induce increased involvement load. However, participants who were motivated to memorize new words benefitted from MCG in form recognition and meaning recall tests. Learners possibly chose to focus on the form and its presented meanings, and in doing so, they accessed a deeper cognitive level (such as contemplating how the presented meanings relate to each other). Such a result would be in accordance with Ajideh, Rahimpout, Amini, and Farrohki (2013), who found that motivational involvement can have positive short-term effects on vocabulary acquisition alongside cognitive involvement.

There is also the possibility that MCG condition was causing a counter-facilitative type of cognitive load to learners who were reading without trying to learn new vocabulary. This kind of cognitive load could constrain their ability to memorize the meanings of the words. The positive effects of MCG with IE could then be explained by the IE, focusing the readers to target words without spending too much cognitive capacity on other words.

5.5.3. The Effects of Frequency

The frequency of target words proved to be a significant predictor of vocabulary acquisition in all three forms of vocabulary tests. This result was expected as previous studies showed a reliable effect of frequency on vocabulary. For example, Eckerth and Tavakoli (2012) confirmed that five times occurring words are more likely to be acquired than one-time occurring ones. Concurrently, they claim that the effect of frequency was not as strong as that of involvement load. This study found the main effect of frequency for all three vocabulary tests, but no main

effect of gloss type was found. It could be another sign that the MCG condition did not induce any significant amount of facilitative involvement load.

5.5.4. Readers' Attention to Unknown Words

Based on Experiment 1 and previous research claims about learners not being able to notice the words they want (Azari, 2012; Izumi, 2013), it was expected that unknown target words would often be skipped (i.e., learners not fixating on them). However, such cases were extremely exceptional in this experiment. One could argue that skipping occurred scarcely because of the special design of this experiment (reading text with a mask obscuring most of the text). However, skipping non-target words was common in this study. After excluding gaze durations shorter than 34ms, it was found that participants only fixated on 895.42 (*SD* = 119.59) words in a 1036-word long text. Therefore, some 140 token words were not fixated on the average.

As participants focused on target words for more than 50 ms in the vast majority of cases, it is assumed that these words were detected, based on the claim that "any word that is fixated on for 50 ms or more is in working memory" (Godfroid & Schmidtke, 2013). If participants were aware of most words, the question of the acquisition lies not in noticing but in the depth of processing and whether they intended to memorize the word.

5.5.5. Attention, Awareness, and Vocabulary Acquisition

As stated in the previous section, participants were aware of almost all target words since they focused on them for longer than 50 ms. In over 96% of cases, unknown target words were being focused on longer than the participant's average

gaze duration length, which should give us the image of how much attention was paid to the unknown target words.

However, participants could only recognize around 43.18% (SD = 49.57%) percent of words. As (S. Song, 2007) stated, form recognition, as an offline test, is influenced by the participants' memory loss. Form recognition is, therefore, not the most accurate tool to measure noticing. Although longer total reading times seemed to produce higher form recognition rates, distinct variation in the data caused the results to lack statistical significance; therefore, the findings of Godfroid and Schmidtke (2013) could not be replicated. Only in the meaning recall test, was a statistically significant influence of total reading times detected. It is quite possible that participants who spent a longer time gazing at the gloss were likely to identify the Japanese gloss as one of the options in the recognition test. Therefore, it is uncertain whether meaning recognition score in an experiment using glossed texts would produce an accurate image of vocabulary acquisition.

6. General Discussion

The principal themes of this study was the pop-up dictionary and how it benefits vocabulary acquisition. The first experiment was designed to show whether giving learners a choice to look up any word will contribute to their comprehension or vocabulary acquisition. In accordance with previous research (Liu & Lin, 2011; Mekheimer, 2018), no effect on reading comprehension was found. Target words in this experiment were more likely to have their meaning recalled in a non-context post-test if the participant had freely chosen to look them up (without the influence of input enhancement) compared with participants who looked them up by input enhancement conditions.

Although participants with pop-up dictionary looked up much fewer target words, their score on a non-context post-test was comparable with the gloss condition. Therefore, the results of Experiment 1 suggest that the pop-up dictionary is a superior tool compared to gloss, because it allows learners to look up words they believe to be relevant.

However, Experiment 1 was not enough to determine whether the decision to look up or not look up words was a conscious decision or if the learners did not look them up because of their lack of attentional resources. In Experiment 2, offline and online measures of noticing were employed to assess whether participants were able to notice all target words.

Experiment 2 showed that participants could pay attention to most unknown target words and that attention was usually above their attentional average. In Experiment 2, participants looked up more target words than in Experiment 1, even when they were not prompted by input enhancement. Participants showing more

interest in target words could have been related to (a) the target words being more related to the text (see Hulstijn, 1993), (b) different percentages of unknown words between the two texts (Jones, 1995), (c) different readability scores of the two texts or (d) the fact that participants read the text in small chunks in Experiment 2 as Hulstijn et al., (1996) point out that learners are more likely to abandon dictionary use in texts longer than few hundred words.

Experiment 2 was also designed to answer how the use strategy of pop-up dictionary would change if the PD group also read the text with visual enhancement for target words. Although participants with input enhancement looked up more target words by a statistically significant margin, the number of target words looked up in groups without input enhancement was also considerably high. It would seem that in Experiment 2, the target words matched words that groups without IE felt compelled to look up. Perhaps the reason why IE groups outperformed groups without IE in a meaning recall test was that the target words were meaningful to them, and therefore, the extrinsic motivation of IE and intrinsic motivation of participants were combined.

The results of Experiment 2 showed that participants were sometimes trying to infer the new words' meanings from context, without depending on the gloss, and this method showed to be the most beneficial for vocabulary acquisition. Lexical inference also took place in Experiment 1 and contributed to a very weak form-meaning connection, which could be reactivated for meaning recall when context was present.

Therefore, it was concluded that learners could pay attention to all unknown words. However, future research is needed to see whether their attention would be

limited in more difficult texts or texts with a higher percentage of unknown words. At the very least, learners paid attention to most unknown words, and consciously decided whether to look them up or not in Experiment 2.

The pop-up dictionary does not only have to serve vocabulary acquisition. Perhaps it could be used as a tool for better comprehension of texts containing unknown words. The pop-up dictionary could be useful for reading foreign materials, such as pdf documents or websites, without an intention to become more proficient in English. Both Experiment 1 and 2 examined the participants' comprehension in a multiple-choice test. However, neither gloss type, the presence of input enhancement, or the number of pop-up dictionary use had any significant effect on comprehension. In terms of Experiment 1, this effectively means that the pop-up dictionary was not superior (or inferior) to a glossed text regarding reading comprehension, which could be influenced by the fact that comprehension questions were designed to avoid the reactivation of target words. Perhaps a future study could explore how looking up a target word with a pop-up dictionary can improve comprehension of concepts or events related to the target word. In Experiment 2, participants who self-reported trying to memorize new words did not score lower on comprehension test, which means there was no trade-off between comprehension and vocabulary acquisition in Experiment 2.

Table 22 shows the cognitive load of each group from Experiments 1 and 2 as measured by the technique feature analysis. Only the main aspects of motivation and noticing are shown because they are solely relevant to the differences between groups. The aspect of retrieval is possibly related to the difference between words appearing only once (F1) and words appearing three times (F3) words but does not

differ between groups. F3 words were consistently more likely to be recalled than F1 words, and such results are in accordance with the technique feature analysis. Differences between groups are, however, difficult to explain with the feature technique analysis. Whether learners choose the words themselves is an important factor in word learning (Hulstijn, 1993; Masoudi, 2017) and is also one aspect of the technique feature analysis. Although participants could choose words to learn in this study, participants were tested on the knowledge of words selected by the author. Another aspect that did not have the expected result was the negotiation of meaning. Based on the technique feature analysis of reading with a dictionary in Alahmadi and Foltz (2020) and some of the participants' reports in the questionnaire, the MCG condition of Experiment 2 was evaluated to have negotiation of meaning. For the MCG IE- group, the positive effect of negotiation could not be confirmed. Based on the growing body of research supporting the validity of technique feature analysis, it is possible that the MCG group did not necessarily have increased meaning negotiation levels. Although some participants reported difficulty choosing the correct meaning, it is unclear whether this was also the case for target words.

Table 22

Assessment of All Experimental Groups Based on Technique Feature Analysis

	Experiment 1		Experiment 2			
Criteria	G	PD	SG IE-	SG IE+	MCG IE-	MCG IE+
Motivation						
Is there a clear vocabulary learning goal?	0	0	0	0	0	0
Does the activity motivate learning?	0	0	0	0	0	0
Do the learners select the words?	0	1	1	1	1	1
Noticing						
Does the activity focus attention on target words?	1	0	0	1	0	1
Does the activity raise awareness of new vocabulary learning?	0	0	0	0	0	0
Does the activity involve negotiation?	0	0	0	0	1	1
Total Score	1	1	1	2	2	3
Non-Context Meaning Recall Score	.53	.58	.19	.19	.11	.24

Note. Experiment 1 scores were measured on a 2-point scale. Therefore, .53 and .58 refer to the accuracy of 26.5% and 29%, respectively.

7. Conclusion

7.1. Overview

This study provided important information about the effectiveness of pop-up dictionary for vocabulary acquisition. Firstly, pop-up dictionaries should not be condemned for being easy to use because it can still contribute to vocabulary acquisition, at least on the same level as static single gloss. However, the two experiments provided mixed results regarding whether it could be more effective than lexical inference. Further research is in place to examine how it compares to other forms of vocabulary acquisition and how the effect of pop-up dictionaries can be increased.

Secondly, the effectiveness of the gloss was deeply connected to the intrinsic motivation of learners and the attractiveness of target words. Input enhancement produced more lookups but only contributed to acquisition in situations where the reader would be compelled to look up the word even without IE.

Another finding was that it is theoretically possible for learners to pay attention to all unknown words, regardless of whether they try to memorize these words or not. It is acknowledged that an appropriate choice of materials might be necessary for a consistent attention level towards new words.

Lastly, although the pop-up dictionary did not contribute to reading comprehension any more than a simple hypertext gloss, it did not seem to hinder comprehension, and there was no trade-off between the two. This result suggests the possible multifold uses of dictionary gloss, such as reading for comprehension or vocabulary acquisition.

7.2. Theoretical Implications

This study supports the claim by Bruton et al. (2011) that incidental vocabulary acquisition, defined as learning new words without the intention to memorize them (Barcroft, 2004), cannot be maintained in an experimental setting because participants may be intrinsically motivated to memorize new words even when they are not instructed to do so. This study has also shown that the presence of an intention to learn new words may significantly interact with the study elements such as the cognitive load of each respective task or input enhancement. Therefore, it is always advised to ask participants whether they were making an effort to memorize words and either include this factor in the analysis or maintain the definition of incidental vocabulary acquisition such as in (Barcroft, 2004) and exclude the data from participants who reported trying to memorize new words.

Reflecting on previous research on pop-up dictionaries, participants' opinions of the pop-up dictionary were overwhelmingly positive. However, some participants voiced concerns that a tool that is as easy to use as pop-up dictionary might demotivate readers from memorizing new words. Due to the high probability that pop-up dictionaries such as the one employed in this study will get more popular with English learning material creators and learners themselves, more studies about pop-up dictionaries and how they influence vocabulary acquisition or reading comprehension should be conducted in the future.

7.3. Pedagogical Implications

In activities where incidental vocabulary learning is expected, such as reading for comprehension or extensive reading activity, teachers should understand that the

effectiveness of pop-up dictionaries and its use strategy, chosen by each student, will differ according to the learners' motivation.

For example, it should not be expected that just because using a pop-up dictionary is easier than using a paper or electronic dictionary, learners will use it to look up the meaning of all unknown words in a text. Some students will choose to infer the unknown word's meaning from the contextual or morphological clues. This method of learning new words is advised over the use of a pop-up dictionary because the cognitive involvement of guessing the meaning of a word might contribute to memorizing the new word. On the other side of the spectrum, learners may sometimes use it to confirm the meaning of previously known words. The nature of a pop-up dictionary being hidden behind a mouse click might be one of its strengths because it gives the learner freedom of choice whether to display the gloss or not. Even in electronic materials, where a pop-up dictionary is not a viable option, it is ideal to hide glosses behind a hypertext to provide this freedom to learners.

The learners' attention to important words can be increased with input enhancement. However, it should be noted that some participants will become so influenced by the presence of input enhancement that they will rarely look up other words.

Developers of pop-up dictionaries must consider how much information to display with each click and how this information should be organized. While learners who do not read to learn new words benefit more from seeing a simple context-fitted translation, more motivated learners might learn more vocabulary with a more complex gloss. It is also advised to provide learners with correct translations and information about etymology, related words (e.g., synonyms and antonyms), or

collocations. Perhaps a pop-up dictionary could be developed so that highly motivated learners may choose to display more information with a click of a button, or this information might be visually separated from the translations so as not to distract all learners from reading.

A function that should not be overlooked in developing a pop-up dictionary is looking up the meanings of collocations in the text. A highly sophisticated pop-up dictionary should parse the whole sentence and decide whether the word being looked up is a part of collocation, such as an idiom or a phrasal verb. A simpler version of a pop-up dictionary could display the most common collocations for the word being looked up.

Input flooding is always advised to bring the learners' attention to important words and help them acquire their meanings. As previous studies show, a task might overshadow input flooding with a high cognitive load or promote the effect of such a task. (Eckerth & Tavakoli, 2012). When the materials allow it, their creators should increase the frequency of important words as high as possible to promote learner acquisition.

7.4. Limitations of This Study

Numerous issues were overlooked or could not have been addressed in this study.

Experiment 2 was based on Experiment 1; however, the materials used in both experiments were different. Since there was a possibility that participants were ignoring target words simply because they did not feel the need to know their meaning, a pilot study comparing few paragraphs from the reading material for Experiment 2 and a candidate material (*The Midas Touch*) for Experiment 1 was

conducted. It confirmed that *The Midas Touch* contains more words that readers would like to know the meaning. The results of Experiment 2 also confirm this, because even without IE, the participants looked up 88% of unknown target words (or 77% of all target words) on average, while in Experiment 1, participants without IE only looked up about 50% of all unknown target words.

This study provided an important insight into the interplay of various features such as IE, gloss type, or the pop-up dictionary. However, limited time and the number of participants only permitted examining these features along with a pop-up dictionary; therefore, an analysis comparing pop-up dictionaries with traditional static gloss or reading with no gloss at all was not possible. Although it is clear from the results of this study that pop-up dictionaries can lead to vocabulary acquisition, there is still a need for research comparing it to other modes of reading.

The self-reported mode of vocabulary learning (incidental vs. intentional) was added to the analysis as it proved to be a significant factor. However, dividing the participants in this manner and the IE and gloss type group distinction resulted in small sample sizes. Perhaps the results of the analysis would be more comprehensive with a larger sample of learners. A larger number of participants could also suggest comparing the results based on each participant's English proficiency.

It should also be noted that IE and gloss type were between-subject variables that could be conducted as within-subject variables in future studies to account for individual differences between participants' attention and aptitude for vocabulary learning.

The software used to measure participants' attention in this experiment was very limited in that it could only record where the participants' cursor had stopped

and the duration of idleness. While it was possible to record each pass duration and total reading time, the software could not measure the length of individual fixations, as the cursor was not dependent on them. Furthermore, no clear distinction could have been made between how much time the participant spent fixating on a word and how long they were reading the gloss. These issues should be addressed in a future research, which will employ a proper eye-tracking camera.

While this study showed that participants could pay attention to all target words and the decision whether to look them up or not is somewhat intentional, it could not address the target words' intrinsic factors and the reasons for different words receiving a different amount of attention or being clicked by a different number of participants. Further studies should be conducted, where the properties of the target words such as length or relevance to the context are compared to provide more insight into how learners distribute their attention between words in a text.

Furthermore, it was concluded that participants could pay attention to all unknown words in a text. However, this study did not consider learners with ADHD or other disorder that affects learning. Future studies related to attention to new words should address the difference in attention given by neurodiverse students.

This study compared the effects of different gloss types based on the involvement load hypothesis (Laufer & Hulstijn, 2001) and the feature technique analysis. However, no such comparison could be achieved, as the MCG condition did not seem to produce a significant amount of cognitive load for most of its users. Perhaps a pop-up dictionary, where the gloss contains incorrect options for the reader to evaluate against the context, could be devised in the future, allowing learners to increase the task involvement for words of interest. As a pop-up dictionary is only

powered in an environment such as a PC or a tablet, feedback could prevent the learners from memorizing an incorrect meaning from the gloss.

List of References

Ajideh, P., Rahimpour, M., Amini, D., & Farrohki, F. (2013). Motivational strategies, task effectiveness and incidental acquisition of second language vocabulary. *Journal of Language Teaching and Research*, *4*, 1044–1052.

Alahmadi, A., & Foltz, A. (2020). Exploring the effect of lexical inferencing and dictionary consultation on undergraduate EFL students' vocabulary acquisition. *PLOS ONE*, *15.*

Alessi, S., & Dwyer, A. (2008). Vocabulary assistance before and during reading. *Reading in a Foreign Language*, *20*, 246–263.

Alharbi, M. A. (2016). Using different types of dictionaries for improving EFL reading comprehension and vocabulary learning. *JALT CALL Journal*, *12*, 123–149.

Committee on Quality Improvement, American Academy of Pediatrics. (2000). Clinical Practice Guideline: Diagnosis and evaluation of the child withattention-deficit/hyperactivity disorder. *Pediatrics*, *105*, 1158–1170.

Arani, S. G., & Yazdanimoghaddam, M. (2016). The Impact of Input Flooding and Textual Enhancement on Iranian EFL Learners' Syntactic Development. *Applied Linguistics*, *16*, 13.

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*, 502–508.

Azari, F. (2012). Review of effects of textual glosses on incidental vocabulary learning. *International Journal of Innovative Ideas*, *12*, 13–24.

Baars, B. J., & Franklin, S. (2007). An architectural model of conscious and unconscious brain functions: Global Workspace Theory and IDA. *Neural networks*, *20*, 955–961.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation, 8,* 47–89.

Barcroft, J. (2004). Second Language Vocabulary Acquisition: A Lexical Input Processing Approach. *Foreign Language Annals*, *37*, 200–208.

Beal, V. (2009). *The weight of Involvement Load in college level reading and vocabulary tasks.* Library and Archives Canada = Bibliothèque et Archives Canada, Ottawa.

Benati, A. (2016). Input manipulation, enhancement and processing: Theoretical views and empirical research. *Studies in Second Language Learning and Teaching*, *6*, 65–88.

Bensoussan, M., & Laufer, B. (1984). Lexical Guessing in Context in EFL Reading Comprehension. *Journal of Research in Reading*, *7*, 15–32.

Bergleithner, J. M. (2007). *Working memory capacity, noticing, and L2 speech production*. Doctoral dissertation, Universidade Federal de Santa Catarina.

Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context Variation and Definitions in Learning the Meanings of Words: An Instance-Based Learning Approach. *Discourse Processes*, *45*, 122–159.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, *20*, 28.

Brusnighan, S. M., & Folk, J. (2012). Combining Contextual and Morphemic Cues Is Beneficial During Incidental Vocabulary Acquisition: Semantic Transparency in Novel Compound Word Processing. *Reading Research Quarterly*, *47*, 172–190.

Bruton, A., López, M. G., & Mesa, R. E. (2011). Incidental L2 Vocabulary Learning: An Impracticable Term? *TESOL Quarterly*, *45*, 759–768.

Chaudron, C. (1985). Intake: On Models and Methods for Discovering Learners' Processing of Input. *Studies in Second Language Acquisition*, *7*, 1–14.

Chen, X. (2013). *Chinese EFL learners' noticing of recasts: Its relation to target structures, uptake, and working memory capacity*. Michigan State University, Second Language Studies.

Cho, M. Y. (2010). The effects of input enhancement and written recall on noticing and acquisition. *Innovation in Language Learning and Teaching*, *4*, 71–87.

Chun, D. M., & Scott Payne, J. (2004). What makes students click: working memory and look-up behavior. *System*, *32*, 481–503.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*, 201–215.

Cowan, N. (2010). The Magical Mystery Four: How is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science*, *19*, 51–57.

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.

De Ridder, I. (2002). Visible Or Invisible Links: Does the Highlighting of Hyperlinks Affect Incidental Vocabulary Learning, Text Comprehension, and the Reading Process? *Language Learning*, *24*.

Eckerth, J., & Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research*, *16*, 227–252.

Educational Testing Service, Princeton, New Jersey, U.S.A.

Fatalaki, J. A. (2014). Involvement Load Hypothesis: Word Meaning Retention across Oral and Written Task Types. *International Letters of Social and Humanistic Sciences*, *37*, 29–45.

Fichtner, F., & Barcroft, J. (2019). The Input-Based Incremental Approach to German Vocabulary: A Sample Lesson Based on the Film Barbara. *Die Unterrichtspraxis/Teaching German*, *52*, 82–102.

Fraser, C. A. (1999). Lexical Processing Strategy Use and Vocabulary Learning Through Reading. *Studies in Second Language Acquisition*, *21*, 225–241.

Frishkoff, G. A., Collins-Thompson, K., Hodges, L., & Crossley, S. (2016). Accuracy feedback improves word learning from context: Evidence from a meaning-generation task. *Reading and Writing*, *29*, 609–632.

Fukkink, R. G., Blok, H., & Glopper, K. de. (2001). Deriving Word Meaning from Written Context: A Multicomponential Skill. *Language Learning*, *51*, 477–496.

Gass, S., Ildikó, S., & Lemelin, S. (2003). Differential Effects of Attention. *Language Learning*, *53*, 497–545.

Gershon, J. (2002). Gender differences in ADHD. *The ADHD Report*, *10*, 8–16.

Godfroid, A., Boers, F., & Housen, A. (2013). An Eye for Words: Gauging the Role of Attention in Incidental L2 Vocabulary Acquisition by Means of Eye-Tracking. *Studies in Second Language Acquisition*, *35*, 483–517.

Godfroid, A., & Schmidtke, J. (2013). What Do Eye Movements Tell Us About Awareness? A Triangulation of Eye-Movement Data, Verbal Reports, and Vocabulary Learning Scores. In *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 183–205). National Foreign Language Resource Center.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*, 404.

Gu, Y., & Johnson, R. K. (1996). Vocabulary Learning Strategies and Language Learning Outcomes. *Language Learning*, *46*, 643–679.

Hamada, M. (2014). The Role of Morphological and Contextual Information in L2 Lexical Inference. *The Modern Language Journal*, *98*, 992–1005.

HAMADA, A. (2015). Improving incidental L2 vocabulary learning with latent semantic analysis. *ARELE: annual review of English language education in Japan*, *26*, 61-76.

Harrington, M., & Sawyer, M. (1992). L2 Working Memory Capacity and L2 Reading Skill. *Studies in Second Language Acquisition*, *14*, 25–38.

Hasegawa, Y. (2012). Incidental Vocabulary Learning and Memory of Context: interaction between informativeness and imageability. *ARELE: Annual Review of English Language Education in Japan*, *23*, 201–216.

Hasegawa, Y. (2013). Immediate or delayed effect of context: Japanese learners' vocabulary learning using translation. *Opening New Lines of Communication in Applied Linguistics*, 195.

Horst, M. (2005). Learning L2 Vocabulary through Extensive Reading: A Measurement Study. *Canadian Modern Language Review*, *61*, 355–382.

H Hu, H. C. M., & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement load hypothesis versus technique feature analysis. *System*, *56*, 28-39.

Hu, H. C. M., & Nation, P. (2000). Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language*, *13*, 403–430.

Huckin, T., & Bloch, J. (1993). Strategies for inferring word-meanings in context: A cognitive model. *Second Language Reading and Vocabulary Learning*, *153*, 178.

Huckin, T. & Coady, J. (1999). Incidental Vocabulary Acquisition in a Second Language. *Studies in Second Language Acquisition*, *21*, 181–193.

Huckin, T., Haynes, M., & Coady, J. (1993). *Second Language Reading and Vocabulary Learning*. Norwood, N.J.: Ablex Publishing Corporation.

Hulstijn, J. H. (1993). When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *The modern language journal*, *77*, 139-147.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 258–286). Cambridge University Press.

Hulstijn, J. H. (2003). Incidental and Intentional Learning. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 349–381). Oxford, UK: Blackwell Publishing Ltd.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental Vocabulary Learning by Advanced Foreign Language Students: The Influence of Marginal Glosses, Dictionary Use, and Reoccurrence of Unknown Words. *The Modern Language Journal*, *80*, 327–339.

Izumi, S., & Bigelow, M. (2000). Does Output Promote Noticing and Second Language Acquisition? *TESOL Quarterly*, *34*, 239–278.

Jacobs, G. M., Dufon, P., & Hong, F. C. (1994). L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading*, *17*, 19–28.

Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 824.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*, 47–77.

Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences*, *103*, 17048–17052.

Jones, F. R. (1995). Learning an alien lexicon : a teach-yourself case study. *Second Language Research*, *11*, 95–111.

Kadota, S., Noro, T., Shiki, O., & Hase, N. (2014). *Eitango un'yo-ryoku hantei soft o tsukatta goi-shido* [English vocabulary instruction using lexical proficiency assessment software]. Tokyo, Japan: Taishukan.

Kasahara, K. (2004). Threshold Vocabulary Size for Incidental Vocabulary Learning by MC Glosses. *HELES Journal*, *4*, 39–52.

Kim, Y. (2003). *Effects of input elaboration and enhancement on second language vocabulary acquisition through reading by Korean learners of English* (Doctoral dissertation, University of Hawaii at Manoa).

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*, 285–299.

LaBrozzi, R. M. (2016). The effects of textual enhancement type on L2 form recognition and reading comprehension in Spanish. *Language Teaching Research*, *20*, 75–91.

Laufer, B., & Goldstein, Z. (2004). Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness. *Language Learning*, *54*, 399–436.

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

Leow, R. P. (2001). Do Learners Notice Enhanced Forms while Interacting with the L2?: An Online and Offline Study of the Role of Written Input Enhancement in L2 Reading. *Hispania*, *84*, 496.

Leung, J. H. C., & Williams, J. N. (2012). Constraints on Implicit Learning of Grammatical Form-Meaning Connections. *Language Learning*, *62*, 634–662.

Levelt, W. J. (1993). Lexical access in speech production. In *Knowledge and language* (pp. 241–251). Springer.

Lin, Y.-G., McKeachie, W. J., & Kim, Y. C. (2003). College student intrinsic and/or extrinsic motivation and learning. *Learning and Individual Differences*, 13, 251–258.

Liu, T.-C., & Lin, P.-H. (2011). What comes with technological convenience? Exploring the behaviors and performances of learning with computer-mediated dictionaries. *Computers in Human Behavior*, *27*, 373–383.

Loewen, S., & Inceoglu, S. (2016). The effectiveness of visual input enhancement on the noticing and L2 development of the Spanish past tense. *Studies in Second Language Learning and Teaching*, *6*, 89–110.

Marefat, H., Ali Rezaee, A., & Naserieh, F. (2016). Effect of Computerized Gloss Presentation Format on Reading Comprehension: A Cognitive Load Perspective. *Journal of Information Technology Education: Research*, *15*, 479–501.

Masoudi, G. (2017). The Effect of Vocabulary Self -Selection Strategy and Input Enhancement Strategy on the Vocabulary Knowledge of Iranian EFL Learners. *English Language Teaching*, *10*, 32.

McKeown, M. G. (1993). Creating Effective Definitions for Young Word Learners. *Reading Research Quarterly*, *28*, 16.

McVay, J. C., & Kane, M. J. (2012). Why Does Working Memory Capacity Predict Variation in Reading Comprehension? On the Influence of Mind Wandering and Executive Attention. *Journal of Experimental Psychology. General*, *141*, 302–320.

Meganathan, P., Thai Yap, N., Paramasivam, S., & Jalaluddin, I. (2019). Incidental and Intentional Learning of Vocabulary among Young ESL Learners. 3L The Southeast Asian Journal of English Language Studies, *25*, 51–67.

Mekheimer, M. A. (2018). Effects of e-dictionaries on reading comprehension and vocabulary learning in EFL college students: a re-examination. *Journal of the Faculty of Education at Beni Suef*, *1*, 413-464.

Miller, G. A. (1956). The magic number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 91–97.

Ministry of Eucation, Culture, Sports, Science and Technology, Japan. (2016). 2020-nendai ni muketa kyouiku no jouhouka ni kansuru kondankai - saishuu matome - kyouiku no jouhouka kasoku puran [Conference on Education Digitalization for 2020's - Final Summary - A Plan for Acceleration of Education Digitalization]. Retrieved from https://www.mext.go.jp/b_menu/shingi/chousa/shotou/110/shiryo/__icsFiles/afieldfile/2016/11/07/1378984_8.pdf

Mitarai, Y., & Aizawa, K. (1999). The effects of different types of glosses in vocabulary learning and reading comprehension. *ARELE: Annual Review of English Language Education in Japan*, *10*, 73–82.

Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: an exploratory study. *System*, *28*, 291–304.

Mondria, J.-A. (2003). The Effects of Inferring, Verifying, And Memorizing on the Retention of L2 Word Meanings: An Experimental Comparison of the "Meaning-Inferred Method" and the "Meaning-Given Method". *Studies in Second Language Acquisition*, *25*, 473–499.

Movahedi, M., & Shourkaee, H. Z. (2018). *Effects of a Pop-up Dictionary Interface on Students' Vocabulary Retention and Attitudes towards Reading Literary Works*. *9*.

Nakagawa, C. (2006). The Effects of Morphological and Contextual Clues on EFL Readers' Lexical Inference. *ARELE: Annual Review of English Language Education in Japan*, *17*, 151–160.

Namaziandost, E., Rezvani, E., & Polemikou, A. (2020). The impacts of visual input enhancement, semantic input enhancement, and inputflooding on L2 vocabulary among Iranian intermediate EFL learners. *Cogent Education*, *7*.

Nassaji, H. (2003). Higher–Level and Lower–Level Text Processing Skills in Advanced ESL Reading Comprehension. *The Modern Language Journal*, *87*, 261–276.

Nation, I. S. (2013). *Teaching & learning vocabulary*. Boston: Heinle Cengage Learning.

Nation, I. S. P., & Webb, S. A. (2011). *Researching and analyzing vocabulary* (1st ed). Heinle, Cengage Learning.

Oxford, R. L., & Scarcella, R. C. (1994). Second language vocabulary learning among adults: State of the art in vocabulary instruction. *System*, *22*, 231–243.

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory & Cognition*, *29*, 344–354.

Park, S., Sungmook Choi, & Lee, M. (2012). Visual Input Enhancement, Attention, Grammar Learning, & Reading Comprehension: An Eye Movement Study. *English Teaching*, *67*, 241–265.

Pellicer□Sánchez, A., Conklin, K., & Vilkaitė□Lozdienė, L. (2020). The Effect of Pre-reading Instruction on Vocabulary Learning: An Investigation of L1 and L2 Readers' Eye Movements. *Language Learning*, *0*. 1–42.

Petchko, K. (2011). Input enhancement, noticing, and incidental vocabulary acquisition. *The Asian EFL Journal Quarterly*, *3*, 228–255.

Peters, E. (2012). Learning German formulaic sequences: The effect of two attention-drawing techniques. *The Language Learning Journal*, *40*, 65–79.

Peters, E., Hulstijn, J. H., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German Vocabulary Through Reading: The Effect of Three Enhancement Techniques Compared. *Language Learning*, *59*, 113–151.

Popov, V., & Reder, L. (2020). *Frequency effects in recognition and recall* [Preprint]. PsyArXiv.

Prichard, C. (2008). Evaluating L2 readers' vocabulary strategies and dictionary use. *Reading in a Foreign Language*, *20*, 216–231.

Qian, D. D. (2002). Investigating the Relationship Between Vocabulary Knowledge and Academic Reading Performance: An Assessment Perspective. *Language Learning*, *52*, 513–536.

Rashtchi, M., & Yousefi, L. M. (2017). Reading input flooding versus listening input flooding: Can they boost speaking skill? *Journal of Language and Cultural Education*, *5*, 39–58.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*, 372.

Rayner, K. (2009). Eye Movements in Reading: Models and Data. *Journal of Eye Movement Research*, *2*, 1–10.

Robinson, P. (2003). Attention and Memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 631–678). Blackwell Publishing Ltd.

Roby, W. B. (1999). " What's in a gloss?": A commentary on Lara L. Lomicka's" To gloss or not to gloss": An investigation of reading compreension online. *Language Learning & Technology*, *2*, 94–101.

Rott, S. (2005). Processing glosses: A qualitative exploration of how form-meaning connections are established and strengthened. *Reading in a Foreign Language*, *17*, 30.

Rott, S. (2007). The Effect of Frequency of Input-Enhancements on Word Learning and Text Comprehension. *Language Learning*, *57*, 165–199.

Rott, S. J. Williams (2003). Making-form-meaning connections while reading: a qualitative analysis of word processing. *Reading in a Foreign Language*, *15*, 45–75.

Sabet, M., Farhoumand, F., Zafarghandi, A., & Naseh, A. (2014). The Effect of Focus Strategies on ADHD Students' English Vocabulary Learning in Junior

High School. *International Journal of Applied Linguistics & English Literature*, *4*.

Sagarra, N., & Abbuhl, R. (2013). Optimizing the Noticing of Recasts via Computer-Delivered Feedback: Evidence That Oral Input Enhancement and Working Memory Help Second Language Learning. *The Modern Language Journal*, *97*, 196–216.

Sánchez Gutiérrez, C. H., Serrano, M. P., & García, P. R. (2019). The effects of word frequency and typographical enhancement on incidental vocabulary learning in reading. *Journal of Spanish Language Teaching*, *6*, 14–31.

Schmidt, R. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, *11*, 129–158.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *Consciousness in second language learning*, *11*, 237-326.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. *Attention and Awareness in Foreign Language Learning*, *9*, 1–63.

Schmidt, R. (2012). Attention, awareness, and individual differences in language learning. *Perspectives on Individual Characteristics and Foreign Language Education*, *6*, 27.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary rresearch manual*. Basingstoke, UK: Palgrave Macmillan.

Simard, D., Foucambert, D., Bergsleithner, J. M., Frota, S. N., & Yoshioka, J. K. (2013). Observing noticing while reading in L2. *Noticing and second language acquisition: Studies in honor of Richard Schmidt*, 207–226.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 1558–1568.

Song, S. (2007). Beginning ESL Learners' Noticing of Morphological and Syntactic Changes in Recasts. *Applied Linguistics*, *7*, 25.

Tabata-Sandom, M. (2016). How do learners of Japanese read texts when they use online pop-up dictionaries. *The Reading Matrix: An International Online Journal*, *16*, 98–109.

Timarova, S. (2008). Working memory and simultaneous interpreting. In Translation and its others: selected papers of the CETRA research seminar in translation studies 2007 (pp. 1–28). KULeuven; Leuven.

Tokuda, M. (2006). Dokkai ni oite suisoku o tomonau gaitekishien wa fuzuitekigoigakushuu o sokushin suru ka: tagisentakugochuu to tan'itsugochuu no hikaku [Does External Support for Lexical Inference support Incidental Vocabulary Acquisition? Comparison of Single Gloss and Multiple Choice Gloss]. Gengobunka to nihongokyouik, *31*, 122–125.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*, 183–203.

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The Effects of Repetition on Incidental Vocabulary Learning: A Meta-Analysis of Correlational Studies. *Language Learning*, *69*, 559–599.

Ushiro, Y., Hasegawa, Y., & Nahatame, S. (2013). Incremental learning of homonyms in multiple contexts among Japanese EFL readers. *Jacet journal*, *57*, 1–19.

Van Daalen□Kapteijns, M., Elshout□Mohr, M., & Glopper, K. de. (2001). Deriving the Meaning of Unknown Words From Multiple Contexts. *Language Learning*, *51*, 145–181.

Van der Schoot, M., Vasbinder, A. L., Horsley, T. M., & Lieshout, E. C. D. M. V. (2008). The role of two reading strategies in text comprehension: An eye fixation study in primary school children. *Journal of Research in Reading*, *31*, 203–223.

Vu, D. V., & Peters, E. (2020). Learning Vocabulary from Reading-only, Reading-while-listening, and Reading with Textual Input Enhancement: Insights from Vietnamese EFL Learners. *RELC Journal*.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*, 130.

Watanabe, Y. (1997). Input, Inatake, and Retention: Effects of Increased Processing on IncidentalLearning of Foreign Language Vocabulary. *Studies in Second Language Acquisition*, *19*, 287–307.

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Wesche, M., & Paribakht, T. S. (1996). *Assessing second language vocabulary knowledge: Depth versus breadth. Canadian Modern Language Review*, *53*, 13-40.

Winke, P. M. (2013). The Effects of Input Enhancement on Grammar Learning and Comprehension: A Modified Replication of Lee (2007) with Eye-Movement Data. *Studies in Second Language Acquisition*, *35*, 323–352.

Yanagisawa, A., Webb, S., & Uchihara, T. (2020). How Do Different Forms of Glossing Contribute to L2 Vocabulary Learning From Reading?: A Meta-Regression Analysis. *Studies in Second Language Acquisition, 42*, 411–438.

Zhang, D., & Koda, K. (2012). Contribution of morphological awareness and lexical inferencing ability to L2 vocabulary knowledge and reading comprehension among advanced EFL learners: Testing direct and indirect effects. *Reading and Writing*, *25*, 1195–1216.

Zhu, Y. (2015). Revisiting word exposure frequency and incidental vocabulary acquisition. *Theory and Practice in Language Studies*, *5*, 968.

Appendices

Appendix A

Text Material Used in Experiment 1

Early People in the Central American Land Bridge

People have been residing in Central and South America for many, many years now. How did ancient people live in this area thousands of years ago? Archaeologists studying the area of the Central American land bridge have been working to answer this question.

What is the Central American land bridge? It is the land that is now the countries of Costa Rica and Panama. Like a modern bridge over a river, this land bridge was used by animals and people to roam back and forth. This Central American land bridge connects the northern land that is now Nicaragua, Mexico, and so on to the southern land that is now Colombia, Brazil, and other South American countries.

People who were already residing in North America traveled down and across this land bridge. Scientists think they traveled there around 11000 BC. They probably were following large animals that they hunted and consumed. These people would have traveled on foot, following the herds of animals. They had no permanent houses. They would pack up their things and bring them along as they hunted. Their homes were like tents and were very easy to take down and put up. Archaeologists can tell these people roamed via the land bridge because they have found similar arrowheads and tools concealed in the ground in both the land bridge and in areas further north and south.

These utensils are the main record of people's movement and settlement. Ancient people at this time made their utensils, by hand, out of stone. Later, they used obsidian, which is a dark glass formed in volcanoes. They made small hatchets, arrowheads and spearheads by knapping. Knapping is when you hit one stone with another to break off little pieces. Slowly, you can shape the stone however you want. When certain stones (like obsidian or flint) break, they leave very sharp edges, which can be used to cut.

Excavators don't find indicia of these utensils very often. When so much time passes, natural things like volcanoes, rain, soil and trees destroy and conceal them. Scientists also think the oceans were lower back then. This means there once was more land that is now underwater. Probably there is more indicia like obsidian tools lying under the sea off the coasts of Costa Rica and Panama.

It is hard to tell when people stopped roaming along the land bridge and began living there. One clue is when people began farming. On the land bridge, this was around

9000 and 7000 BC. In Panama, scientists have found indicia that people were growing bottle gourds, squash, and a few other things around that time. These would be in small gardens, not big farms. People weren't consuming just the plants they grew at this time. They would gather fruits and nuts from the forest, hunt deer, fish for crabs and fish, and consume vegetables from their gardens.

Archaeologists now think that some forests were actually farms too, which makes it harder to determine when hunting and gathering stopped. Places we once thought were wild were perhaps cultivated by ancient people. This would be like an apple orchard. It looks like a forest, but has actually been planted and taken care of by humans— you wouldn't know it was a farm for apples unless you looked more closely. The same thing probably happened on the Central American land bridge and elsewhere.

Slowly, people built more permanent houses on the land bridge. The first small village excavators have found is in Costa Rica, in a place called Tronadora Vieja. There are round pole and thatch houses, which are simple huts made out of long tree branches covered in leaves and grasses. These houses date to 3800 BC and were destroyed when a nearby volcano exploded and concealed them in ash. Scientists also found the earliest maize kernels on the land bridge, as well as metates. Metates are small stone tables used to grind maize into powder. This was used in cooking and baking, like flour. Nearby at the Zoncho lagoon, excavators found a bigger village, with more houses and farms. People started residing here after Tronadora Vieja, around 3240 BC.

People were still hunting and gathering at this time, though. Why did some people settle down to grow maize and some people continue moving around? There are lots of ideas, but on the land bridge, archaeologists think it had to do with the amount of rain and water in the area. Places that had less water made it harder for the forests to remain full of fruits and animals. So in these dry areas, people started to farm more, build houses and stay put. So some groups kept hunting and gathering while others were starting to build and farm.

Archaeologists are still studying this part of the world. Hopefully in the years to come, we can find out more about how ancient people lived on the Central American land bridge.

Note. Material was adapted from https://www.readworks.org/article/Early-People-in-the-Central-American-Land-Bridge/

Appendix B

Comprehension Questions used in Experiment 1

1. What is the Central American land bridge?

○ a manmade bridge between North and South America

○ the ancient name for the country of Mexico

○ the land that is now the countries of Costa Rica and Panama

○ the strip of sea between Central and South America

2. What does the author describe in the passage?

○ the Central American land bridge and the people who traveled it

○ the gradual rise in ocean level on the Central American land bridge

○ the species of animals that crossed the Central American land bridge

○ the cultures of North American peoples before 11,000 BCE

3. Why do scientists think that people traveled the land bridge around 11,000 BCE?

○ because scientists found fossilized human footprints in the land bridge

○ because scientists found similar arrow heads and tools in the land bridge and in areas further north and south

○ because scientists found remains of tent-like residences in the land bridge

○ because scientists found evidence of cattle herding in the land bridge

4. Why don't archeologists often find evidence of tools in the Central American land bridge?

○ because ancient people threw them in the water

○ because they are buried under ground

○ because they decomposed very fast

○ because ancient people carried their tools away

5. What is this passage mostly about?

○ the rise of agriculture in South America from 9,000 BCE to today

○ how global warming has affected wildlife on the Central American land bridge

○ the development of civilization on the Central American land bridge

○ knapping and its importance in ancient Central American culture

6. How can archeologists tell when people started to settle on the land bridge, as opposed to just moving across it?

○ they found evidence of farming

○ they found arrows and spears on the bridge

○ they found animal bones

○ they found boats

7. Which sentence most accurately describes knapping, a technique used by ancient people to make tools?

○ they melted metals in stone furnaces

○ they found stones with sharp edges

○ they hit one stone with another to break off little pieces

○ they used metates to sharpen the stones they found

8. Why did some of the ancient people continue moving around?

○ because there were many active volcanoes

○ because they were attacked by wild animals

○ because dry land provided them with animals

○ because wet land provided them with fruits

Appendix C

Unknown Target Words Used in Experiment 1 With Their Descriptive Statistics

	Translation	*N*	*Non-context score*	*Context score*	*Chance of being looked up*
F1				*M* (*SD*)	
herd	群れ	11	.55 (.93)	1.45 (.93)	.64 (.50)
hatchet	斧	9	.22 (.67)	1.22 (.97)	.78 (.44)
consume	食す	3	.00 (.00)	.67 (1.15)	.33 (.58)
flint	火打ち石	10	.40 (.84)	1.60 (.97)	.50 (.53)
soil	土	5	.40 (.89)	1.20 (1.10)	.00 (.00)
orchard	園	10	.00 (.00)	.10 (.99)	.50 (.53)
hut	小屋	9	.22 (.67)	.89 (.93)	.44 (.53)
kernel	穀粒	12	.17 (.58)	.50 (.90)	.58 (.51)
lagoon	潟	7	.29 (.76)	.57 (.98)	.29 (.49)
F3				M (SD)	
reside	住む	6	.67 (1.84)	1.67 (.03)	.83 (.42)
conceal	隠す	5	.40 (.00)	.8 (1.00)	.60 (.44)
utensil	道具	10	.40 (1.03)	.8 (1.03)	.80 (.45)
excavator	掘削者	9	1.33 (1.04)	1.33 (1.04)	.78 (.50)
indicia	証拠	12	1.17 (1.05)	1.17 (1.00)	.75 (.44)
maize	コーン	11	.91 (1.00)	1.09 (1.00)	.64 (.00)
roam	歩く	9	1.11 (1.00)	1.33 (1.00)	.78 (.00)
Sum		138	.55 (.90)	1.03 (.99)	.61 (.49)

Note. Number of observations (N) is equal to portion of 12 participants who did not have previous knowledge of this word. Test scores are on a 2-point scale (0-2).

Appendix D
Context Meaning Recall Test Used in Experiment 1

1. These people would have traveled on foot, following the herds of animals.

2. Later, they used obsidian to make their tools.

3. Scientists also found the earliest grain kernels on the land bridge, as well as metates.

4. They made small hatchets, arrowheads and spearheads by knapping.

5. Ancient people at this time were growing gourds.

6. People who were already residing in North America traveled down and across this land bridge.

7. Natural things like rain, soil and trees destroyed the tools.

8. People were growing bottle gourds, squash, and a few other things around that time.

9. Probably there is more indicia lying under the sea off the coasts of Costa Rica and Panama.

10. Ancient people maintained orchards, which now look like wild forests.

11. Archaeologist have found similar arrowheads and tools concealed in the ground.

12. This land bridge was used by animals and people to roam back and forth.

13. Some people continued gathering as their main food source.

14. These houses date to 3,800 BC and were destroyed when a nearby volcano exploded.

15. When certain stones like flint break, they leave very sharp edges.

16. The first small village scientists have found consisted of pole and thatch houses.

17. Places we thought were wild were perhaps cultivated by ancient people.

18. The excavators found a bigger village.

19. In these dry areas, people started to farm more, build huts and stay put.

20. Some people settled down to grow maize.

21. People were probably following large animals that they hunted and consumed.

22. People started residing at the Zoncho lagoon around 3,240 BC.

23. Ancient people at this time made their utensils, by hand, out of stone.

24. Some groups of people continued hunting animals.

Appendix E

Text Material Used in Experiment 2

The Midas Touch

There was once a dreadfully ugly beast called Silenus. He pranced over the mountains on a pair of hairy legs, each ending with a goat's hoof. A long tail swished behind him, but from the waist up he was a man, more or less. His big belly bounced up and down as he ran along. A pair of horns sprouted out of his bald and shiny head. Quite often, slobber dribbled from his thick and purple lips. In short, this delightful creature was a satyr.

Silenus was a crony of Dionysus, the God of Wine. Dionysus often used to gather his wild band of followers in the grove for a noisy, riotous party. They included satyrs as well as Maenads, who were wild women. They would bang drums, blow pipes and horns, and crash cymbals and they danced themselves into a mad frenzy. But above all, they liked to drink wine.

One time after Silenus had been partying all night, he staggered out of the grove and into the chateau grounds of Midas, king of Phrygia. He lay down between the bushes and fell into a deep sleep. Around mid-morning Princess Zoe was walking through the gardens collecting blossoms. She saw the hairy hoof of Silenus sticking out from amongst the bushes, and she thought that a poor sick goat had come into the garden to lie down. As he was dirty and smelled not very nice, she called the steward. When he came, he pulled on the leg and found not a goat, but a satyr.

"Ugh, he's horrible," exclaimed Zoe. "Throw him on the compost heap."

"Ah, I'd better ask the king before I do that," said the steward. "After all, a satyr can bring good fortune."

When King Midas learned that there was a satyr sleeping in the hydrangea garden, he ordered that he be given a bed in the chateau until he felt better. The servants carried him on a stretcher to the best guest room. There he remained, snoring loudly and smelling like – well, a goat for almost another day.

When finally he arose, he staggered into the kitchen and noisily demanded cheese, eggs, and wine.

The cook wanted to chase him out with a cleaver, but the steward held him back saying that the satyr was a guest of the King. Silenus took the wine and went wandering around the place, leaving dirty hoof prints as he went.

When the marquise saw him, she was horrified. "Who or what is this vile creature that has come to stay with us?" she asked the King.

Midas replied that he was a crony of Dionysus, and everyone must treat him with great courtesy.

Although Princess Zoe and the marquise did their best to stay out of the way of the satyr, King Midas entertained his guest, eating and drinking with him until late at night, and playing music on the pan pipes. All in all, Silenus stayed with Midas for a week.

No one was more pleased about this show of hospitality than Dionysus, because in his eyes, anyone who honoured his crony Silenus honoured Dionysus.

A few days after Silenus had returned to the grove. Midas was walking in his hydrangea garden when he heard some strange but lovely music. He followed the sound and discovered a perfectly beautiful man sitting on the grass and playing a pipe. He knew right away that the stranger was one of the gods and he fell down on one knee.

The god said, "Get up man. I’m not one for rites. I wish to reward you. What gift would you like more than any other in the world? Power isn’t really my thing, but I can offer you wine, women or song."

"I need money," said Midas.

"Money. What good comes of money?" asked the god.

"Well of course a god like you has no use for money," said Midas, "but we mortals can never have enough of it. I wish that everything I touched turned to gold."

Although Dionysus thought it was a foolish wish, he granted it with the words, "Midas, all that you touch shall turn to gold."

The god disappeared, and King Midas rejoiced in his curse. He reached out and touched a hydrangea blossom and it turned to gold. He picked up a stone, and that too became golden. Even a clod of earth became gold.

He plucked an apple from a low branch, and it immediately became cold and shiny. He held it in his hand and said,

"Oh, how pure and perfect it is."

Then he tossed the golden apple over his shoulder, and hurried into the chateau to try his touch on random objects: columns, statues, furniture and doorknobs.

The servants heard his voice laughing and shouting, "Gold, glorious gold!" And they wondered what had gotten into the king.

The marquise heard him too. She found him turning peas into little golden nuggets.

"What has happened, dear?" she asked.

"The most wonderful thing," he replied, and he hugged her.

But this was not what he had expected. He was holding not his wife in his arms, but a cold, still statue.

Distraught, he went to the fountain to wash his hot tears from his face. But as he scooped up the water in his hands, it turned into liquid gold.

Now he realised the cruelty of his gift. He called out, "Lord Dionysus, save me from this golden curse!"

Dionysus heard him and took pity on the foolish King. He appeared sitting on the edge of the fountain and said,

"Go to the river that flows by the great city of Sardis. Make your way upstream until you come to the source. Plunge your head and body at the same moment into the foaming fountain, where it gushes out, and wash away your curse."

Midas did as he was told, and when he plunged into the torrent the banks and the flowers that grew on them became yellow and golden. But Midas emerged from the waters free of his wish for riches and gold. So as long as he lived, he rejoiced in all that was simple and natural.

Note. Material was adapted from https://www.storynory.com/the-midas-touch/

Appendix F

Comprehension Questions Used in Experiment 2

1. A satyr is a creature that is half goat and half human.

True | False

2. Dionysus was a God of Wine.

True | False

3. Princess Zoe thought Silenius was very handsome.

True | False

4. The king wanted to chase Silenius out of the castle.

True | False

5. As a reward for helping him, Silenius granted the king one wish.

True | False

6. After the king turned an apple to gold, he immediately threw it away.

True | False

7. If the king touched a living person, the person would become gold.

True | False

8. The king has enjoyed his new power for many years.

True | False

9. The king realized his power was bad when he turned his daughter to gold.

True | False

10. The king could only cure himself if he went into a river.

True | False

Appendix G

Target Words Used in Experiment 2 and Their Translations

	SG Translation	*MCG Translations*
F1		
cleaver	肉切り包丁	ナイフ、刀、包丁、出刃、出刃包丁
clod	塊	塊、一括、瘤、一丸、団塊
courtesy	礼儀	礼儀、愛想、行為、義理、エチケット
frenzy	祭り	狂気、気違い、発狂、癲狂、狂い
heap	山	山、山積み、山盛り、てんこ盛り、重ねる
mortal	人間	人間的、人間らしい、人間、仁、仁的
rite	式典	式典、祭儀、祭式、祭事、葬送
slobber	唾	唾、唾液、津液、生唾、固唾
torrent	流れ	川、流れ、小川、流れる、流水
F3		
booze	お酒	酒、アルコール、お酒、水割り、左利き
chateau	城	城、飛車、城郭、城下町、城
crony	旧友	旧友、親友、茶飲み友達、クローニー、縁故政治
curse	呪い	呪い、祟り、呪う、呪縛、罵り
fife	笛	菅、パイプ、笛、筒、導管
fount	泉	泉、万年筆、噴水、噴泉、泉水
grove	木立	い木立、藪、森、雑木林、杏林
hoof	ひづめ	爪、蹄、装蹄師、蹄叉、テクシーで行く
hydrangea	紫陽花	紫陽花、ハイドランジア、糊空木、甘茶、甘茶の木
marquise	女王	蹄女王、王妃、国王、クイーン、后
steward	執事	執事、家従、家令、家父、スチュワード

Note. All translations in kanji were shown with hiragana or katakana transcription.

Appendix H

Descriptive Statistic of Words Used in Experiment 2

	N	*Form Recognition*	*Meaning Recall*	*Meaning Recognition*	*Chance of being looked up*
F1		*M (SD)*			
cleaver	42	.36 (.48)	.05 (.22)	.36 (.48)	.81 (.40)
clod	40	.38 (.49)	.10 (.30)	.9 (.30)	.83 (.38)
courtesy	36	.25 (.44)	.06 (.23)	.83 (.38)	.86 (.35)
frenzy	43	.26 (.44)	.09 (.29)	.65 (.48)	.84 (.37)
heap	40	.23 (.42)	.03 (.16)	.73 (.45)	.83 (.38)
mortal	36	.31 (.47)	.11 (.32)	.47 (.51)	.92 (.28)
rite	42	.14 (.35)	.07 (.26)	.95 (.22)	.93 (.26)
slobber	43	.35 (.48)	.09 (.29)	.84 (.37)	.84 (.37)
torrent	40	.33 (.47)	.13 (.33)	.43 (.50)	.80 (.41)
F3					
booze	43	.53 (.50)	.23 (.43)	.88 (.32)	.93 (.26)
chateau	41	.41 (.50)	.17 (.38)	.90 (.30)	.95 (.22)
crony	43	.63 (.49)	.21 (.41)	.86 (.35)	.98 (.15)
curse	32	.59 (.50)	.41 (.50)	.91 (.30)	.88 (.34)
fife	44	.57 (.50)	.25 (.44)	.95 (.21)	.89 (.32)
fount	33	.24 (.44)	.12 (.33)	.70 (.47)	.64 (.49)
grove	38	.71 (.46)	.32 (.47)	.68 (.47)	.82 (.39)
hoof	42	.76 (.43)	.26 (.45)	.98 (.15)	.95 (.22)
hydrangea	39	.67 (.48)	.28 (.46)	.79 (.41)	1.00 (.00)
marquise	44	.55 (.50)	.39 (.49)	.93 (.25)	.98 (.15)
steward	38	.34 (.48)	.29 (.46)	1.00 (.00)	.97 (.16)
Sum	799	.43 (.50)	.18 (.39)	.79 (.41)	.88 (.32)

Note. Number of observations (N) notes how many of 44 participants did not previously know this word. Test scores are on a 1-point scale (0-1).

Appendix I
List of Target and Filler Words Used in Meaning Recognition Test

steward, adherent, cleaver, scaffold, caterer, rite, toll, heap, frenzy, clod, mandate, hoof, parity, hydrangea, enclave, grove, matinee, crony, fife, canopy, liner, mortal, cognate, torrent, bachelor, slobber, sewage, chateau, curse, proxy, fount, dean, booze, rote, courtesy, podium, marquise, berth, wicket, shoehorn

Note. Words were displayed in random order.

Appendix J

Meaning Recognition Test from Experiment 2

hoof
丘 | **ひづめ** | 半分
torrent
権力 | **流れ** | 独裁者
slobber
スリッパー | お願い事 | **唾**
heap
卵 | **山** | 希望
chateau
台所 | **お城** | 影
cleaver
花 | **包丁** | 賢さ
grove
木立 | お墓 | 宮殿
clod
雲 | 贈り物 | **塊**
mortal
庭 | **人間** | 道徳
marquise
女王 | 蛍光ペン | ご褒美
rite
式典 | 文房具 | 睡眠
fife
お金 | 五角形 | **笛**
fount
山羊 | **泉** | 発見
curse
自然 | 方向性 | **呪い**
courtesy
自由 | **礼儀** | 川
frenzy
友達 | 急ぎ | **狂気**
crony

王冠 | 川 | **旧友**
booze
よそ者 | **お酒** | 男子
hydrangea
水素 | **あじさい** | 半獣神
steward
開始 | お客 | **執事**