UNIMODAL AND BIMODAL INPUT IN INCIDENTAL VOCABULARY LEARNING: COGNITIVE PROCESSES AND THE DEVELOPMENT OF DIFFERENT KNOWLEDGE TYPES

By

Ayşen Tuzcu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Second Language Studies – Doctor of Philosophy

2023

ABSTRACT

Researchers have investigated the promise of unimodal and bimodal input in enhancing vocabulary learning from meaning-focused activities. Compared to unimodal input, the simultaneous presentation of written and aural input in bimodal input has been argued to direct L2 learners' attention to words and enhance the form-meaning links for new vocabulary (e.g., Long, 2017; Malone, 2018). However, so far, the research focus has been on the learning outcomes, and there is a lack of studies showing the processing of unimodal and bimodal input for vocabulary learning. Moreover, the learning outcomes have traditionally been assessed through paper-based accuracy tests that tap into conscious, verbalizable knowledge of target words. How bimodal input contributes to fluent form and meaning retrieval of the newly learned vocabulary during a real-time activity (e.g., reading) is still not clear and warrants attention. Therefore, the main motivation of this study was to fill these research gaps by comparing the effectiveness of unimodal and bimodal input in facilitating different cognitive processes and developing different types of vocabulary knowledge.

Sixty-three adult L2 English speakers were randomly assigned to reading only and reading while listening groups and read a 9500-word graded reader containing 24 pseudowords with or without audio over two days. Cognitive processes were operationalized using two variables: attention (eye fixations on target words) and awareness (reported recall of encountering a target word in the text). Reading processes were recorded using eye-tracking; awareness in learning the pseudowords were probed using retrospective interviews. Learning outcomes were measured with three written untimed tests (form recognition, meaning recall, and meaning recognition) and one sentence-reading test that indicated fluent retrieval of word form and meaning in real time. Four eye-tracking measures, gaze duration, regression path duration,

rereading time, and total reading time, were used to assess the effectiveness of reading only and reading while listening on lexical access and context integration of target pseudowords. The results from eye-tracking measures were analyzed using growth curve models; the results from retrospective reports regarding awareness levels were analyzed using mixed logistic models; and vocabulary tests were analyzed using mixed logistic models (form recognition, meaning recall, and meaning recognition) and linear mixed effects models (sentence-reading test).

The growth curve models indicated a nonlinear decrease in reading times from the first encounter with target pseudowords to the last encounter. Moreover, the reading behaviors of the two groups differed for the early and late reading processes. The audio in the reading while listening condition augmented the attention to target pseudowords in early reading processes as reflected in gaze durations, whereas reading only condition promoted increased attention to target pseudowords in late reading processes as reflected in rereading times. Both groups were similar in terms of their awareness levels for the target words in the reading text. They also had comparable learning gains on the meaning recall and meaning recognition tests, but the reading while listening group had significantly higher form recognition scores, particularly on the delayed test. Moreover, the reading while listening group had more robust lexical representations for the target pseudowords as reflected in the reading times in the sentence-reading test. These results pointed to a positive effect of access to phonologic forms through audio in vocabulary learning. Overall, the findings of this study provided a broader and more comprehensive understanding of the cognitive processes and the types of knowledge promoted by unimodal and bimodal input.

ACKNOWLEDGEMENTS

I have been very lucky to receive guidance and help from many great individuals in various stages of my PhD journey. Their constant support has made this long journey eye-opening and entertaining for me. I would like to express my gratitude to the following individuals for their constant support and guidance.

First, I would like to express my deepest gratitude to my advisor Dr. Shawn Loewen for all his support and encouragements. His encouragements have helped me tremendously stay on track with the dissertation process. He believed in me and supported me academically in the best way possible. I would also like to thank the members of my dissertation committee: Dr. Aline Godfroid, Dr. Paula Winke, and Dr. Meagan Driver. Their insightful and constructive feedback has helped me improve the quality of this project to a great extent.

I would also like to thank the members of the Second Language Studies (SLS) community, SLS students and faculty, for providing a very supportive environment. I am especially thankful for my colleagues and friends for always cheering me on – SLS 2019 cohort and "Booze Clues" team. My special thanks go to Brittany Finch being the greatest friend anyone could ask for. I am also grateful for "SSLA Energy Family" - Dr. Susan Gass, Dr. Luke Plonsky, Lizz Huntley, and Hima Rawal - for their emotional and academic support for the past two years. The relationships and memories that I have developed with them all will always be priceless for me.

Most importantly, none of this would have been possible without the love and encouragement of my family, my dear mother, father, and sister. I am grateful for their unconditional love and belief in me. My special thank goes to Adil Han Orta for his constant encouragement that helped keep going whenever I felt lost or discouraged.

Last but not least, this dissertation would not have been possible without with the financial support of the following grants: Fulbright Doctoral Degree Grant, the Language Learning Dissertation Grant, Duolingo English Test Doctoral Dissertation Research Award, and the research and conference support from the SLS program.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Knowledge of vocabulary is an essential aspect of achieving a high proficiency in a second language, and it is one of the predictors of productive (i.e., speaking and writing) and receptive (i.e., reading and listening) second language (L2) skills (Qian & Lin, 2020; Zareva et al., 2005). Research has demonstrated that L2 learners can acquire vocabulary as a by-product of meaning focused activities such as reading (e.g., Godfroid et al., 2013; Pellicer-Sánchez, 2016; Webb & Chang, 2015a), listening (Van Zeeland & Schmitt, 2013; Vidal, 2011) and watching TV (Montero Perez, 2020; Peters & Webb, 2018; Puimège & Peters, 2020). Although this type of incidental learning activities has provided promising evidence that vocabulary can be learned without instruction, they have also been criticized for requiring a high amount of exposure to target words (Webb et al., 2013). Moreover, this type of learning is generally more suitable for L2 learners who have a higher proficiency in the L2 (Puimège & Peters, 2020).

To enhance learning from meaning-focused activities, many advocated the use of audio along with text (e.g., Chen, 2021; Malone, 2018; Webb & Chang, 2020). One of the main arguments for the facilitative effects of bimodal input is that simultaneous presentation of written and aural input directs L2 learners' attention to words, enhances the form-meaning links for new vocabulary, and results in better learning outcomes (e.g., Malone, 2018). However, how bimodal input is processed by learners is still unclear since researchers have only recently started to explore the processing of bimodal input (e.g., Conklin et al., 2020). For example, in an eye-tracking study, Conklin et al. (2020) reported moderating effects of audio on the reading speed of L1 and L2 speakers of English. L1 speakers but not L2 speakers read the given reading texts significantly slower while reading with audio. Moreover, there was a discrepancy between participants' gazes and the audio. Specifically, participants' eye movements were ahead of the

audio. Although this study provided important insights into how L1 and L2 learners cognitively engage with a reading text while reading with or without audio, the text used in the study did not contain any novel words. Therefore, it is still not clear exactly how audio support provided during reading while listening influences the reading behaviors of L2 learners on unknown words compared to reading alone.

Moreover, the learning gains in unimodal and bimodal input studies have traditionally been assessed through paper-based accuracy tests that tap into conscious, verbalizable knowledge of target words (i.e., declarative knowledge). However, these tests do not necessarily indicate spontaneous language use of the learned vocabulary, which is critical for proficient L2 use (Godfroid, 2019). Therefore, how simultaneous exposure to text and audio contributes to fluent retrieval of form and meaning of the newly learned vocabulary during a real-time activity such as reading (i.e., nondeclarative knowledge) is still unknown.

To fill in these gaps in the previous research, in this dissertation, I compared the effectiveness of bimodal input in facilitating different cognitive processes and developing different types of vocabulary knowledge with that of unimodal input by using complementary data collection instruments that indicate cognitive processes (i.e., attention, awareness and intentionality) and learning gains in incidental vocabulary learning. This dissertation consists of seven chapters, including this introduction chapter. In the next chapter, Chapter 2, I review literature on incidental vocabulary learning by focusing on the constructs of attention and awareness, the development of different aspects of vocabulary knowledge in incidental learning environments, and the factors influencing the learning of unknown vocabulary. While summarizing previous research, I also highlight the research gaps that motivated the current study and provide the research questions addressed in this study. In Chapter 3, I describe the

methodology of this study, detailing the participants, study design, the materials used for learning and testing phases, procedure, and data analysis. In Chapter 4, I provide the results for the eye-tracking measures recorded during the learning phase; in Chapter 5, I report the results from the retrospective verbal reports regarding the awareness levels of the participants; and in Chapter 6, I give the results for the learning gains assessed through four vocabulary tests. Finally, in Chapter 7, I discuss the findings in relation to previous research, provide information about the theoretical, methodological, and pedagogical implications, and conclude the dissertation.

CHAPTER 2: LITERATURE REVIEW

As explained in the introduction, the present study aims to explore the processing of unimodal and bimodal input, and the learning of new vocabulary from these two input types. Therefore, the goal of the current chapter is to provide a concise review of the literature on incidental vocabulary learning with a focus on the learning processes and outcomes. Throughout this chapter, I provide a summary of related research and highlight the main motivations for the present study. First, to provide a bigger picture of cognitive processes in L2 learning, I review the theoretical conceptualizations of attention and awareness in cognitive psychology and second language acquisition fields while presenting the most prominent and popular models of each construct. Second, I define different second language learning types (i.e., incidental, intentional, explicit, and implicit learning) and discuss the learner-oriented and method-oriented definitions of incidental vocabulary learning by highlighting the shortcomings in each of these definitions. Then, I summarize the studies that examined the constructs of attention and awareness in incidental vocabulary learning while presenting the two data collection methods that have been used by the researchers: eye tracking for measuring attention and retrospective verbal reports for assessing awareness. Third, I discuss the effects of input modality in vocabulary learning from incidental learning conditions. I emphasize that, in addition focusing on the learning outcomes, researchers need to conduct empirical studies examining the cognitive processes underlying learning from bimodal input. I underline that the arguments for the facilitative effects of bimodal input in processing incidental vocabulary learning can only be confirmed with the use of appropriate data collection tools that provide information about attention and awareness levels. Finally, I discuss several item-related and learner-related factors that have been indicated to affect incidental vocabulary learning from different input modalities. At the end of the chapter, I

highlight the main goals of the present dissertation and present my research questions.

Attention and Awareness in Cognitive Psychology

The construct of attention is one of the primary factors that affect the processing of a stimulus (Chun et al., 2011). Increased attention to stimuli results in the creation of stronger associations, which in turn, facilitates binding of the ideas (Cowan, 2014), forming new knowledge, and retrieving them from memory (Jiménez, 2003). Therefore, attention plays an important role in learning of an input (Robinson et al., 2012).

In a review article, Chun et al. (2011) define attention as "a core property of all perceptual and cognitive operations" (p. 73). In the presence of a stimulus, attention acts as a filter and allows for selecting the information to be processed, sustaining focus on a particular information among competing stimuli, and activating related concepts in long-term memory. Therefore, attention plays a main role in the selection of which stimulus to be processed and how deeply it is processed. Attention is a continuous variable; people may allocate different amounts of attention to different stimuli (Cowan, 1995).

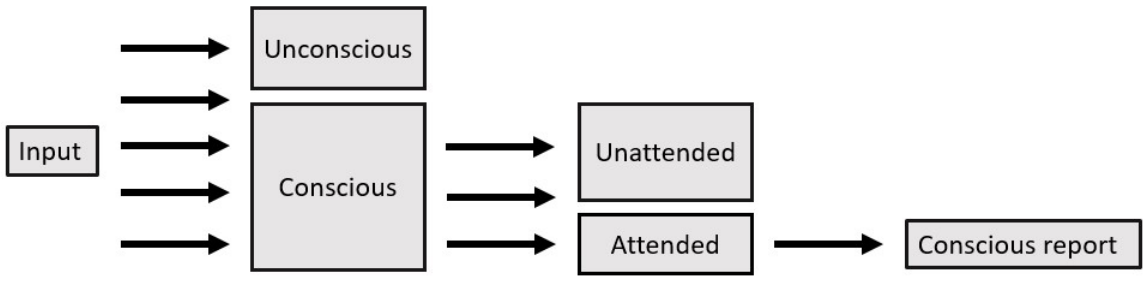
Chun et al. (2011) provide a useful taxonomy where they consider attention as a construct of multiple systems rather than a unitary system. They argue that there are two attentional systems through which attentional targets are processed: external attention and internal attention. External attention is related to the selection and adjustment of the information perceived through senses (e.g., an auditory input coming from a radio or a written text presented on a computer screen). External attention to a stimulus can be automatic and involuntary such as attending to noises coming from outside, i.e., perceptual attention, or it may involve voluntary control to some extent such as attending to the words of a friend while they are speaking, i.e., focal attention (Robinson, 2017). The processing of the stimuli through external attention can be a

stimulus-driven bottom-up process or a goal-direct top-down process. This means that external attention given to an external stimulus can change depending on the factors such as task demands (Robinson et al., 2012). Internal attention, on the other hand, is responsible for selecting and adjusting internally existing information such as the representations in working and long-term memories; therefore, internal attention is a goal-directed top-down process.

Awareness, which can be defined as "the subjective, contentful feel of experience that can be reported to others, to varying extents" (Robinson et al., 2012, p. 247), is a construct that is intricately related to attention. Although being closely related, Lamme (2003) argues that these two constructs are distinct from each other and provides a cognitive model showing how awareness is different from attention. This model consists of three steps, each step tapping into the constructs of consciousness, attention, and awareness respectively (see Figure 1). In this model, Lamme first distinguishes between conscious and unconscious processing and asserts that a stimulus can be processed consciously or unconsciously. The consciously processed input, then, can either receive attention or remain unattended. Only the attended stimuli can be reported to the others, which is considered as awareness. In other words, attention is considered to be a necessary condition for awareness. However, not all attended stimuli reaches the level of awareness (Chun et al., 2007).

Figure 1

Lamme's Model of Consciousness, Attention, and Awareness



Unlike attention, awareness is considered to be absolute (Cowan, 1995); however, instead of being a dichotomy (aware vs. unaware), awareness may exist in different degrees (Andringa, 2020; Spit et al., 2021). For example, by citing the work of Cleeremans (2011), Spit et al. (2021) categorizes awareness into two levels: awareness at the phenomenal level and awareness at access level. Phenomenal awareness refers to the awareness at the level of verbalizable, subjective awareness (e.g., being able to describe a grammatical rule); access awareness exists when people have a cognitively accessible but subjectively inexpressible awareness (e.g., perceiving the existence of a grammatical rule but not being able to describe it verbally).

To summarize, cognitive psychologists consider attention and awareness as related but distinct constructs. While attention is argued to be necessary for reaching awareness, individuals might not be aware of all attended stimuli. If attention and awareness reflect separate processes, one question that is of particular importance for L2 learning, is the extent to which each of these constructs influence learning different aspects of an L2 such as grammar and vocabulary.

Attention and Awareness in Second Language Acquisition

Just like in cognitive psychology, the constructs of attention and awareness and their relation to learning have been widely investigated by second language acquisition (SLA) researchers (see Robinson et al., 2012 for a review). Attention has been one of the primary predictors of various L2 learning related phenomena. For example, several researchers have found that instructional conditions or tasks with different demands can influence the availability of attentional resources, which in turn can predict the quality and quantity of the learning of various L2 grammatical structures (e.g., Indrarathne & Kormos, 2017; Issa & Morgan-Short, 2019) and L2 vocabulary (e.g., Godfroid et al., 2013; Godfroid et al., 2018; Pellicer-Sánchez, 2016; Jung & Révész, 2018). In these studies, attention has been conceptualized as a mechanism

that regulates and facilitates the use and learning aspects of an L2, and it has been operationalized as a continuous variable, e.g., more attention leads to more learning. Awareness, on the other hand, has been mostly used to investigate explicit and implicit learning of L2 grammatical structures and the development of explicit and implicit knowledge (e.g., Andringa & Rebuschat, 2015; Suzuki & DeKeyser, 2017). The role of awareness in learning L2 vocabulary has been investigated as well (e.g., Godfroid & Schmidtke, 2013), but not as extensively as the learning of L2 grammatical structures.

A third construct that has been of considerable interest to cognitively-oriented SLA researchers is noticing. According to Schmidt (1990), "noticing can be operationally defined as availability for verbal report, subject to certain conditions" (p. 132). When the definition of awareness as the ability to verbally report an experience is considered (e.g., Lamme, 2003), Schmidt relates noticing to awareness. Schmidt (1990, 2010) categorizes awareness into two: awareness as noticing and awareness as understanding. According to Schmidt (1990), although understanding is not a required condition for learning, awareness at the level of noticing is essential for learning. In his later publications, Schmidt puts more emphasis on the construct of attention and argues that noticing is "nearly isomorphic with attention" (Schmidt, 1995, p. 1). From these definitions, noticing, as conceptualized by Schmidt (1990, 1995, 2001, 2010), can be considered as a combination of attention and awareness. In fact, Godfroid et al. (2013) considers awareness and attention to be the two sides of the same noticing coin.

In the early version of the hypothesis, Schmidt (1990) asserts that noticing is "the necessary and sufficient condition for the conversion of input to intake" (p. 129). However, this early version of the hypothesis is criticized due to a lack of definition of the term "intake" (e.g., Godfroid et al., 2013). Moreover, Schmidt's argument regarding lack of learning without

awareness is challenged in studies and reviews where researchers have reported evidence for learning without awareness (e.g., Williams, 2005). In the later version of the hypothesis, Schmidt (2001, 2010) weakens his hypothesis by stating that noticing plays at least a facilitative, if not necessary and sufficient, role in L2 learning. To prevent any theoretical confusions, following the discussions of Godfroid et al. (2013), the current study focuses on the constructs of attention and awareness rather than noticing to investigate the cognitive processes underlying vocabulary learning from unimodal and bimodal input.

Incidental Vocabulary Learning

The terms incidental and intentional have been used by cognitive psychologists mainly for methodological reasons since 1960s to investigate issues regarding different types of information processing and memory (Hulstijn, 2003, 2013). Towards the end of the century, these two terms have been adopted by researchers in the SLA field to investigate L2 learning, and they are still widely used to examine different issues in L2 learning, the learning of L2 vocabulary being the most prominent and popular one (e.g., Godfroid et al., 2013, 2018; Laufer & Girsai, 2008; Pellicer-Sánchez & Schmitt, 2010; Peters et al., 2009). However, the term incidental, in particular, have been interpreted in various ways depending on the perspectives taken by the researchers, which triggered discussions among SLA researchers regarding its exact definition (Gass, 1999; Hulstijn, 2003).

From a learner-oriented perspective, incidental vocabulary learning is defined as the learning of vocabulary items unintentionally, i.e., as a by-product of the other activities such as speaking, listening, reading, and writing, where the main focus is not on learning or teaching vocabulary items (Hulstijn, 2013; Schmitt, 2010). According to this definition, one example of incidental vocabulary learning can be learning new words while reading a book for pleasure

without aiming to do so. In this perspective, incidental vocabulary learning is generally contrasted with intentional vocabulary learning, which refers to the learning or teaching situations where there is an intention to learn vocabulary items (Hulstijn, 2013). In intentional vocabulary learning, learners generally use different strategies and rehearsal techniques to memorize the target vocabulary (Hulstijn, 2013). Therefore, during incidental vocabulary learning, learners are expected to focus on meaning, whereas during intentional vocabulary learning their focus is expected to be on form.

One example study examining the incidental learning of vocabulary was conducted by Horst and colleagues (1998). In this study, the researchers investigated the incidental vocabulary learning from a simplified novel. Participants were 34 low-intermediate English learners in an intensive English course. To promote incidental learning and preclude intentional learning of unknown words in the novel, the novel was read aloud in class by the teacher, and participants were asked to follow along in their books. The learning gains were measured using a form recognition and a word association test given before and after the reading phase. The researchers reported small gains on both tests (22% and 16% respectively), and the gains were higher for the words with higher frequency of occurrence in the novel.

As exemplified in the study by Horst et al. (1998), the main assumption underlying the learner-oriented perspective to incidental vocabulary learning is the absence of intention in learning the new words in an activity. For example, a learner may pick up new words while reading a text or listening to an audiobook; all these learning gains are considered as by-products of the main activity. However, what exactly learners do when they encounter new words in a text or an audio can be different from what they are expected to do. For example, while reading or listening, learners may focus on comprehending the general idea in a text and simply attend to

new words without intending to learn them, or they may try to infer their meanings from context and even to keep these words in their minds, which makes their learning intentional. In other words, it is difficult to decide if learners learn the new words incidentally or intentionally while reading or listening. Barcroft (2004) argues that in a real-world context, vocabulary learning is neither purely intentional nor purely incidental; therefore, instead of a dichotomous approach, intentionality can be perceived as a continuum. Moreover, the degree of intentionality in incidental learning conditions may show differences from one learner to another (Ender, 2016), which further challenges the learner-oriented approach to incidental vocabulary learning. Gass (1999) and Hulstijn (2003) also point out the lack of consensus on the role of attention and awareness in the learner-oriented definition of incidental vocabulary learning.

From the method-oriented perspective, incidental vocabulary learning is operationalized as the learning of vocabulary in meaning-focused activities where learners are not forewarned about the existence of a vocabulary posttest (Hulstijn, 2003, 2013). In other words, the only criterion that distinguishes between incidental and intentional learning conditions is the presence or absence of an explicit information about the vocabulary posttests. In studies where this methodological definition is adopted, participants in the incidental condition are informed about the comprehension test after the learning phase in advance, so they are encouraged to read for meaning (e.g., Godfroid et al., 2013; Pellicer-Sánchez, 2016). However, these participants are not forewarned about the vocabulary posttests. Participants in the intentional condition, on the other hand, are informed about the vocabulary posttests along with comprehension posttests.

One study that adopted Hulstijn's (2003, 2013) methodological definition to investigate the learning gains in incidental and intentional learning conditions was conducted by Peters et al. (2009). In this study, the researchers examined the effects of three vocabulary enhancement

techniques (vocabulary test announcement, post reading vocabulary task, and word relevance in the comprehension test) on 137 L1 Dutch L2 German learners' learning gains of new words and their behavior of looking up word meanings from an online dictionary. These learners were assigned to four groups and their attention to target words in the treatment text were manipulated in three ways: 1) whether learners were informed about the vocabulary posttests before the reading phase, 2) whether participants were asked to complete an additional vocabulary task after the reading phase, and 3) whether the target words were necessary to complete the comprehension questions. The results indicated increased word look up behaviors for the learners who were forewarned about the vocabulary posttests. Moreover, test announcement had a positive effect on the word form recognition levels of the learners; however, test announcement was not a significant predictor of recalling the meaning of new words. This study showed that test announcement itself may not be as facilitative in promoting participants' learning of new words.

Although the methodological operationalizations of incidental and intentional vocabulary learning are clear in terms of experimental design, these operational definitions are far from providing theoretical grounds of the two learning types. That is, the experimental designs based on the methodological operationalizations of the two terms do not provide any insights regarding learners' attention and awareness levels in each type of learning. For example, in an incidental vocabulary learning study using Hulstijn's (2013) methodological definition, Pellicer-Sánchez and Schmitt (2010) reported that intentional learning is possible even in incidental learning conditions. In their study, the participant who had the highest learning gains from reading a book reported that after seeing the Igbo words (the language of an ethnic group in Nigeria) in the book she expected to be tested on these words. Therefore, she paid more attention to these words while

reading by underlining them and reviewing them after finishing the reading. Pellicer-Sanchez and Schmitt's (2010) study clearly shows that even in incidental vocabulary learning conditions where learners are not informed about the main aim of the study, i.e., learning new vocabulary, learners may still try to learn the new words intentionally by paying more attention to them while reading, trying to infer their meanings from context, checking their meanings from a dictionary, or going back to those words for a review. These findings regarding the availability of intentionality in learning was corroborated in other studies such as the one by Godfroid et al. (2018). In this study, researchers reported that L2 learners who expected to receive a vocabulary test after finishing reading intentionally tried to learn the Dari words in the book *A Thousand Splendid Suns*, and their learning gains for the meanings of new words were higher than those who remained naive to the main aim of the study. In other words, even in incidental learning conditions created by teachers or researchers, learners may approach the task and input differently than expected, and they can still engage in the intentional learning of novel words. Or learners may also prefer to not focus too much on unknown vocabulary even when they are informed about the vocabulary posttests. Therefore, Hulstijn (2003) comments that these methodological definitions should only be used to explain and discuss experimental learning conditions and should not be expanded to explain learners' attention, awareness, and intentionality levels.

In summary, whether defined using a learner-oriented or method-oriented perspective, investigating incidental vocabulary learning from a meaning-focused input can be challenging for researchers. First, the methodological definition, i.e., lack of information about vocabulary posttests, cannot guarantee a lack of expectation for vocabulary posttests by L2 learners as exemplified in studies by Pellicer-Sánchez & Schmitt (2010) and Godfroid et al. (2018). Second,

even when incidental learning is defined using the learner-oriented definition, determining the intentionality levels of the L2 learners in engaging with novel words in the input can be difficult, particularly if there are no measures assessing L2 learners' intentionality levels for learning the target words in a text. It is also important to note that the term incidental vocabulary learning used in the majority of the vocabulary studies refers to the learning conditions (Loewen, 2020) rather than the cognitive processes underlying the learning. In fact, as explained in the following section, if the aim is to delve more into cognitive processes underlying learning from a meaning-focused activity, implicit and explicit learning terms can be more informative for researchers.

Cognitive Processes Underlying Incidental Vocabulary Learning

Explicit learning refers to acquisition which involves conscious intention to learn the rules, concept, and regularities in the input (Hulstijn, 2005), whereas implicit learning refers to acquisition without conscious intention or awareness (N. Ellis, 1994; Hulstijn, 2005; Reber et al., 1999). As pointed out by Ender (2016), incidental learning refers only to learning without intention, and it does not imply learning without awareness, so incidental learning does not correspond to implicit learning.

Incidental vocabulary learning can result from both explicit and implicit learning. For example, learners may try to create form-meaning links of unknown vocabulary in a reading or listening text using different strategies such as looking up a word in a dictionary to ensure successful comprehension. Although looking up the meaning of a word from a dictionary indicates explicit focus on the word, i.e., explicit learning, if the main intention in using the dictionary is for comprehension rather than memorizing the meaning of the word, the learning can still be considered as incidental. Moreover, in the same incidental learning condition, learners, particularly the ones with high proficiency, may process the meaning of a word without

awareness, i.e., they may "implicitly absorb the meaning of an unknown lexical item from the context and map it onto a given form" (Ender, 2016, p. 556), which results in implicit learning in an incidental condition. For example, in an incidental learning condition, Ender (2016) investigated the lexical processing strategies of 24 L2 French learners who were given a newspaper article and a literary text to read in L2 French, summarize the content of the newspaper article, and answer several comprehension questions about the literary text. The learners were also asked to think aloud while reading to examine their strategies for coping with unknown vocabulary and were given a surprise vocabulary test to assess their retention levels. The main lexical processing strategies examined were: (1) ignoring the unknown word, (2) consulting a dictionary, (3) inferring the word's meaning with the help of various cues, and (4) inferring the word's meaning and consulting a dictionary subsequently to check the inferred meaning. The results indicated that participants mostly ignored the unknown words while reading (38%) or they consulted a dictionary for their meaning (39%), whereas inferring strategies (alone or with subsequent dictionary consultation) were used less (%23 combined). However, although not used extensively, inferring strategies alone or combined with consulting a dictionary led to the higher retention rates than ignoring or consulting a dictionary, which provided an evidence for the existence and effectiveness of explicit learning processes in incidental learning conditions. Moreover, despite being small in size, participants learned the meaning of several words which they ignored while reading, i.e., did not explicitly focus on during think alouds, indicating an evidence for implicit learning of vocabulary.

Ellis (1994) also comments that vocabulary learning involves both conscious and unconscious processes: "the recognition and production aspects of vocabulary learning rely on unconscious processes, whereas meaning and mediational aspects of vocabulary heavily involve

explicit, conscious learning processes" (p. 39). So, according to this view, encountering a new word in the input can leave traces in memory regarding its form (orthographic representation for written input and phonologic representation for aural input) without any conscious processing, whereas creating form-meaning links through inferring requires conscious processing. Therefore, depending on the type of processing, learners may acquire different aspects of new vocabulary in an incidental vocabulary learning activity. Moreover, type of processing can influence the types of knowledge acquired. While conscious, explicit learning processes can result in the development of explicit knowledge, unconscious, implicit learning processes can be more conducive for the development of implicit knowledge (Toomer & Elgort, 2019). Despite these hypothesized variations in cognitive processes in incidental vocabulary learning conditions, most studies have focused on the end products, i.e., learning gains in posttests, whereas the underlying cognitive processes in these conditions, i.e., attention, awareness, and intentionality, have only been investigated by a handful of studies (e.g., Godfroid et al., 2013; Godfroid et al., 2018; Pellicer-Sánchez, 2016). Moreover, the number of studies examining the learning of different vocabulary knowledge types (i.e., explicit knowledge or implicit knowledge) is limited. The cognitive processes of L2 learners while engaging with novel words in an input can be examined using different measures. The two tools that have been used to measure L2 learners' attention and awareness levels include eye-tracking and retrospective verbal protocols respectively. How these tools are used in vocabulary research is explained in the next section.

Eye Tracking and Verbal Protocols as Measures of Attention and Awareness

Following the arguments regarding the importance of attention (Robinson, 1995; Tomlin & Villa, 1994) and awareness (Schmidt, 1990, 1995) in L2 learning, researchers have investigated how incidental vocabulary learning conditions influence L2 learners' attention to

and awareness of unfamiliar words in a text while reading. One of the most prominent and popular tools that is used in recent years to investigate the cognitive processes of L2 learners while reading or listening to L2 input is eye tracking (e.g., Conklin et al., 2020; Godfroid et al., 2013; Godfroid et al., 2018; Jung & Révész, 2018; Pellicer-Sánchez, 2016).

Eye tracking refers to the recording of an individual's eye movements, i.e., the location and movement of their pupil on a screen, while processing a visual input (Godfroid, 2020). The main underlying assumption in eye tracking is that there is a link between eyes and mind (Rayner, 1998; Rayner et al., 2004; Reichle et al., 2012). That is, eye gaze, which is an indicator of overt attention, can shed light on cognitive processes, i.e., covert attention. According to this assumption, there is a close relationship between cognitive processes and where and when eyes move on visual stimuli. For example, increased processing demands (e.g., processing of low-frequency words) can be reflected in eye movements as larger number of fixations or longer fixation durations.

Eye-tracking methodology provides information about various aspects of eye movement data such as fixations, saccades, and regressions, which tap into early or late stages of cognitive processes. Godfroid et al. (2013) argued that eye tracking can be a more sensitive measure of attention as it can provide information about both the locus and amount of attention during processing. Moreover, eye movements occur naturally in reading or processing visual stimuli, so eye tracking does not require individuals to complete a secondary task such as thinking aloud, which might alter the cognitive processes of the individuals or their performance in the main task (Godfroid, 2020). In fact, eye tracking "does the best job of revealing moment-to-moment processes in reading" (Rayner & Pollatsek, 2006, p. 614), and it can reflect the natural reading processes of individuals (Cop et al., 2015).

Godfroid et al. (2013) was one of the earliest studies that investigated the attention to unfamiliar words in an incidental vocabulary learning condition using eye tracking. In this study, 28 Dutch L2 learners of English read 20 short paragraphs (8 fillers and 12 experimental) containing 12 target words in four conditions while their eye movements were being recorded. The conditions differed in terms of the presence of pseudowords and the availability of contextual cues for inferring the meaning of the pseudowords. The results indicated increased fixation durations on the pseudowords than control words, regardless of the availability of appositive contextual cues. Moreover, there was a positive relationship between the attention allocated to target pseudowords and the vocabulary learning gains in a surprise vocabulary test. That is, participants were more likely to recognize the pseudowords that they spent more time looking at while reading in a vocabulary posttest.

Following Godfroid et al.'s (2013) landmark study, several studies investigated the processing of unfamiliar words in longer texts (Elgort et al., 2018; Mohamed, 2018; Pellicer-Sánchez, 2016). For example, Pellicer-Sánchez (2016) investigated the online reading behaviors of 37 L2 speakers of English while reading a 2300-word short story using eye tracking. The short story contained six pseudowords and six control words (real words), and each pseudoword and control word was repeated eight times. The vocabulary learning gains were measured with a form recall, a form recognition, and a meaning recall test. The results indicated decreased reading times for the pseudowords from the first encounter to the last. Particularly, there was a significant decrease in the reading times after three or four encounters, and the pseudowords were read similarly to control words at the last encounter. Similar to the findings of Godfroid et al. (2013), Pellicer-Sachez (2016) also reported a positive relationship between total reading times of the target pseudowords and the learning gains in the vocabulary tests. Godfroid et al.

(2018), later, extended these findings to the learning of Dari words in a novel. Elgort et al. (2018) also investigated the effects of exposure frequency on the reading times of low-frequency words in a nonfictional book and reported decreased reading times after around eight exposure. Additionally, after the reading phase, they examined the reading times for the target words in neutral contexts using a sentence-reading task and reported longer reading times in neutral sentences, suggesting weak lexical-semantic representations for the new words even after multiple exposures.

While eye-fixation times can reflect the amount of attention paid by individuals to words or grammatical structures in a written text, they are not direct indicators of awareness levels of individuals. Therefore, Godfroid and Winke (2015) emphasize the importance of triangulating eye-tracking data with verbal reports gathered from individuals to shed light on the attention and awareness constructs together. These verbal reports can be collected from individuals retrospectively using stimulated recall protocols or individual interviews.

One incidental vocabulary learning study that investigated the awareness levels of L2 learners along with their attentional levels was conducted by Godfroid and Schmidtke (2013). In this study, the researchers examined L1 Dutch- L2 English speakers' learning of 12 pseudowords in 20 short English paragraphs. Attention was measured with eye-tracking and operationalized as the total fixation time on the pseudowords. Awareness was measured with retrospective verbal reports, and it was operationalized as the extent a participant remembered reading a particular pseudoword in the paragraphs. In the light of participant reports and following the literature on episodic and semantic memory, the researchers divided awareness into three categories: (1) autonoetic awareness, (2) noetic awareness, and (3) no awareness. Autonoetic awareness, which is related with the episodic memory, is defined as "the mental reinstatement of personal

experiences of previous events at which one was present" (Gardiner, 2001, 1351). Noetic awareness refers to familiarity and knowing in the lack of such recollection of personal experiences, and it is a property of semantic memory (Gardiner, 2001). Based on these definitions of autonoetic and noetic awareness, Godfroid and Schmidtke operationalized autonoetic awareness as the instances where participants reported a conscious remembrance of reading a word in one of the treatment paragraphs. The instances where participants reported familiarity with the target word forms without providing any self-recollection of their presence in the treatment paragraphs were considered to illustrate noetic awareness. Finally, all other instances where participants did not report any familiarity or remembrance of target words were considered as no awareness instances. The researchers reported a positive relationship between attention and the possibility of recognizing target words in a posttest. Moreover, both noetic and autonoetic awareness was found to predict learning gains in the posttest. Finally, the results showed that the level of awareness from no awareness to autonoetic awareness increased as the amount of attention, reflected in longer reading times, increased. In other words, more attention paid to a target pseudoword induced greater awareness, autonoetic awareness in particular, for that pseudoword. Based on these findings, Godfroid and Schmidtke argued that autonoetic awareness "arguably represents a higher level, or richer quality, of awareness than noetic awareness" (p. 198), and both attention and awareness data can be informative for exploring the cognitive processes of L2 learners while learning new L2 vocabulary.

Factors that have been reported to affect L2 learners' eye fixations and reading times while reading and their verbal reports after reading includes frequency of occurrence, predictability from context, word length, part of speech, order of presentation, vocabulary test announcement, textual enhancement, and L2 learners' proficiency and their vocabulary size (S.

Choi, 2017; I. Choi, 2018; Godfroid et al., 2018; Joseph et al., 2014; Mohamed, 2018; Puimège et al., 2023). All being highly informative, these previous studies mostly focused on the learning from written input. How attention and awareness levels are influenced by input modality is still unclear.

Input Modality in Incidental Vocabulary Learning

L2 learners can acquire vocabulary as a by-product of meaning focused activities such as reading (e.g., Godfroid et al., 2013; Pellicer-Sánchez, 2016; Webb & Chang, 2015a), listening (Van Zeeland & Schmitt, 2013; Vidal, 2011), and watching TV (Montero Perez, 2020; Peters & Webb, 2018; Puimège & Peters, 2020). These activities differ in terms of the mode of the input L2 learners receive which is a factor that may influence the vocabulary learning amounts in incidental learning conditions. That is, several input modes can be more effective in promoting the learning of new words.

The effectiveness of bimodal input, i.e., the simultaneous presentation of written and aural input, has received considerable attention from L2 researchers. In bimodal input, the written input is accompanied with aural input; that is, aural input is used as a way to enhance written input or vice versa. Therefore, aural input accompanying written input can be considered as a form of input enhancement, i.e., lexical focus on form (Loewen, 2020). Bimodal input is generally contrasted with that of unimodal input (either written or aural input) to investigate their effectiveness on the learning of single words (e.g., Brown et al., 2008; Chen, 2021; Malone, 2018; Teng, 2018) or multiword items (Dang et al., 2022; Tuzcu, 2023; Vu & Peters, 2021; Webb & Chang, 2020) in a language.

In an extensive reading study, Brown et al. (2008) examined how reading, listening, and reading while listening affected the incidental learning of 84 pseudowords in three graded

readers (28 pseudowords in each) by 35 L2 learners of English. The researchers reported that both reading and reading while listening resulted in similar vocabulary gains in learning the meaning of new words, and both were superior to listening. In other words, being exposed to written input alone were equally effective in promoting the learning of new words as being exposed written and aural input together. However, in this study, participants were informed about the vocabulary tests, which might have made the learning more intentional than incidental.

Malone (2018) also investigated the effects of input mode along with those of frequency of occurrence on the learning of 32 low-frequency English words embedded in four stories (eight words in each) in an incidental learning condition. Eighty university-level L2 English speakers were randomly assigned to four groups that differed in terms of input modality and frequency: (1) reading-only with two exposures, (2) reading-only with four exposures, (3) reading-while-listening with two exposures, and (4) reading-while-listening with four exposures. Although these participants were informed about the comprehension posttests, they were not warned about the vocabulary posttests. The vocabulary gains were measured using one form-recognition test and one form-meaning connection test. The results showed that reading-while-listening was effective in increasing the knowledge of new form-meaning connections regardless of how many times these words were encountered in the text. However, reading-while-listening led to higher form-level recall than reading-only only when words were encountered less than four times in the text. In other words, reading-while-listening was particularly advantageous for linking the form and meaning of new words when the frequency of exposure was less than four times. In a partial replication and extension of Malone (2018), Chen (2021) also reported higher learning gains for the participants who read and listened to the treatment texts simultaneously. Moreover, Chen also reported statistically significant effects of self-rated L2 listening and L2 reading

proficiencies on the learning of form by the reading-while-listening group, suggesting that reading-while-listening induce a heavier cognitive load.

Teng (2018) compared the effectiveness of reading-only and reading-while-listening in the learning of four aspects of vocabulary knowledge: form recognition, grammar-functions recognition (i.e., parts of speech), meaning recall, and collocation recognition. Participants were 60 university-level EFL students with low-proficiency in English, and they were randomly assigned to reading-only and reading-while-listening groups. The treatment text was a level 2 graded reader for elementary learners, *The Love of a King*. Twenty-four real words in this graded reader were replaced with pseudowords, and these pseudowords were categorized into four frequency groups according to the number of the encounters: 1–2 times, 4–5 times, 9–10 times, and 14–16 times. The results showed that in both conditions, participants had highest gains for the form recognition, followed by grammar-functions recognition, meaning recall, and collocation recognition, respectively. Moreover, reading-while-listening led to higher vocabulary gains than the reading-only on all four aspects of vocabulary knowledge. This study provides a bigger picture regarding the effects of reading-only and reading-while-listening as it explores vocabulary learning at different sensitivity levels: form and meaning, part-of-speech, and collocational use. Moreover, it builds on the study by Malone (2018) by showing that reading-while-listening has beneficial effects on vocabulary learning even for lower-proficiency English learners.

The majority of the empirical studies reviewed above pointed to an advantage of bimodal input in achieving higher vocabulary learning gains, and there have been various arguments and theories regarding how and why simultaneous exposure to aural and written input can be facilitative for L2 vocabulary learning. With bimodal input, L2 learners get exposed to both the

orthographic and phonological forms of new words. Therefore, the simultaneous presentation of written and aural input may result in deeper processing of the input (Craik & Lockhart, 1972), which in return enhances the creation of the form-meaning links for novel words (Long, 2017; Malone, 2018). Researchers also argued that in incidental learning conditions, bimodal input can direct learners' attention to words (Malone, 2018). Long (2017) considered aural enhancement in bimodal input as a way of speeding up detection of target words in written input, i.e., noticing at the level of attention. Long also argued that bimodal input can be used as a way to keep learners on task and while doing that, it can also "reduce the likelihood that they will switch from incidental to intentional learning, as it is hard for them to 'keep up' with the spoken version if they pause to try consciously to figure out the meaning of a particular word or collocation" (p. 34). In other words, L2 learners' reading behaviors of new words can be different when they are exposed to bimodal input compared to those of unimodal input. However, there are still no concrete confirmations for these arguments in the literature as no studies have examined the amount of attention that L2 learners allocate to new words in a bimodal input as well as their intentionality levels so far. To my knowledge, the only published study that examined the overall reading patterns in bimodal input was conducted by Conklin et al. (2020).

Conklin et al. (2020) investigated the reading patterns of L1 and L2 speakers of English in reading and reading while listening to two 1500-word stories in English using eye tracking. The results indicated moderating effects of aural input in the reading while listening condition on the reading speed of both L1 and L2 speakers of English. In other words, participants, L1 speakers in particular, read significantly slower in the reading while listening condition than in the reading only condition (e.g., total reading time for L1 speakers: 322.77 ms for reading while listening, vs. 242.90 ms for reading only; total reading time for L2 speakers: 325.04 ms for

reading while listening vs. 336.10 ms for reading only). Despite the decrease in reading speed in the reading while listening condition, however, there was still a discrepancy between the participants' gazes and the audio. To be specific, participants' eye movements were ahead of the audio. This discrepancy was larger for L1 speakers and high proficiency L2 speakers than low proficiency L2 speakers. In the light of these findings, the researchers speculated that seeing the written form of words before hearing the audio can help L2 learners link the form and meaning of the words. However, the misalignment between visual and auditory input, particularly for high proficiency L2 learners, may cause L2 learners to process two different words simultaneously, which in turn may hinder comprehension (and possibly vocabulary learning).

Although Conklin and colleagues' (2020) study provided insights for the reading behaviors of L1 and L2 speakers in the presence of audio, there are still several unanswered questions regarding the effects of bimodal input on the reading and learning of new words in a text. With regard to cognitive processes in and vocabulary learning from the bimodal input, the key questions that remain to be answered are the following: (1) how does bimodal input influence L2 learners' attention to novel words in a text?; (2) how does bimodal input influence L2 learners' awareness of novel words in a text?; and (3) is bimodal input more effective than unimodal input in developing nondeclarative knowledge of novel words (i.e., the ability to fluently retrieve the form and meaning of newly learned vocabulary spontaneously) along with the development of declarative knowledge (i.e., conscious, verbalizable knowledge of a word)? Exploring these issues is of paramount importance for laying a foundation for a broader, more comprehensive view of bimodal input in incidental vocabulary learning.

Other Factors Affecting Incidental Vocabulary Learning

Incidental vocabulary learning from unimodal or bimodal input is a complex process that

is affected by a variety of learner-related and item-related factors. In this study, the influence of some of the most prominent factors in incidental vocabulary learning were investigated.

Item-Related Factors

One item-related factor that influences the learning of vocabulary in incidental learning conditions is the frequency of occurrence (Brown et al., 2008; Elgort & Warren, 2014; Elgort et al., 2018; Godfroid et al., 2018; Malone, 2018; Mohamed, 2018; Tekmen & Daloğlu, 2006; Vidal, 2011; Webb, 2007; Zahar et al., 2001). The majority of studies reported increased learning gains for items that were encountered multiple times in a text. For example, Webb (2007) investigated the effects of frequency (1, 3, 7, or 10 encounters) while controlling for the contexts the words appeared and reported higher learning gains for L2 learners' who encountered target words more times in different informative contexts than the ones who encountered the words fewer times. Vidal (2011) also reported frequency of occurrence to be the best item-level predictor for the vocabulary learning gains from reading. Frequency of occurrence accounted for 47% of variance in the learning gains from reading; the predicting effects of frequency of occurrence was a little bit lower for learning from listening with 24%. In a meta-analysis, Uchihara et al. (2019) reported a medium effect size for frequency of encounters (r = .34), and this effect was larger for unimodal input (r = .41) than bimodal input (r = .28). In addition to studies focusing on the learning outcomes, recent eye-tracking studies examining incidental vocabulary learning from reading (e.g., Godfroid et al., 2018; Elgort et al., 2018; Mohamed, 2018; Pellicer-Sánchez, 2016) demonstrated that that after eight to ten encounters, novel words are read faster and similar to familiar words. These findings suggest that incidental vocabulary learning is incremental, and frequency of occurrence in an input affects the learning gains significantly. However, no studies examined how frequency of occurrence influences the

processing of target words in bimodal input.

Another item-related factor is the target words' part of speech. Previous research has shown differences in learning gains of nouns, verbs, or adjectives (Ellis & Beaton, 1993; Godfroid et al., 2018; Puimège & Peters, 2019; Van Zeeland & Schmitt, 2013). In an intentional vocabulary learning study, Ellis and Beaton (1993) indicated that for L2 German learners, nouns were easier to learn than verbs when participants were exposed to the written form of German words and their English translations. Van Zeeland and Schmitt (2013) extended these findings for the learning from aural input. After listening to four passages, L2 English speakers were found to learn the form and meaning of target nouns better than those of verbs and adjectives. Godfroid et al. (2018) also examined the effects of part of speech in an incidental vocabulary learning condition. Differently from Van Zeeland and Schmitt, participants in Godfroid and colleagues' study read a novel containing 29 Dari words. Findings showed that the meanings of nouns were recalled better than the meanings of other words, but part of speech was not a significant predictor of form recognition. Differently from previous studies, Puimège and Peters (2019) reported that part of speech was not significant predictor of learning gains from audiovisual input. However, how part of speech influences the learning of form and meaning of target words from bimodal input has not been examined yet.

Concreteness of the target words is also a factor that affects vocabulary learning (Crossley et al., 2016; de Groot, 2006; Elgort & Warren, 2014; Puimège & Peters, 2019; Van Zeeland & Schmitt, 2013). For instance, in an intentional vocabulary learning study, de Groot (2016) indicated that L2 learners had higher meaning recall scores for concrete words (69.4%) than abstract words (55.9%). Moreover, the effects of concreteness of learning was larger for the earlier stages of learning. In an incidental vocabulary learning study from reading a nonfiction

book, Elgort and Warren (2014) also reported that the meanings of more concrete words were recalled better than the ones with less concrete words. Also, concreteness was a significant predictor of implicit knowledge development of the target words, measured with a primed lexical decision task. Using aural and audiovisual input respectively, Van Zeeland and Schmitt (2013) and Puimège and Peters (2019) also reported concreteness to be a significant predictor of learning the form and meaning of L2 single words. The influence of concreteness in learning new words from bimodal input still remains unclear.

Learner-Related Factors

The main learner-related factor that has been shown to affect vocabulary learning in incidental learning conditions is learners' prior vocabulary knowledge and their vocabulary size (Elgort & Warren, 2014; Horst et al., 1998; Tekmen & Daloğlu, 2006; Webb & Chang, 2015; Zahar et al., 2001). For example, Horst et al. (1998) investigated the relationship between L2 learners' vocabulary size, shown by their scores on the Vocabulary Levels Test (Nation, 1983), and their learning gains through reading and listening to a graded reader measured. The researchers reported a medium-sized positive relationship between the VLT scores and the learning gains on a form recognition test. That is, the chances of learning new words from reading and listening was higher for the participants with larger vocabulary size. In a longitudinal extensive reading study, Webb and Chang (2015) examined whether L2 learners' vocabulary size, measured with the bilingual version of the VLT (Schmitt et al., 2001), influences the learning gains of 60 L2 English learners in Taiwan when they read and listen to multiple graded readers over a longer period of time. Webb and Chang used a matching test (L2 word form – L1 meaning) to measure the learning gains and reported a positive relationship between participants' vocabulary size and their learning gains on this matching test. In other

words, participants with larger vocabulary size had higher vocabulary gains than the ones with smaller vocabulary size. Elgort and Warren (2014) extended these findings to the relationship between vocabulary size and learning gains on a meaning recall test. Using an authentic expository text for learning, Elgort and Warren (2014) indicated that vocabulary size, indicated by the scores on the Vocabulary Size Test (Nation & Beglar, 2007), was a statistically significant predictor of L2 learners' learning gains. They also reported that learners with larger vocabulary size were likely to learn the meanings of unknown words after fewer encounters than the ones with smaller vocabulary size. Taken together, these studies suggest a positive effect of prior vocabulary knowledge in learning unknown words in incidental learning conditions. One reason for this can be the increased comprehension levels for learners with larger vocabulary size (Hu & Nation, 2000; Schmitt et al., 2011). L2 learners with larger vocabulary size can comprehend the texts better and more easily, which in turn can allow them to allocate more attention to unknown words in the text. It is also more likely for these learners to better understand the surrounding words, which can help with inferring the meaning of unknown words.

Related to frequency of exposure, L2 learners' summed total reading times on target words during reading is another factor that can influence the learning gains (Godfroid et al., 2013; Godfroid et al., 2018; Mohamed, 2018; Pellicer-Sánchez, 2016). For example, Godfroid et al. (2013) reported a positive relationship between the total reading times on target pseudowords in a list of sentences and the scores in a multiple-choice, gap-fill form recognition test. The similar positive relationship between the summed total reading times and learning gains were also reported by Mohamed (2018) for the form recognition, meaning recognition, and meaning recall tests. The positive effects of attention to target words while reading on the learning gains were found only for participants' ability to recall the meanings of target pseudowords in a study

by Pellicer-Sánchez (2016) showing a weaker relationship between attention and learning gains compared to Mohamed (2018). Godfroid et al. (2018) also reported a positive effect of summed total reading times on recognizing and recalling the meanings of the target Dari words in the reading text. The results from these studies pointed that increased attention to target words while reading can influence the learning target words, in particular recognizing and recalling their meanings. However, a few studies failed to find any relationship between attention to target words and the learning gains (e.g., Elgort et al., 2018), showing that the strength of the relationship between attention and learning may not be stable across different learning environments. None of the studies investigated the relationship between eye-movements and learning gains from bimodal input.

In addition to prior vocabulary knowledge, vocabulary size, and attention levels of L2 learners, different affective factors can also play a role in the vocabulary acquisition process (Hulstijn & Laufer, 2001). Two factors that can interact with the L2 vocabulary learning are L2 learners' level of enjoyment and their topic interest. Both of these factors have been shown to be significant predictors of L2 learning (e.g., Li & Wei, 2022), L2 comprehension (e.g., Erçetin, 2010), and L2 vocabulary learning. For example, Lee and Pulido (2017) examined the vocabulary learning gains of Korean L2 English learners from reading two passages with different topics. The results indicated that the form and meaning of target words that were encountered in the passage with a more interesting topic were recognized better in the vocabulary posttests than the ones encountered in the less interesting passage. Elgort and Warren (2014) also examined the predictive effects of interest and enjoyment levels. In their studies, L2 learners who had higher levels of interest and enjoyment for the reading text were found have higher vocabulary learning gains than the ones with lower interest and enjoyment levels.

The Present Study

In the light of the literature reviewed in this chapter, in this study, I compared the effects of unimodal and bimodal input in learning vocabulary in an incidental learning condition. There are three overall goals of this study: first, to investigate the alignment between the hypotheses regarding the facilitative effects of bimodal input in processing a written input, as theoretically conceptualized vocabulary researchers, and the empirical evidence, as provided in this study; second, to examine the potential of unimodal and bimodal input in developing L2 learners' nondeclarative vocabulary knowledge along with that in developing their declarative knowledge; third, to explore the effects of different learner- and item-related factors on the processing of and learning from unimodal and bimodal input. For the first goal, I investigated the attention and awareness levels of L2 English learners while reading (unimodal input) or reading while listening (bimodal input) to a graded reader using eye tracking and verbal retrospective reports. For the second goal, I used a set of vocabulary measures tapping into different types and aspects of vocabulary knowledge. For the third goal, I focused on the effects of three learner-related factors (vocabulary size, enjoyment level, and interest level) and three item-related factors (frequency of occurrence, part of speech, and concreteness). I added participants' summed total reading times as a predictor of learning from unimodal and bimodal input as well. This study addresses the shortcomings in the previous studies by triangulating data with the use of complementary data collection instruments. The results of this study will adduce evidence for the hypothesized benefits of bimodal input for processing and learning of new vocabulary items by considering the possible effects of learner- and item-related factors. The research questions that guide my study and the hypotheses for each of these questions are given below.

The first set of questions were related to the *attention* levels of L2 learners while reading

the treatment text with or without audio:

(1a) How does input modality and repeated exposure affect L2 learners' attention to target words while reading?

Hypothesis (1a.1): Based on the arguments regarding the attention-directing feature of bimodal input to novel words in a text (Long, 2017), participants in the reading while listening group were predicted to have longer fixation durations on the target words than the participants in the reading only group.

Hypothesis (1a.2): Participants in the reading only group were predicted to spend less time fixating on target words from the first exposure to the last exposure. Participants in the reading while listening group were predicted to show similar fixation times across exposures.

(1b) Do learner-related factors (vocabulary size, interest level, and enjoyment level), and item-related factors (part of speech and concreteness) predict L2 learners' attention to target words while reading?

Hypothesis (1b.1): Participants with higher vocabulary size were expected to have shorter fixation durations on the target novel words than participants with lower vocabulary size.

Hypothesis (1b.2): Participants with higher interest and enjoyment levels were expected to spend more time looking at the target words in the reading text than participants with lower interest and enjoyment levels.

Hypothesis (1b.3): Target pseudowords replacing nouns were expected to receive shorther fixation durations than target pseudowords replacing verbs and adjectives.

Hypothesis (1b.4): Target pseudowords replacing more concrete words were expected to receive longer fixation durations than target pseudowords replacing less concrete words.

The second set of questions were related to the *awareness* levels of L2 learners while

reading the treatment text with or without audio:

(2a) How does input modality affect L2 learners' awareness of target words while reading?

Hypothesis (2a): Based on the arguments regarding the attention-directing feature of bimodal input to novel words in a text (Long, 2017), participants in the reading while listening group were predicted to have greater awareness of the target words than the participants in the reading only group.

(2b) Do summed total reading times, learner-related factors (vocabulary size, interest level, and enjoyment level), and item-related factors (frequency of occurrence, part of speech, and concreteness) predict L2 learners' awareness of target words while reading?

Since previous research showed a positive relationship between attention and awareness, I expected to find similar effects of learner-related and item-related factors on participants' awareness levels.

Hypothesis (2b.1): Increase in summed total reading times were predicted to influence awareness levels positively.

Hypothesis (2b.2): Participants with higher vocabulary size were predicted to have greater awareness of the target novel words than participants with lower vocabulary size.

Hypothesis (2b.3): Participants with higher interest and enjoyment levels were expected to have greater awareness of the target words than participants with lower interest and enjoyment levels.

Hypothesis (2b.4): Target words with higher frequency of exposure were expected to induce greater awareness than target words with lower frequency of exposure.

Hypothesis (2b.5): Target pseudowords replacing nouns were expected to induce greater awareness than target pseudowords replacing verbs and adjectives.

Hypothesis (2b.6): Target pseudowords replacing more concrete words were expected to induce greater awareness than target pseudowords replacing less concrete words.

The third set of questions were about L2 learners' *learning gains on the vocabulary tests* from the treatment text that they read with or without audio.

(3a) How does input modality affect the development of declarative knowledge of target words?

Hypothesis (3a): Participants in the reading while listening group were predicted to have higher learning gains on all three tests assessing the declarative knowledge development (i.e., form recall, meaning recall, and meaning recognition) than the participants in the reading only group.

(3b) Do summed total reading times, learner-related factors (vocabulary size, interest level, and enjoyment level), and item-related factors (frequency of occurrence, part of speech, and concreteness) predict the learning gains on the declarative knowledge tests?

Hypothesis (3b.1): Higher summed total reading times were expected to result in higher learning gains on all three declarative knowledge tests.

Hypothesis (3b.2): Participants with higher vocabulary size were expected to have higher learning gains on all three tests of declarative knowledge than participants with lower vocabulary size.

Hypothesis (3b.3): Participants with higher interest and enjoyment levels were expected to have higher learning gains on all three declarative knowledge tests than participants with lower interest and enjoyment levels.

Hypothesis (3b.4): Participants were expected to have higher learning gains for target words with higher frequency of exposure than words with lower frequency of exposure on all three declarative knowledge tests.

Hypothesis (3b.5): Participants were expected to have higher learning gains for target words replacing nouns than target words replacing verbs and adjectives on all three declarative knowledge tests.

Hypothesis (3b.6): Participants were expected to have higher learning gains for more concrete target words than less concrete target words on all three declarative knowledge tests.

(3c) How does input modality affect the development of nondeclarative knowledge of target words?

Hypothesis (3c): Participants in the reading while listening group were expected to have higher nondeclarative knowledge development (i.e., quicker reading of target pseudowords than nonwords in the sentence reading test) than the participants in the reading only group.

(3d) Do learner-related factors (vocabulary size, interest level, and enjoyment level) predict the learning gains on the nondeclarative knowledge test?

Hypothesis (3d.1): Higher vocabulary size were expected to affect the nondeclarative knowledge development positively.

Hypothesis (3d.2): Higher interest and enjoyment levels were expected to affect the nondeclarative knowledge development positively.

CHAPTER 3: METHODOLOGY

Overview of the Research Design

This study had a between-participants design with two conditions: reading only (Group 1) and reading while listening (Group 2). The two conditions differed in terms of the available input modalities. In the reading only condition, participants had access to the written input only, and in the reading while listening condition, participants were exposed to both written and aural input. Apart from the difference in input modalities, both groups completed the same data collection instruments. See Table 1 for the operationalization of all variables in the study.

Table 1

Operationalization of All Variables in the Study

	Variables	Operationalization	Instruments	Type
Cognitive processes	Attention	Eye fixations on the target words while reading the treatment text	Four eye-tracking measures	Continuous
	Awareness and intentionality	Participants' retrospective reports of their cognitive processes during reading	Retrospective verbal interviews	Categorical
Vocabulary learning measures	Form recognition	Accuracy of form recognition	Form recognition test	Categorical
	Meaning recall	Accuracy of meaning recall	Meaning recall test	Categorical
	Meaning recognition	Accuracy of meaning recognition	Meaning recognition test	Categorical
	Fluency of word retrieval	Eye fixations on target pseudowords compared with those on real words and nonwords	Sentence-reading posttest – two eye tracking measures	Continuous
Learner-related factors	L2 proficiency	Receptive vocabulary size	14k Vocabulary Size Test	Continuous
	Enjoyment and interest levels	Participants' enjoyment and interest ratings for the reading text	Their enjoyment and interest levels on a 100-point Likert scale	Continuous

Table 1 (cont'd)

	Variables	Operationalization	Instruments	Type
Item-related factors	Frequency of Occurrence	Target words' number of occurrences in text	-	Continuous
	Part of speech	Target words' part of speech	-	Categorical
	Concreteness	Target words' concreteness levels on a 5-point scale (taken from Brysbaert et al., 2014)	-	Continuous

Participants

Participants were 63 L2 English speakers at a Midwestern university in the United States. Using the data request form on the university's office of the registrar website, participants were recruited via recruitment emails. The following three inclusion criteria were applied to the eligible participants: (1) their first language and reported dominant language was a language other than English, (2) they attended primary and secondary education in a non-English speaking country, and (3) their length of residence in an English-speaking country was less than five years. All participants were affiliated with the university; 14 were undergraduate students, 39 were graduate students, and 10 were visiting scholars. All participants were pursuing a degree in fields other than language-related disciplines. The participants came from different L1 backgrounds (see Appendix A), and they all had normal or corrected-to-normal vision and normal hearing. Their mean vocabulary size was 10298 (SD = 1738) out of 14000 word families, measured using the Vocabulary Size Test (Nation & Beglar, 2007). The results of the VST indicated that the participants had advanced proficiency in English as PhD students who are L2 speakers of English are estimated to know around 9000 word families (Nation, 2012). Table 2 summarizes the demographic information for the participants.

Table 2

Demographic Information of the Participants in the Study

	Reading Only				Reading While Listening			
Variable	Mean	SD	Min.	Max.	Mean	SD	Min.	Max.
Age (years)[1]	28.07	7.93	18	53	26.90	6.10	18	41
LOR (months)	18.69	19.07	3	60	12.40	11.68	3	48
Vocabulary size	106.54	15.26	79	135	102.52	13.81	77	130
Self-rated proficiency[1]								
Overall	4.59	0.91	3	6	4.60	0.72	3	6
Reading	4.76	0.99	2	6	4.60	0.97	3	6
Listening	4.62	1.08	2	6	4.67	0.84	3	6

Notes. LOR = Length of residence; Min. = Minimum; Max. = Maximum; [1] Age at the time of testing; [2] participants self-reported on a scale where 1 represented "beginner" and 6 represented "Native-like"

All participants were assigned to one of the two reading groups that differed in terms of modality: reading only (unimodal) and reading while listening (bimodal). Participants with different profiles were evenly distributed to two groups. There were 31 participants in the reading only group and 32 participants in the reading while listening group. Four participants were removed from the final sample for two reasons: two participants experienced technical problems during data collection (i.e., the eye-tracking camera calibration was not successful) and two participants scored very poorly on the reading comprehension test (accuracy lower than 60%). Thus, the final sample consisted of 29 participants in the reading only group and 30 participants in the reading while listening group. Two participants (one from each group) did not

complete the Vocabulary Size Test; however, these participants were still kept in the final sample. The two groups were comparable in terms of their mean vocabulary size, $t = 1.04$, $df = 54.02$, $p = .30$, and their self-rated proficiency in English, $W = 436$, $p = .99$.

All participants received monetary compensation for their participation. Ethical clearance was obtained from the university's Institutional Review Board (IRB).

Materials

The Reading Text

The reading text used in this study was the Level 4 Pearson graded reader *The Time Machine*. This graded reader is an adapted version of the classic science fiction novel by H. G. Wells. The text is around 21000 words, but only the first seven chapters (9761 words) were used in this study. I chose this reader as the reading text in the study because I believed that it was interesting and engaging.

Twenty-four real words in the text were replaced with pseudowords. Then, the text was slightly adapted to increase the frequency of occurrence of a few pseudowords in the text. Moreover, all sentences with the target pseudowords were checked and adapted when necessary to ensure that target pseudowords never occurred as the first or last word in a sentence as well as before or after a punctuation, as these sentence positions were found to affect eye movements of readers (i.e., a tendency to skip or overlook the words in these areas) (Godfroid, 2020; Hirotani et al., 2006; Rayner & Pollatsek, 2006). Since the existence of multiple pseudowords in one sentence can decrease the possibility of guessing from context, several sentences in the text were also slightly modified so that only one target pseudoword occurred in one sentence. Finally, the chapters where each target pseudoword occurred were checked, and one or two more instances of target pseudowords were added to the first or last chapters to ensure that participants were

exposed to all target words on both data collection days at least one time (see Joseph et al., 2014 for an example study indicating the effects of acquisition order on incidental vocabulary learning). The entire adapted reading text can be found in Appendix B.

The text was analyzed with VocabProfile on the Compleat Lexical Tutor (https://www.lextutor.ca/vp/comp/) (Cobb, n.d.) to investigate the corpus frequency of the words used in the text. This tool provides information about the corpus frequency of all words in a text by breaking the text into 25 frequency bands and calculating the lexical coverage percentage for each band based on the BNC-COCA word lists. For example, the words in the K-1 band consist of the 1000 most frequent words in English. Before analyzing the text, all the target pseudowords and proper nouns in the text were categorized as off-list words. The vocabulary profile of the reading text is given in Table 3. As can be seen in this table, around 97% of the words in the text (without proper nouns and the target pseudowords) are in the most frequent 3000 word families in the BNC/COCA word frequency list. In other words, the words that make up the reading text are high frequency words, and a reader who knows the first 3000 word families in English can understand 97% of this text. Considering the high vocabulary size of the participants and the less demanding lexical profile of the reading text, the reading text was suitable for comprehension for the L2 speakers of English in the current study.

Table 3

Vocabulary Profile of the Reading Text

	K-1	K-2	K-3	K-4	K-5	Others	Off-list	Total
Number of tokens	8887	485	94	51	7	27	210	9761
Cumulative tokens (%)	91.1	96.1	97.1	97.6	97.7	97.9	100	

The Audio

For the reading while listening group, the audio to be used was recorded by an L1 female speaker of American English. To find the optimum audio speech rate for the current study, I checked the speech rates used in the previous studies and piloted different speech rates with three L2 English speakers.

The audio speech rates in the studies that examined vocabulary learning from reading while listening were different from one another (see Table 4 for a summary of these studies). For example, in Malone (2018), the speech rate of the audio was 120 – 140 words per minute (i.e., 2 to 2.3 words per second). Malone argued that this speech rate was appropriate for examining the effects of audio on learning without being unnaturally slow. Although Malone did not report the proficiency levels of the participants in the study, all participants were ESL learners at two Intensive English Programs in the United States, so they had a lower proficiency in L2 English. In a reading study with advanced learners of L2 English, on the other hand, Conklin et al. (2020) used audios with a speech rate of 210 words per minute (i.e., 3.5 words per second). They justified their speech rate by citing the study by Griffiths (1990) who reported that for low-intermediate level English learners the maximum speech rate should be 190 – 200 words per minute (i.e., 3.3 words per second). Speech rates faster than this rate were found to reduce the comprehension levels. Since their participants' proficiency level was higher than low-intermediate, Conklin et al. argued that 210 words per minute was a suitable speech rate for their participants for comprehending the text without being distracted by a slow speech rate.

Table 4

Summary of the Studies Investigating Reading While Listening

Study	Reading Text	Participant Profile	Speech Rate
Brown et al.	Three graded readers containing 28 target words (84 target words in total)	EFL learners at a university in Japan (proficiency not reported)	93 words per minute
Chen (2021)	Same as Malone (2018) – low frequency words replaced with pseudowords	High-intermediate level L2 English learners (self-reported)	120 – 140 words per minute
Conklin et al. (2020)	Two 1500-word long stories (no target words)	L2 English speakers with advanced proficiency in the UK	210 words per minute
Dang et al. (2022)	An academic text containing 19 target collocations	EFL learners in China (proficiency not reported)	Not reported
Malone (2018)	Four stories containing 8 low-frequency target words (32 target words in total)	ESL learners at two intensive English programs in the US	120 – 140 words per minute
Teng (2018)	A graded reader containing 24 pseudowords as target items	EFL learners in China (proficiency not reported)	90 words per minute
Vu & Peters (2021)	Three graded readers containing 32 target collocations	Pre-intermediate level EFL learners	Not reported
Webb and Chang (2020)	A graded reader containing 17 target collocations	Low to high-intermediate EFL learners	120 words per minute
Webb et al. (2013)	A graded reader containing 18 target collocations	EFL learners at a university in Taiwan (proficiency not reported)	140 – 170 words per minute

The proficiency profile of the participants in the current study were similar to the ones in Conklin et al. (2020). However, the main goals of the two studies were different. In Conklin et al.'s study, achieving a high reading comprehension was the main aim, whereas in the current study the main goal was incidental vocabulary learning although participants were not informed about this goal. Therefore, I considered that the speech rate used by Conklin and colleagues (i.e., 210 words per minute) could be a bit fast for learning the target pseudowords incidentally as participants would have no or minimal opportunities to fixate on target words due to the fast speed of the audio. Instead, I predicted that 170 - 180 words per minute (i.e., 3 words per second) could be a better speech rate, as this speech rate would be neither too slow nor too fast for high proficiency L2 English learners. So, participants could still pay attention to target pseudowords while comprehending the text without being distracted by an unnaturally slow audio. Therefore, I piloted the experiment with three different audio speech rates to find the optimum speech rate for the current study: 120 – 140 words per minute, 170 – 180 words per minute, and 200 – 210 words per minute. The speech rates of the audio were modified using Audacity.

Three pilot participants (L2 English speakers with a similar profile to main participants in the study) were asked to read and listen to one chapter from the reading text with each speech rate. As expected, their eye movements were ahead of the audio with the first speech rate (120 – 140 words per minute), and participants stated that the audio was too slow for them. Participants' eyes were more aligned with the audio when the speech rate was either 170 – 180 words per minute or 200 – 210 words per minute, and they stated that they liked the pace of the audio with both speech rates. I also checked their mean reading rates without the audio. Their mean reading rates ranged from 150 words per minute to 240 words per minute. Participants with slower reading rates reported that they can read faster with the audio and achieve a similar

comprehension whereas the participant with higher reading rate reported that 170 – 180 words per minute was not too slow for her. Therefore, I decided that 170 – 180 words per minute would be the optimum speech rate for the current study as it would not be too fast for slow readers or too slow for fast readers. This speech rate could also provide considerable latitude for participants in attending to particular sentences and words in the reading text whenever they wanted to.

Target Words

Twenty-four words with a frequency of occurrence ranging from 2 to 16 in the adapted reading text were replaced by pseudowords. The use of pseudowords allowed controlling for prior knowledge and out-of-experiment learning. The pseudowords' parts-of-speech remained the same, and they replaced nouns, verbs, or adjectives (nine nouns, eight verbs, and seven adjectives). All the pseudowords were selected from the English Lexicon Project (https://elexicon.wustl.edu/) (Balota et al., 2007). This project contains information about the descriptive lexical statistics (word length, orthographic neighbors, bigram frequency) and behavioral data (i.e., reaction times (RTs) and accuracy in a lexical decision task) of 40,481 real English words and of 40,481 nonwords which were created by changing one or two letters in the real words stimuli. Using this list, I chose five to six letter-long pseudowords which are comparable in the number of orthographic neighbors and mean bigram frequency. Choosing pseudowords that are similar in terms of these lexical statistics is important to ensure similar word processing (Keuleers & Brysbaert, 2010). Moreover, choosing words with similar word length (i.e., five to six letters) is essential because previous research (e.g., Godfroid et al, 2018) has shown that eye movements are influenced by word length. Finally, I checked the information about mean reaction times (RTs) in a LDT for each pseudoword because this information has

been reported to be affected by the "wordlikeness" of a pseudoword (Keuleers & Brysbaert, 2010), and choosing pseudowords with an RT closer to the general mean (which was reported to be 855.78 in the English Lexicon Project) can ensure that the selected pseudowords are neither too similar to nor too different from real words. The target pseudowords, their characteristics obtained from the English Lexical Project, and the words they replace are given in Table 5.

I also investigated the predictability of words in the text, particularly their predictability levels on their first occurrences with a pilot test. I prepared short excerpts for the first occurrences of each target pseudoword in the reading text. In all excerpts, target pseudowords were deleted, and seven L1 English speakers of English were asked to fill in each blank with a word that fits in the context. They were also instructed to rate how easy or difficult it was to guess the word from the context on a 10-point scale. Except for six target words, all target pseudowords were found to be guessable from their first occurrence. These six target words and the sentences they appeared in were modified, and the new contexts were piloted one more time with three different L1 English speakers. On average, the number of correct responses was high (M = .70, SD = .24), and the majority of words were reported to be easy to guess (M = 3.88, SD = 1.44). These findings indicated that it was possible for L1 English speakers to infer the meanings of the missing words from their first occurrences in the reading text. See Appendix C for more information about the predictability of each pseudoword on its first occurrence.

Table 5

Target Pseudowords and Their Characteristics

		Original word characteristics in text			Pseudoword characteristics			
Target Pseudoword	Original Word Replaced	Part of Speech	Frequency of Occurrence	Concreteness	Length (letters)	Orthographic neighbors	Bigram Frequency	Mean Reaction Time
hilder	seat	noun	8	4.58	6	4	2,322.00	848.208
mantil	intelligence	noun	3	2.24	6	3	2,878.00	875.273
plear	story	noun	9	3.30	5	4	2,243.25	860.152
fleak	danger	noun	6	2.68	5	4	1,508.75	871.75
rolley	world	noun	16	4.36	6	3	1742.00	848.931
spoll	mind	noun	9	2.50	5	4	1,204.75	853.545
mittle	face	noun	9	4.87	6	3	1,611.40	853.233
merse	group	noun	5	4.12	5	4	2,850.75	837.192
demuse	journey	noun	4	2.57	6	4	1,493.40	849.417
sorge	eat	verb	8	4.44	5	4	1,263.25	838.625
halker	explain	verb	8	1.97	6	4	2,275.00	846.926
thrine	apologize	verb	2	2.63	6	4	2,723.40	858.727
prome	think	verb	13	2.41	5	5	1,611.25	862.464
crimb	understand	verb	12	2.28	5	4	1,246.50	870.407
slair	wash	verb	2	4.35	5	4	1,044.25	832.519
tickel	recognize	verb	3	2.07	6	4	2,128.60	839.692
maive	speak	verb	3	3.70	5	5	1,343.25	803.519
vandy	great	adjective	10	1.81	5	5	1,468.00	879.633
fanish	important	adjective	2	2.14	6	5	1,896.60	841.68
tifled	green	adjective	5	4.07	6	3	2,730.80	858.333
merky	small	adjective	9	3.22	5	5	2,066.50	841.824
fanky	weak	adjective	5	2.79	5	5	1,072.50	850.345
croom	new	adjective	11	2.81	5	4	1,270.50	841.433
tartle	clear	adjective	7	3.55	6	3	1,936.20	845.143

Eye-Tracking Measures

Four eye-tracking measures, namely, *gaze duration, rereading time, regression path duration, and total reading time,* were used to examine participants' eye-movements on the target pseudowords words during reading. Gaze duration is the sum of all fixations on a word before the eyes move to another word. This measure is considered to be an early measure of processing (Godfroid, 2020), and when used in a reading study, it provides information about the initial stages of reading, particularly lexical access (Elgort et al., 2018). Late measures of processing include rereading duration, regression path duration, and total reading time. Rereading time refers to any non-first-pass fixations in an interest area. Regression path duration is the sum of all fixations on a word until the eyes move to the right of the word including the fixations made to previous words in the sentence (i.e., regressions to previous words). Since regression path duration provides information about the regressions made to the previous context after fixating on a target word, it can be an indicator of processing difficulty and the time it takes to overcome this difficulty (Godfroid, 2020). Because of this, regression path duration can reflect the word-to-text integration in reading (Elgort et al., 2018; Rayner & Pollatsek, 2006), and in this study, it is used to understand whether/how the participants use the preceding context to figure out the meaning of the target pseudowords. Total reading time is considered as another measure indicating word-to-text integration, and it refers to the sum of all fixations made on a word including regressions made back to the word (i.e., regressions in). This measure is a composite measure of gaze duration and rereading time. The main reason of examining the total reading time in this study is because it has been found to be a predictor of vocabulary learning gains in previous studies (e.g., Godfroid et al., 2018), and it was the only measure that distinguished between words that were less and more fully acquired in a sentence-reading posttest (Joseph et

al., 2014).

In their study focusing on the general reading patterns in reading only and reading while listening conditions, Conklin et al. (2020) reported shorter and fewer fixations on real English words in the reading only condition for L1 speakers of English. However, for L2 speakers of English, the two reading conditions elicited similar amounts and durations of fixations. The results of this study indicated that audio in reading while listening conditions influences the reading patterns of readers, in particular those of L1 readers, by promoting increased number and duration of fixations on words in a text. But for L2 speakers, who generally have more and longer fixations on words than L1 speakers, the effects of additional audio support may not be as pronounced. However, differently from reading familiar words, while reading a text containing novel words, the additional audio support in reading while listening conditions can have a distinct effect on readers' reading patterns. Additionally, in line with the results of previous incidental vocabulary learning studies, for the reading only group, I predict that both early and late measures on target pseudowords indicating lexical access and word-to-text integration will decrease from the first exposure to the last during the course of reading (Elgort et al., 2018; Godfroid et al., 2018; Joseph et al., 2014; Pellicer-Sánchez, 2016). However, for the reading while listening group, the same decreasing pattern may not be observed due to the audio support. As argued by Long (2017), if audio can direct participants' attention to target words while reading, participants' attention to target pseudowords may stay similar across exposures, showing a more stable reading pattern.

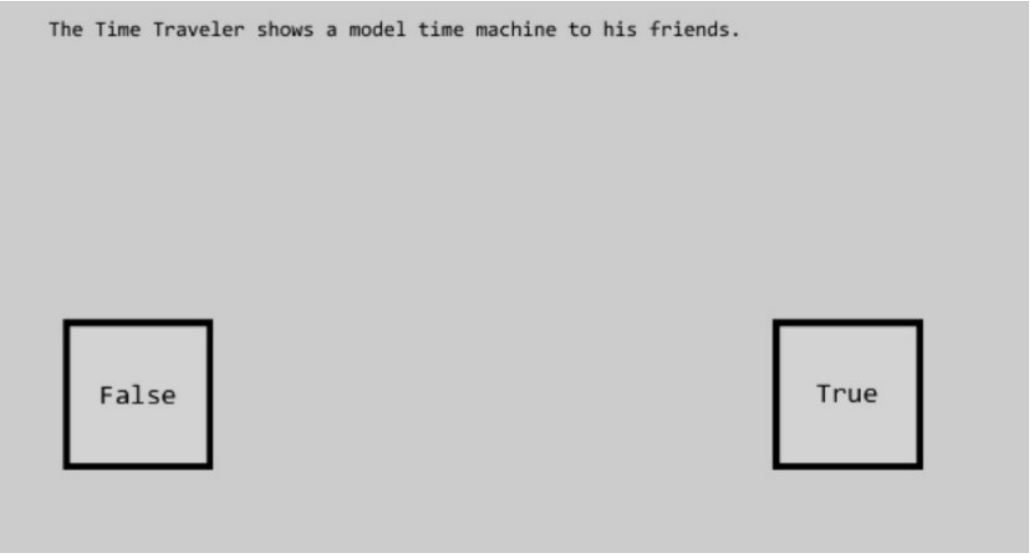
Comprehension Test

To encourage participants to read for meaning, participants were given comprehension questions after reading each chapter. There were four comprehension questions related to each

chapter (28 questions in total), and participants answered these questions before moving on to the following chapters. All questions were true/false statements (Appendix D). An example comprehension question is given in Figure 2. All items on the test were piloted with two L1 English and three L2 English speakers. Comprehension questions did not contain any of the target pseudowords. Participants were given one point for correct answers and zero points for incorrect answers. The data from this test were used only to see if participants completed the reading task according to the instructions given to them, i.e., read for comprehension.

Figure 2

Example Comprehension Question



Vocabulary Tests

The learning of the target pseudowords was examined at different sensitivity levels with the use of four tests: a sentence-reading test, a form recognition test, a meaning recall test, and a meaning recognition test. With the sentence-reading test, the aim was to assess participants' ability to fluently retrieve word meanings during real-time reading. With the form recognition test, the aim was to examine the initial stages of learning, i.e., the learning of form. Finally, with

the meaning recall and recognition tests, I assessed participants' ability to recall and recognize word meanings. All tests were given to participants at two times: immediately after the second reading session as immediate posttests and one week after the reading sessions as delayed posttests. The reliability of these tests were assessed using Cronbach's α coefficients and calculated using the *CTT* package (Willse, 2018) in R. All tests can be found in Appendix E.

Sentence reading test. The aim of the sentence-reading test was to compare the processing of the pseudowords encountered in the reading text with that of real words and nonwords (i.e., words participants never encountered before). Participants' eye movements for the target pseudowords were compared to those for real words and nonwords to examine the quality of knowledge established for the target pseudowords. The main assumption with this test is that reading times for words with strongly established representations should be shorter than words with no or weakly established representations. Therefore, if participants have established the form and semantic representations of the target pseudowords, they should be able to fluently retrieve the meanings of the target pseudowords and integrate them while reading a sentence, just like they do for higher frequency real words. In other words, participants' eye movements on the target pseudowords are expected to be similar to those of real words and different from (i.e., shorter fixations) those of nonwords.

For this test, 24 sentences were created, and each sentence had three versions which differed only in terms of the critical word, i.e., a target pseudoword, a real word, or a nonword in the sentences. As with pseudowords, real words and nonwords used in this sentence-reading test were selected from the English Lexicon Project (Balota et al., 2007), and all selected words and nonwords were matched to the target pseudowords in terms of word length. I also matched the pseudowords and nonwords in terms of their orthographic neighbors, mean bigram frequency,

and mean RT using the lexical characteristics and behavioral data information on the nonwords list retrieved from the English Lexicon Project. See Appendix F for the characteristics of each critical word used in the test.

All the created sentences were neutral sentences; there were no contextual cues in these sentences which could help infer the meaning of the word in the critical positions (i.e., pseudoword, real word, or nonword). Three versions of sentences were distributed to three experimental lists so that participants encountered the same sentence only once. Therefore, each list contained eight sentences with pseudowords, eight sentences with real words, and eight sentences with nonwords. In addition to these experimental items, there were 24 filler sentences and four practice sentences. In total, each list contained 52 trials (4 practice trials, 24 experimental trials, and 24 fillers). One of the practice sentences also contained a nonword. To keep participants' focus on comprehension during the experiment, one fourth of these sentences (12 sentences + one practice sentence) was followed with a comprehension question. The order of the question was random, so participants did not know which sentences were followed with a question. This randomization was done to keep participants from developing a task specific reading strategy. Moreover, to not draw participants' attention to the critical words in the sentences, the comprehension questions did not contain any of the critical words. Similar to the comprehension questions for the reading text, all comprehension questions in the sentence reading test were also in true/false format.

Participants were randomly assigned to each stimuli list, so each list was used with one third of the participants. Moreover, participants were tested with different experimental lists on the immediate and delayed posttest. For example, if one participant completed List 1 as the immediate test, they completed List 2 as the delayed test. An example distribution of three

sentences with critical items to three stimuli lists is given in Figure 3. Also, see Appendix E for the complete stimuli lists.

In the sentence reading test, all sentences were visually presented to the participants. Each sentence appeared in a separate screen, and participants read all sentences one by one for comprehension. After reading each sentence, participants pressed the SPACE key on the keyboard, and the following sentence appeared. Participants were instructed to read each sentence for comprehension while their eye movements were being recorded, and they were also informed about the presence of comprehension questions.

Figure 3

Example Critical Items and Experiment Lists in the Sentence Reading Test

Example Critical Items

	Sentence	Pseudoword	Real Word	Nonword
1	Kara saw a (critical position) next to the window.	hilder	plant	biddle
2	They always value (critical position) more than anything.	mantil	respect	coster
3	Jason listened to Mary's (critical position) before leaving the building.	plear	secret	snash

Example Experiment Lists

List 1	List 2	List 3
Kara saw a hilder next to the window.	Kara saw a plant next to the window.	Kara saw a biddle next to the window.
They always value respect more than anything.	They always value coster more than anything.	They always value mantil more than anything.
Jason listened to Mary's snash before leaving the building.	Jason listened to Mary's plear before leaving the building.	Jason listened to Mary's secret before leaving the building.

The main interest areas in the test were the critical words, i.e., pseudoword, real word, or nonword, in each experimental sentence. To investigate the possible processing differences between the real words, pseudowords, and nonwords in the sentences, I used one early measure (i.e., gaze duration) and one late eye tracking measure (i.e., total reading time). These measures are described in the "Eye-Tracking Measures" section above. If participants developed strong mental representations of form and meaning of the target pseudowords, I predicted to find similar reading times for pseudowords and real words. Additionally, I expected to find shorter reading times for pseudowords and real words than nonwords.

Form recognition test. This test assessed participants' ability to recognize the form of the target pseudowords encountered during reading. In this test, participants were given a randomized list of 24 target pseudowords and 24 distractors and were asked to choose the words they remember in the reading text. Distractors were the same as the nonwords used in the sentence reading posttest. Since the form recognition test was given after the sentence-reading test, participants were instructed explicitly to choose the words that they remembered seeing in the reading text and not the ones in the sentence-reading test. All words on this test were treated as separate test items, and the scoring was binary. All the correctly selected target pseudowords were given one point; all non-selected target pseudowords received zero points. Following the scoring procedures for the form recognition posttest used in Godfroid et al. (2018), participants were not penalized for selecting distractor items in the test. Overall, the mean of selected distractor items (false positives) was 3.57 (around 14%): 2.71 in the reading only group and 4.4 in the reading while listening group. The overall reliability of this test was .79 (for 24 target items, 59 participants, and two testing times), indicating good reliability. The test reliability for the reading only and reading while listening groups were .80 and .78 respectively.

Meaning recall test. This test assessed learners' ability to recall the meaning of the target pseudowords they encountered during reading. In the meaning recall test, participants were given a randomized list of the 24 target pseudowords, and they were asked to provide an explanation or synonym of these words in English. All target pseudowords were given without any contextual support. The test items only consisted of the 24 target words. See Figure 4 for an example and Appendix E for the test. Scoring of the test was completed in two steps. First, the researcher and a second rater discussed the meanings of each pseudoword based on the contexts they appeared in the reading text and created a list of acceptable definitions for each target pseudoword. Then, the researcher and the second rater coded the answers as correct and incorrect separately. All correct answers were rewarded one point, and all incorrect answers or "I don't know" answers were given zero points. There were no partial points. There were only seven instances of disagreement between two raters (out of 2832 data points), and the initial interrater reliability, measured using the *kappa2* function in the R package *irr* (Gamer et al., 2019), was very strong ($\kappa = 0.98$, $p < .001$) (Cohen, 1960). Later on, the raters compared all the seven instances of disagreement one by one and resolved all the conflicts, reaching 100% agreement. The overall reliability of the test was .68 for all 24 items, 59 participants, and two testing times, which was considered to be acceptable. The test reliability for the reading only group and the reading while listening group was .72 and .64, respectively.

Figure 4

Example Meaning Recall Test Item

Please have a look at the words and write their English definitions or synonyms. Then choose a percentage for your confidence level.

If you do not know the meaning of the word, please choose "I don't know".

	Word	**English definition or synonym**	**I don't know**	**How confident are you with your answer?** 0 - 50% = not very confident 60-70% = somewhat confident 70-80% = confident 80-90% = very confident 100% = absolutely certain
1	hilder		☐	
2	prome		☐	

Meaning recognition test. This test assessed participants' ability to recognize the correct meaning of target pseudowords they encountered during reading. In this test, participants were given the target pseudowords along with four options, and they were asked to choose the option that matched the meaning of the target pseudoword. The test items only consisted of the 24 target pseudowords that participants encountered in the reading text. To reduce the effect of guessing, each multiple-choice item also contained an "I don't know" option. The distractor options given in each multiple-choice question were plausible in the context of the reading text, and they were matched with the correct answer in terms of part of speech. Therefore, guessing the meaning of each pseudoword by looking at the options only was not possible. See Figure 5 for an example and Appendix E for the complete test. The scoring was binary: correct answers were awarded one point, and zero points were given for incorrect or "I don't know" answers. The overall reliability of the test was .71 (24 items, 59 participants, and two test times), which was acceptable. The reliability estimates by group were .67 for the reading only group and .74 for the reading while listening group.

Figure 5

Example Meaning Recognition Test Item

Look at the following target words and circle the word that has the closest meaning to the target word. Then choose a percentage for your confidence level.

If you do not know the meaning of the word, please choose "I don't know".

	Word	1	2	3	4	I don't know.	How confident are you with your answer? 0 - 50% = not very confident 60-70% = somewhat confident 70-80% = confident 80-90% = very confident 100% = absolutely certain
1	**mantil**	☐ intelligence	☐ material	☐ existence	☐ trick	☐	
2	**tifled**	☐ dusty	☐ green	☐ daily	☐ ordinary	☐	

L2 Proficiency Measure

Participants were asked to complete the Vocabulary Size Test (Nation & Beglar, 2007), which is a test for assessing L2 learners' vocabulary size in English, to measure their vocabulary knowledge and proficiency in English. This test consists of 140 multiple-choice vocabulary questions from the first 14000 word families. There are 10 questions for each frequency level. In each question, the vocabulary item is given in a neutral sentence, and participants are asked to select the option that best fits the meaning of the target item in the question out of four options. There is no time limit in completing this test. To calculate the participants' vocabulary size, the sum of their correct answers was multiplied by 100.

Background Questionnaire

To screen participants and to learn about their educational background and language use, participants were asked to complete a background questionnaire. This questionnaire consisted of three parts. The first part was about participants' demographic information and educational background. The second part contained questions about participants' language learning history and use. The third part focused on participants' English proficiency, and it contained a self-

assessment of their overall English proficiency, and their reading, writing, listening, and speaking skills in English on a 6-point scale. See Appendix G for the background questionnaire.

Retrospective Verbal Report

After completing the reading text on the second session, participants also took part in an interview with the researcher. The interviews were conducted in English, and they were audio-recorded. The aim of these interviews was to understand the ways participants interacted with unknown vocabulary in the text, and participants' reports were expected to be informative regarding their awareness of target pseudowords. The awareness levels of the participants in the current study were operationalized as their reported ability to recall a particular target word from the reading text.

I based the interview questions on the interview protocol in Godfroid and Schmidtke (2013) and the post-experiment survey in Godfroid et al. (2018) while considering the lexical processing strategies in Ender (2016). During this interview, I showed each target pseudoword to participants one by one and asked them a set of questions that were intended to tap into their awareness and recall levels. In addition to 24 target pseudowords, there were also six distractor items in the interview to control for the guessing levels of the participants during the interview. These six distractors were randomly selected from the distractor items used in the sentence reading test and form recognition test. During the interviews, except for two – three instances where participants made things up for the distractor words, participants stated that they did not recall reading these distractor words in the reading text. On several occasions, participants stated that they recalled reading these distractor words in the sentence-reading test or the form recognition test but not in the reading text. Very low false positive rates for the distractor items (i.e., participants making things up for distractor words) indicated that participants were good at

reporting their noticing and awareness levels for the target pseudowords that they encountered in the reading text. The questions that were used to probe participants' awareness and intentionality levels are given in Appendix H.

Following the coding scheme in Godfroid and Schmidtke (2013) and considering the concepts of noetic and autonoetic awareness, the reported awareness levels of the participants were coded into three categories:

1. When a participant reported not recalling reading a particular target word in the reading text, their reported awareness levels were coded as no awareness. The cases where participants reported being unsure were also coded as no awareness.
2. When a participant reported a particular word as familiar to them or reported recalling a particular target word somewhere in the reading text without stating their experiences with that particular word in the reading text, their reported awareness level was coded as noetic (or having a sense of familiarity).
3. When a participant reported recalling a particular target word somewhere in the reading text, and they also reported something they experienced while reading that word in the reading text, their reported awareness was coded as autonoetic (or recollection).

Around 30% of all the awareness data (data from 19 participants with a total of 456 data points) were coded with a second rater - a PhD student in the Applied Linguistics field. First, the main researcher shared the coding scheme containing several example excerpts with the second rater, and they discussed which participant reports can be considered as instances of no awareness, noetic awareness, and autonoetic awareness. Then, the researcher and the second rater coded the reports into these three categories separately. The following excerpts show

several examples for the classifications:

1. [no awareness]: target word: *spoll* – In this excerpt, the participant states that they do not recall reading the target word in the reading text.

   R What about this word “spoll”?

   P10 I don't know this word either.

   R So, you don't recall reading this word in the book?

   P10 I don't recall it, no.

2. [no awareness]: target word: *croom* – In this excerpt, the participant states that they are not sure if they recall reading the target word in the reading text.

   R What about “croom”? Do you recall reading this in the book?

   P39 Maybe, I'm not sure.

3. [noetic awareness]: target word: *mittle* – In this excerpt, the participant states that they recall reading the target word in the reading text, but they do not provide any further information regarding what they experienced while reading this target word in the reading text.

   R What about this word “mittle”? Do you recall reading this in the book?

   P14 I think I've seen it. But I don't remember where in the text exactly, in which context it was.

   R Okay. So you remember seeing this word in the text?

   P14 Yeah, but I don't remember any context.

4. [noetic awareness]: target word: *maive* – In this excerpt, the participant states the target word looks familiar to them without providing any further information regarding their experiences while reading this target word in the reading text.

R What about this word “maive”? Do you recall reading this in the book?

P13 It looks familiar.

5. [autonoetic awareness]: target word: *mittle* – In this excerpt, the participant states that they recall reading the target word in the reading text. They also provide further information regarding what they experienced while reading this target word in the reading text by stating that they saw this word multiple times and tried to figure out what it means.

R What about this word “mittle”? Do you recall reading this in the book?

P31 Yeah, I definitely remember and I was trying to figure out the meaning of the word but like, I remember that word was, like, pretty much redundant. So it appeared several times in the book, but yeah, but I'm not sure about the meaning.

R Okay. So, you think you tried to figure out what this word means while reading?

P31 Yes, because it appeared several times in the context. So I think I was trying. Yeah. So I was wondering what the word is meaning.

6. [autonoetic awareness]: target word: *merky* – In this excerpt, the participant states that they recall reading the target word in the reading text. They also provide further information regarding what they experienced while reading this target word in the reading text by stating where this word was used in the book. They also state that they found this word a funny word.

R What about “merky”?

P3 Yeah, I think that's a word he used to describe the people (in the future).

R Okay, so you noticed this word when you were reading, right? Did you try to guess its meaning while reading?

P31 Yeah like not strong. And, and like skinny. Like, they looked not strong. Like, yeah, that's the best way I can describe it. It was a funny word.

7. [autonoetic awareness]: target word: *rolley* – In this excerpt, the participant states that they recall reading the target word in the reading text. They also provide further information regarding what they experienced while reading this target word in the reading text by stating where they remember seeing this word in the reading text.

R What about "rolley"?

P57 Rolley, yeah. It was in the readings from today. Today's chapters. Rolley… It was in the context when the time machine was lost, I guess. When he was describing the lawn, and the Time Machine was lost. And when he was trying to find it.

R So, did you try to figure out the meaning of this word while reading?

P57 Yeah. When he saw the time machine was… I think they also like use this word to describe a place where, where a time machine was lost in. So he was trying to put his time machine out from this place.

8. [autonoetic awareness] target word: *crimb* – In this excerpt, the participant states that they recall reading the target word in the reading text. They also provide further information regarding what they experienced while reading this target word in the reading text by stating where they remember seeing this word in the reading text. The participant even goes one step beyond by stating what they think the word meant in the context.

R What about this word “crimb”? Do you recall reading this in the book?

P30 Yeah, I saw this. Crimb is probably like understand, that's what I think it means because he (Time Traveller) said, “Oh, they could not crimb what I was saying”. And it's like, they couldn't understand what I was saying.

The initial interrater reliability between the two raters, measured using the *kappa2* function in the R package *irr* (Gamer et al., 2019), showed 89% reliability ($p = .000$). The cases of disagreement (32 data points out of 456 data points) were resolved through discussions. After this initial two-rater coding, the main researcher continued coding the remaining data following the same coding scheme.

Enjoyment and Interest Questionnaire

On Day 2, participants also completed one final questionnaire about their enjoyment and interest levels while reading. This questionnaire contained two questions: one question about their overall enjoyment levels for the experiment and one about their interest levels for the reading text. They were asked to indicate their enjoyment and interest levels on a 100-point scale (0 = not enjoyable at all or not interesting at all, 100 = very enjoyable or very interesting). See Appendix I for the enjoyment and interest questionnaire.

Procedure

The data for this study was collected over three sessions in an eye-tracking lab at a Midwestern US university. The first two sessions were conducted on two consecutive days, and the third session was completed a week after the first two sessions. Participants met individually with the researcher in the eye-tracking lab. See Figure 6 for a summary of the complete procedure of the study.

Figure 6

Procedure

<table>
<tr><th>Day 1</th><th>Day 2</th><th>Online</th><th>One week later</th></tr>
<tr><td>• Background questionnaire
• Consent form</td><td>• Text (remaining three chapters) + comprehension questions (eye-tracking)
• Sentence- reading test (eye-tracking)</td><td rowspan="5">• Vocabulary size test</td><td>• Sentence- reading test (eye-tracking)</td></tr>
<tr><td rowspan="2">Random assignment to groups: Reading only or reading while listening</td><td>5-minute break</td><td>5-minute break</td></tr>
<tr><td>• Enjoyment and interest questionnaire
• Form recognition test</td><td>• Form recognition test
• Meaning recall test
• Meaning recognition test</td></tr>
<tr><td rowspan="2">• Text (first four chapters) + comprehension questions (eye-tracking)</td><td>5-minute break</td></tr>
<tr><td>• Retrospective interview
• Meaning recall test
• Meaning recognition test</td></tr>
</table>

On Day 1, participants completed the consent form and the background questionnaire. After completing the background questionnaire, participants were randomly assigned to one of the two groups: the reading only group and the reading while listening group. Then the main learning experiment started. During the experiment, participants read the first four chapters of the graded reader *The Time Machine* while their eye movements were being recorded. Participants in the reading only group were instructed to read the chapters at a normal pace; participants in the reading while listening group were instructed to read and listen to the audio simultaneously. All participants were also informed about the comprehension questions at the end of each chapter, so they were encouraged to read for comprehension. Before the main reading text, participants were also given a short passage about the author of the book to familiarize them with reading in the eye-tracking lab and moving between screens in the experiment. To decrease participants' surprise levels when they encountered the target

pseudowords in the main reading text, the practice passage also contained three pseudowords. The first session of the study was complete once participants finished reading chapter 4 of the reader and answered the related comprehension questions. This first session lasted around 45 minutes. On both days, after reading each chapter participants were asked how they felt, and if they reported any tiredness or boredom, they were allowed to take a 5-minute break before moving to the next chapter.

On Day 2, participants read the remaining three chapters of the reader and answered the comprehension questions after each chapter. Once the reading part was over, participants were asked to complete the surprise vocabulary tests, the enjoyment and interest questionnaire, and the retrospective interview. To minimize response contamination, participants completed the sentence-reading test first, which was followed with a five-minute break. After the short break, participants completed the enjoyment questionnaire and the form recognition test. While completing the form recognition test, participants were explicitly asked to choose the words they remembered seeing in the reader and not the ones in the sentence-reading test. After the form recognition test, there was a second five-minute break. This break was followed by the retrospective interview and the meaning recall and meaning recognition tests, in that order. At the end of Day 2, participants were asked to complete the Vocabulary Size test online before the third and final session. For this test, I used the online data collection platform Gorilla (www.gorilla.sc) (Anwyl-Irvine et al., 2019). I added participants' university email addresses to Gorilla which sent a message to the participants, directing them to log into the system to complete the test. This test was completed by participants on their own, outside the lab.

One week after the second session, participants visited the eye-tracking lab one last time. During this final session, participants completed all of the vocabulary tests one more time as

delayed posttests in the same order: sentence-reading test, form recognition test, meaning recall test, and meaning recognition test.

Apparatus

During the reading sessions and the immediate and delayed sentence-reading posttests, participants' eye-movements were recorded using a desk-mounted Eyelink 1000 at 1000 Hz sampling rate (SR Research Ltd.; http://www.sr-research.com/). During the reading sessions, participants sat in front a computer screen (Dell, 1920 × 1080, with a 60 Hz refresh rate) with a 66 cm seating distance. A chin and forehead rest were used to reduce head movements while reading and to increase data recording quality. Although participants read binocularly, only their dominant eye was monitored (right eye for 61 participants and left eye for 2 participants). Eye dominance was determined using the Porta test described in Godfroid (2020). Each participant was asked to extend one arm with their thumb pointing up, and they were asked to align their thumb with an object on the wall of the lab with both of their eyes open. Then, they were instructed to close their right eye while their left eye was open, and they were asked if their thumb stayed aligned with the object on wall when their right eye was closed (testing the dominance of left eye). Then, the same procedure was repeated with the left eye (testing the dominance of right eye). The eye for which they reported their thumb staying aligned with the object when open was their dominant eye.

The entire text as well as the sentences in the sentence-reading test were presented in black (RGB: 0, 0, 0) Consolas font, size 18, double-spaced on a light gray background (RGB: 204, 204, 204). The first four chapters in Session 1 were divided into 59 screens (plus three practice text screens), and the remaining three chapters in Session 2 were divided into 38 screens. Each screen consisted of five to twelve lines of text. The text on each screen were left-

aligned with a one-inch margin inserted around the edges of the screen to minimize spatial recording errors (Godfroid, 2020). There were zero to four target pseudowords on each screen depending on the distribution of the target pseudowords in the reader. To control for measurement errors, the target pseudowords were not presented in the first or last line, as the first or last word in a line, nor as the first or last word in a sentence (Godfroid, 2020; Godfroid et al., 2018; Rayner & Pollatsek, 2006). Each comprehension question was also presented at a separate screen individually. There were 18 comprehension question screens in Session 1 (two practice questions and 16 main questions) and 12 question screens in Session 2.

The screen settings were the same for the sentence-reading test (black, 18-point Consolas font against a light gray background with one-inch margin). Differently from reading sessions, sentences in this test were displayed individually at the center of the screen. The order of the sentences was randomized. As in the reading sessions, all sentences were left-aligned. Sentences were not longer than one line, and the target pseudowords and control words (real words and nonwords) did not appear at the beginning or end of a sentence.

All sessions started with instruction screens, which were followed with a calibration screen. A nine-point calibration were performed at the beginning of each reading session and after each break. Each screen in the experiments started with a drift correction, i.e., a check for spatial recording error. For drift correction, a black dot appeared at the top-left side of the screen, and participants were asked to fixate on the dot. The text was presented after a successful drift correction. During the entire experiment, I monitored the data recording quality from the host computer and performed additional calibrations whenever I detected any drifts during drift correction or any drifts from the text line while participants were reading.

Participants in the reading only group moved from one screen to the next using the

SPACE key on the keyboard. For the participants reading while listening group one screen moved to the next one automatically once the audio ended. All participants could reread the text on the current screen if they wanted to, but they were not allowed to go back to previous screens while reading. Rereading was possible even for the participants in the reading while listening group because this group were not forced to follow the audio strictly when reading. In other words, rereading some sentences on the same screen was possible for this group as well since all sentences stayed on the screen until the end of the audio (see Malone (2018) for a more controlled study where sentences disappeared from the screen after some time). However, as reported during interviews, for this group, rereading was more challenging because whenever participants wanted to go back to a previous sentence, this meant not being able to read some words in the upcoming sentences to keep up with the audio pace.

For the sentence-reading test, participants moved from one sentence to the next by pressing the SAPCE key, and as in the reading sessions, they could not go back to the previous sentences. For all comprehension questions (the ones after each chapter and the ones after sentences in the sentence-reading test), participants made a true/false choice using the arrow keys on the keyboard (RIGHT-ARROW for true and LEFT-ARROW for false).

Data Analysis

Eye Tracking Data Preprocessing and Cleaning

I used the default cognitive configuration of Eyelink for the parsing of eye-movements in both the main reading text and the sentence-reading posttest (Godfroid & Hui, 2020). There were technical problems (i.e., problems with camera calibration) in the data from two participants (one from each group). In addition, two participants (one from each group) scored very poorly on the reading comprehension test (accuracy lower than 60%). The data from these four participants

were removed from the eye-tracking dataset. All eye-tracking data were cleaned using the default four-stage fixation cleaning procedure of Eyelink Data Viewer program. In the first two steps of this cleaning procedure, overly short fixations were merged with neighboring fixations (i.e., fixations immediately before or after) within a specified distance based on a threshold value (first stage = fixations <80 ms and within 0.5 degrees; second stage = fixations <40 ms and within 1.25 degrees). Then, during the third step, any instances of three consecutive fixations <140 ms in an interest area were merged into one fixation. Finally, any fixations lower or higher than specified thresholds were removed from the data. After checking the removal threshold values used in previous studies and examining the fixation durations in the current dataset, I deleted the fixations <80 ms and >800 ms from the dataset as these fixation values reflect minor location errors rather than cognitive processing and losses of concentration, respectively (Carrol & Conklin, 2020; Conklin & Pellicer-Sánchez, 2016; Godfroid, 2020). I also visually inspected the remaining eye-tracking data trial-by-trial for any track loss or drift instances (Godfroid, 2020). Trials with track loss were removed from the data, and any vertical drifts were corrected manually. The initial pool of eye-tracking data for the target pseudowords contained 10647 data points (169 interest areas for target pseudowords × 63 participants). After the data preprocessing and cleaning stages, 826 data points (676 data points for removed participants and 150 data points for track loss) were discarded, leaving 9821 data points for analysis.

Eye-Tracking Measures as Outcome Variable

Before conducting any inferential statistics, I visually inspected the eye-movement data of each group using (a) plots of the means and (b) plots that illustrate key patterns in the data by applying smoothers that reflect the effect of the order of occurrence (from the first encounter to the last encounter with the target pseudowords). I created separate plots for each eye-tracking

measure.

The plots for the eye-tracking measures indicated non-linear (S-shaped) changes in participants' eye movements on target pseudowords across different exposures. Therefore, for the inferential statistics, following Godfroid et al. (2018), I treated the repeated occurrences of target pseudowords as a time-course variable and conducted a growth curve analysis for each eye-tracking measure (Godfroid, 2020; Mirman, 2014). The main aim of these analyses was to explore how eye-tracking measures changed over time as participants were exposed to the same target pseudowords from the beginning till the end of the reading text. These models can account for the non-linear increases in vocabulary learning over time (Godfroid et al., 2018). In growth curve models, in addition to the linear term of a variable (in this case Exposure^1), higher-order polynomial terms, such as quadratic and cubic effects (e.g., Exposure^2 and Exposure^3) are also added to the models as predictors.

All the models were fit using the *lmer* function in the R package *lme4* (Bates et al., 2015). Restricted maximum likelihood (REML) was used as the estimation method for model fitting. In all models, eye-tracking measures, i.e., gaze duration, regression path duration, rereading time, and total reading time, were the outcome variables. Since the distribution of all these measures were positively skewed, I transformed all the data using log transformation with base e. The main independent variables in all models were group (i.e., reading only vs. reading while listening), linear exposure, quadratic exposure, cubic exposure, and the exposure terms' interaction with group to the models. I selected exposure terms a priori based on visual inspection. The graphs for all eye-tracking measures indicated an S-shaped pattern, so I decided to include quadratic and cubic terms to the models. All exposure terms were changed into orthogonal polynomials before adding to the models to decrease the collinearity between the

three exposure terms. I also added the learner-level (vocabulary size, enjoyment level, and interest level) and item-level (part of speech, concreteness, and word length in letters) factors to the models as fixed effects. All continuous covariates were changed into z scores. During the standardization process, the mean of each covariate was subtracted from all values and were divided by two standard deviations (Gelman & Hill, 2007). Moreover, all models contained random intercepts for participants and target items. I fit a minimally adequate statistical model to each eye-tracking measure using a stepwise variable selection. In other words, with a backward stepwise approach any non-significant fixed effects variables that did not improve the model fit were removed from the model one by one. Changes in model fit were assessed through log-likelihood ratio tests and comparison of the Akaike information criterion (AIC) values. The final model was the most parsimonious model (i.e. with the fewest covariates) with the lowest AIC value.

The best fitting models for all eye-tracking measures were subjected to model criticism (Baayen & Milin, 2010). The potentially harmful outliers (i.e., data points with absolute standardized residuals exceeding 2.5 SDs) were removed from the data set, and the models were refitted. I report the results of these models in this dissertation. Effect sizes were calculated following Westfall et al. (2014).

Awareness Levels as Outcome Variable

To have a general understanding of participants' awareness of target words, I first investigated percentages for each awareness level, i.e., no awareness, noetic awareness, and autonoetic awareness, for both groups. After checking these percentages, for inferential statistics, I conducted two separate mixed-effects logistic regressions (Jaeger, 2008) using the *glmer* function in the *lme4* package in R. First, I wanted to examine how input modality affected the awareness

versus unawareness levels of participants by comparing the reported unaware (i.e., no awareness) and aware (i.e., noetic + autonoetic awareness) cases for each group. Therefore, I coded no awareness data points as 0 and noetic and autonoetic data points as 1 and changed the data into binary data. For the second analysis, I was interested in investigating how input modality influenced the participants' reported noetic and autonoetic awareness levels considering that these two awareness categories are argued to differ from each other in terms of awareness level and quality (i.e., autonoetic awareness representing a higher level awareness than noetic awareness) (Godfroid & Schmidtke, 2013). Therefore, to directly compare the noetic and autonoetic awareness levels, I first removed all data points where participants reported no awareness from the data (603 data points out of 1416 points). Then I coded noetic awareness data points as 0 and autonoetic data points as 1.

In both models, group (reading only vs. reading while listening) was the main predictor. The standardized summed total reading times on target words, learner-level (vocabulary size, enjoyment, and interest) and item-level (part of speech, concreteness, and word length in letters) factors were also added to the models as covariates. Random intercepts for participants and items were also included in the models. As in the models for eye-tracking measures, a backward stepwise approach was used in model selection. Firstly, all possible variables were included in the models as fixed effects, and then, any non-significant variables that did not improve the model fit were removed from the model one by one. Changes in model fit were assessed through log-likelihood ratio tests and comparison of the Akaike information criterion (AIC) values. The best fitting models were the most parsimonious ones with the fewest covariates. The dependent variable in each model was logit (also known as log odds), so to increase interpretability of the coefficient estimates, I calculated the odds ratios (OR) by exponentiating the coefficients

estimates, which measured the accuracy levels relative to the reference categories. An OR of 1 represents no effect, whereas an OR < .33 or OR > 3 indicates a moderately strong effect (Ferguson, 2009). In addition to OR, I also reported standardized effect sizes (Cohen's *d*), which was calculated by dividing the log odds ratio by 1.81 (Chinn, 2000).

Vocabulary Test Results as Outcome Variables

Sentence-reading test. Two eye-tracking measures, gaze duration and total reading time, were the two outcome variables for the sentence-reading test. The main aim with this test was to investigate the differences and similarities between the processing of pseudowords, real words, and nonwords in neutral sentences by two reading groups. I compared the two groups' processing of different word types by fitting linear mixed effects models to two eye-tracking measures data using *lmer* function in the *lme4* package in R. List was not a significant factor in neither of the models, so I collapsed all data across lists (i.e., List 1, List 2, and List 3). The main predictor in models were group (i.e., reading only vs. reading while listening), type of the critical word (i.e., pseudoword, real word, or nonword), test time (i.e., immediate vs. delayed), and the interactions among these three predictors. Learner-related (vocabulary size, enjoyment, and interest) factors were added to the models as covariates. Both models also included random intercepts for participants and items. I used a stepwise backward approach to find the best fitting model for the data using likelihood ratio tests and AIC values. The final models were the most parsimonious ones. I subjected the best fitting models to model criticism using the same steps explained in the previous section.

Form recognition, meaning recall, and meaning recognition tests. The outcomes for each test consisted of each participants' answers (correct or incorrect) for each target pseudoword, so the outcome variables in all three tests were binary. Therefore, to analyze the

learning gains on form recognition, meaning recall, and meaning recognition tests, I conducted three separate mixed logistic regressions (Jaeger, 2008) using the *glmer* function in the *lme4* package in R. The main analysis for the form recognition test contained only the binary scores for the target items.

Group (reading only vs. reading while listening), total exposure (range: 2 – 16 times), summed total reading time (as a measure of attention) and their interactions were the main predictors in all models. To increase the comparability of the effects of total exposure and summed total reading time, these variables were changed into z-scores through centering and scaling. The standardized learner-level (vocabulary size, enjoyment, and interest) and item-level (part of speech, concreteness, and word length in letters) factors were also added to the models as covariates. Random intercepts for participants and items were included in the models as well. Model selection consisted of backward, stepwise selection of predictors. Model comparison was done using likelihood ratio tests and AIC values, and the best fitting models were the most parsimonious ones with the fewest covariates. As in mixed-effects logistic regression for the awareness data, I again calculated the odds ratios (OR) for each coefficient estimate in the models and reported standardized effect sizes (Cohen's *d*) for each estimate.

CHAPTER 4: RESULTS (EYE-TRACKING MEASURES)

In this chapter, first, I provide preliminary descriptive statistics where I report reading only and reading while listening groups' comprehension scores and their mean reading times (mean screen time and mean summed total reading time) on the reading experiment. Then, I move on to the results for the eye-tracking measures and examine the reading times of each group on the target pseudowords from the first encounter to the last (research question 1a). I also report to what extent learner- and item-related factors predict the reading times indicated by four eye-tracking measures (research question 1b).

Comprehension Test

Except for two participants, all participants showed adequate comprehension (mean above 60%) of the reading text. As stated in the methodology section, these two participants were removed from the study. The remaining participants (n = 59) answered 81% of the comprehension questions correctly (SD = .39). The comprehension levels of the two groups were similar. The reading only group answered 82% of the questions correctly (SD = .39, *95% CIs* [0.79, 0.84], min. = 0.61, max. = 1.00); the accuracy level was 80% for the reading while listening group (SD = .40, *95% CIs* [0.78, 0.83], min. = .61, max. = 1.00).

Time on the Reading Text

To examine the comparability of the two groups in terms of the amount of time they spent on the reading text, I calculated the average time each group spent on each screen in the reading experiment and the average time they spent reading each target word. As explained in the methodology section, for the reading while listening group, the time on each screen in the experiment was determined by the length of the audio, so the participants in this group had no control on screen time. For the reading while listening group, the average time on each screen

was 30286 ms, and this number was the same for all participants. Differently from the reading while listening group, the reading only group had control on the screen time; they moved from one screen to the next by pressing a key on the keyboard. For the reading only group, the average time on each screen was 31138 ms (*SD* = 10769.76 ms) which was similar to the average screen time for the reading while listening group. For the reading only group, however, the participant with the minimum average screen time spent 13264 ms on each screen whereas the participant with maximum average screen time spent 63552 ms on each screen, indicating a large variability in the mean amount of time spent on each screen by different participants.

The average summed total reading time per target pseudoword was 3016.08 ms (*SD* = 1025.78 ms, *95% CIs* [2630, 3410]) for the reading only group and 2906.00 ms (*SD* = 313.49 ms, *95% CIs* [2790, 3020]) for the reading while listening group. The average number of occurrences of all target words was in the reading text 7.04 (*SD* = 3.71). These numbers indicate that participants in the reading only group spent an average 428 ms (*min.* = 170, *max* = 824) on each occurrence of a target word, and the reading while listening group spent an average 413 ms (*min.* = 310 ms, *max* = 507 ms) on each occurrence of a target word. Considering that the mean fixation duration for L1 readers on familiar words is 225 – 250 ms for silent reading and 275 – 375 ms for oral reading (Rayner, 1998, 2009), these average scores of L2 readers in the reading only and reading while listening groups indicate that L2 readers spend longer time fixating on novel words in a text. Moreover, although the mean scores for the two groups were similar, the minimum and maximum number values indicated a higher variability among the participants in the reading only group than the participants in the reading while listening group.

Eye-Tracking Measures

Graphs and Descriptive Statistics

Before conducting any inferential statistics for the eye-tracking data, I first visually inspected the data using (a) plots of the means and (b) scatter plots with fitted loess lines (i.e., smoothers) and 95% CIs, shown as gray shaded areas around the lines, to have a general understanding of how input modality influences participants' eye movements on target pseudowords while reading and how their eye movements change from the first encounter to the last encounter. The data points from all target words were included in the plots (Figures 7, 8, 9, and 10). The exact means, standard deviations, and 95% CIs for each occurrence are given Tables 6, 7, 8, and 9. The observations where the target pseudowords were skipped were removed from the data before calculating the means and standard deviations. Additionally, as can be seen from the expanded gray areas indicating the 95% CIs in the plots with fitted loess lines, there was a larger variability for the mean scores after 12 occurrences because only two target pseudowords appeared more than 12 times on the reading text.

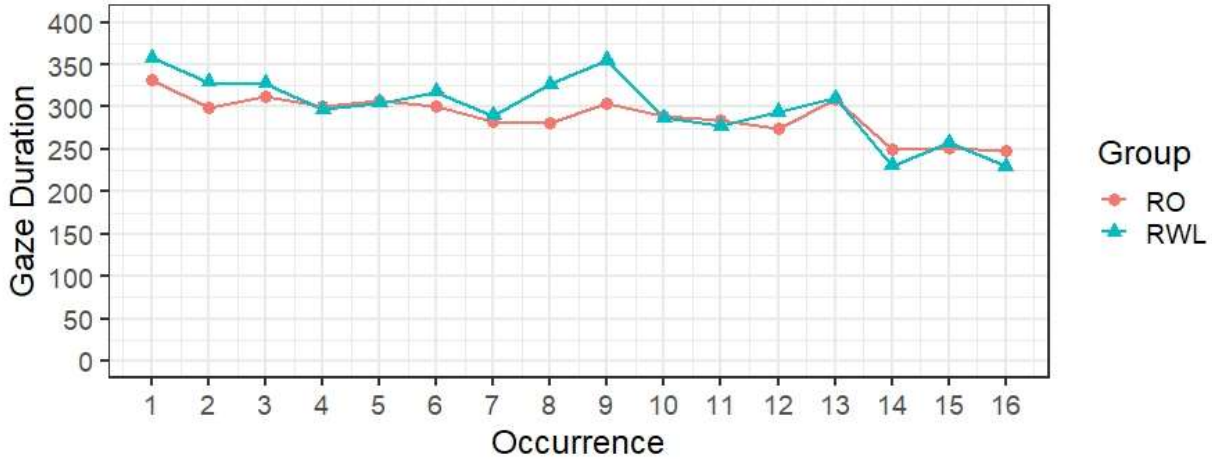
For gaze duration, indicating lexical access, there was an initial 25 ms difference between the two groups (see Table 6). The participants in the reading while listening group had higher gaze durations on the target pseudowords on the initial occurrences than the reading only group. As indicated in the plot with fitted loess lines in Figure 7, there was an overall decrease in gaze durations on the first 4 occurrences for both groups, and from the 5^{th} to the 7^{th} occurrence, the gaze durations were stable for both groups. From the 8^{th} to the 10^{th} occurrence, the two groups differed in their gaze duration patterns, the reading while listening group showing an increase in gaze durations and the reading only group showing a stable pattern. After the 10^{th} observation on the target pseudowords, the gaze durations of both groups gradually decreased again, and the

decrease was sharper for the reading while listening group.

Figure 7

Two Groups' Gaze Durations on the Target Pseudowords in the Reading Text

a) Plot of means



b) Plot with fitted loess lines and 95% CIs

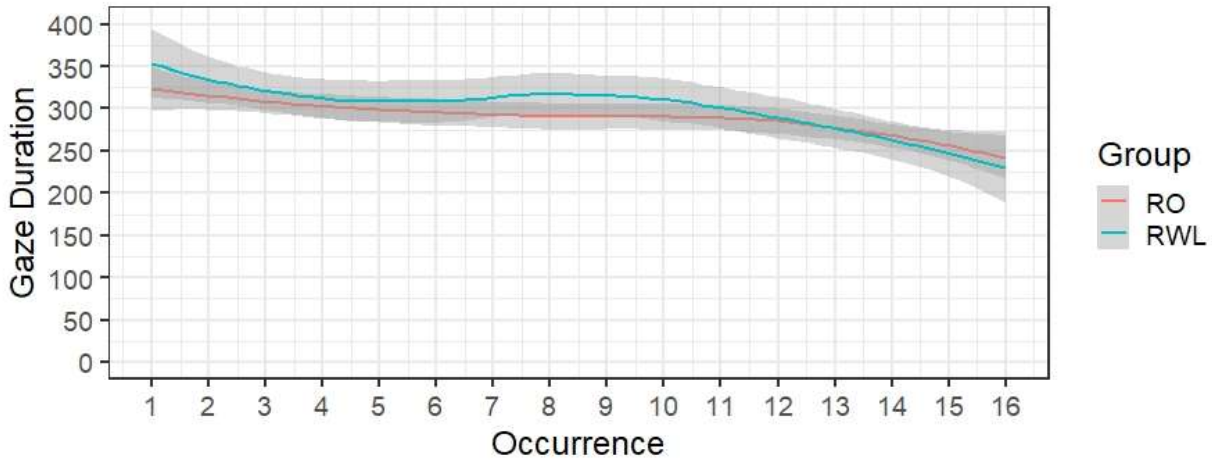


Table 6

Means, Standard Deviations, and 95%CIs of Two Groups' Gaze Durations

	Reading Only				Reading While Listening			
Encounter	*n*	*M*	*SD*	*95% CIs*	*n*	*M*	*SD*	*95% CIs*
1	657	332.68	208.19	[317, 349]	659	358.12	194.53	[343, 373]
2	656	299.69	164.92	[287, 312]	670	329.10	177.48	[316, 343]
3	583	312.32	168.17	[299, 326]	582	327.53	164.09	[314, 341]
4	490	300.00	149.82	[287, 313]	490	297.05	152.35	[284, 311]
5	458	307.08	175.88	[291, 323]	462	304.67	159.79	[290, 319]
6	379	301.49	157.51	[286, 317]	376	317.77	141.31	[303, 332]
7	340	282.66	163.07	[265, 300]	359	289.63	148.31	[274, 305]
8	326	281.83	155.09	[265, 299]	341	326.87	163.45	[309, 344]
9	237	303.37	135.84	[286, 321]	259	355.26	172.08	[334, 376]
10	133	289.68	131.47	[267, 312]	135	287.82	132.88	[265, 310]
11	105	284.82	177.91	[250, 319]	109	277.82	141.20	[251, 305]
12	82	275.04	195.99	[232, 318]	79	294.47	165.03	[258, 331]
13	57	309.33	133.65	[274, 345]	56	310.02	134.75	[274, 346]
14	28	249.18	78.57	[219, 280]	25	230.84	74.28	[200, 262]
15	26	251.81	95.80	[213, 291]	29	257.93	103.23	[219, 297]
16	27	247.44	138.61	[193, 302]	27	230.26	125.27	[181, 280]

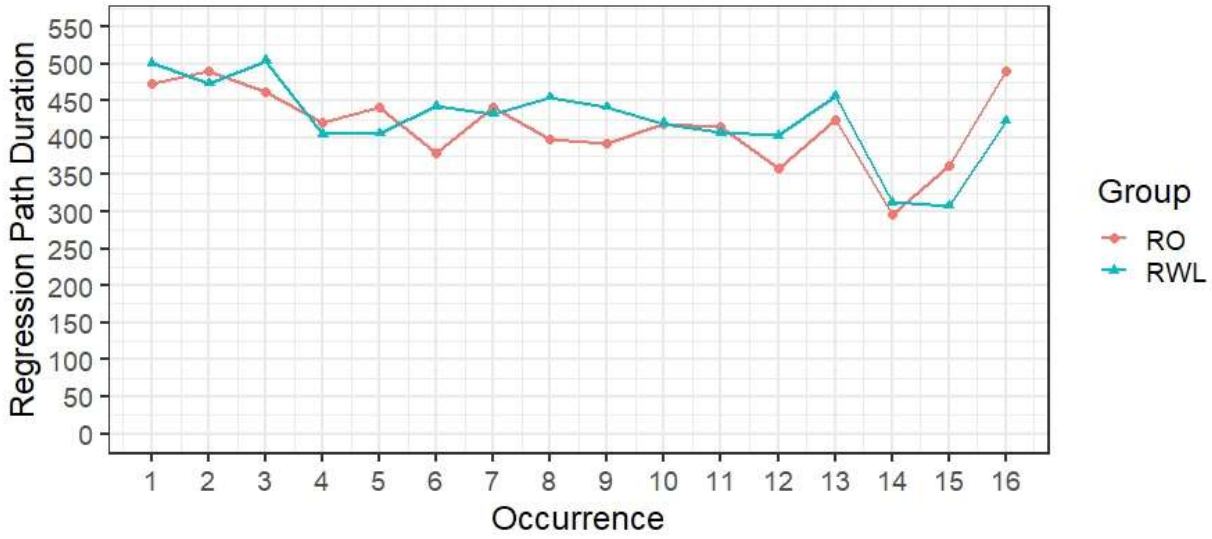
The reading patterns for the regression path duration were somewhat similar (see the plot with fitted loess lines in Figure 8). The reading while listening group had higher regression path duration than the reading only group on the first occurrence, with a 28 ms difference (see Table 7). Until the 5th occurrence, there was a decrease in regression path durations of both groups, which was followed with a more stable pattern until the 10th occurrence. After the 10th encounter with the target words, the regression path duration decreased for the reading while listening

group. The reading only group indicated a stable pattern until the 13th occurrence, which was followed with an increasing pattern. The difference between the two groups on the last occurrence was 68 ms.

Figure 8

Two Groups' Regression Path Durations on the Target Words in the Reading Text

a) Plot of means



b) Plot with fitted loess lines and 95% CIs

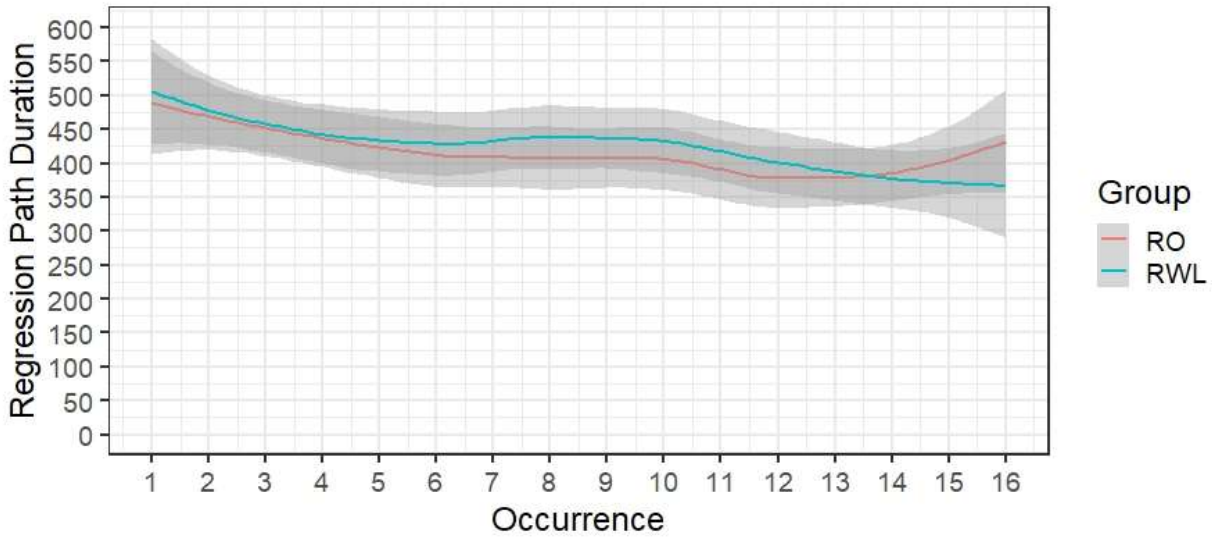


Table 7

Means, Standard Deviations, and 95% CIs of Two Groups' Regression Path Durations

	Reading Only				Reading While Listening			
Encounter	*n*	*M*	*SD*	*95% CIs*	*n*	*M*	*SD*	*95% CIs*
1	657	473.09	484.55	[436, 510]	659	501.26	369.60	[473, 530]
2	656	489.85	605.03	[443, 536]	670	473.45	338.65	[448, 499]
3	583	462.60	427.65	[428, 497]	582	504.02	882.83	[432, 576]
4	490	420.26	317.03	[392, 448]	490	404.62	308.47	[377, 432]
5	458	440.31	360.53	[407, 473]	462	405.92	320.25	[377, 435]
6	379	378.78	262.73	[352, 405]	376	442.66	551.48	[387, 499]
7	340	441.02	444.60	[394, 488]	359	432.49	328.62	[398, 467]
8	326	397.07	290.14	[365, 429]	341	454.64	387.09	[413, 496]
9	237	391.41	315.64	[351, 432]	259	441.47	317.76	[403, 480]
10	133	417.73	360.79	[356, 480]	135	419.32	398.44	[351, 487]
11	105	415.32	294.12	[358, 472]	109	407.30	279.32	[354, 460]
12	82	357.39	241.35	[304, 410]	79	402.70	224.04	[353, 453]
13	57	423.72	303.67	[343, 504]	56	456.36	360.61	[360, 553]
14	28	296.18	186.81	[224, 369]	25	312.36	182.46	[237, 388]
15	26	362.96	231.29	[270, 456]	29	307.72	172.71	[242, 373]
16	27	490.07	908.05	[131, 849]	27	422.44	357.58	[281, 564]

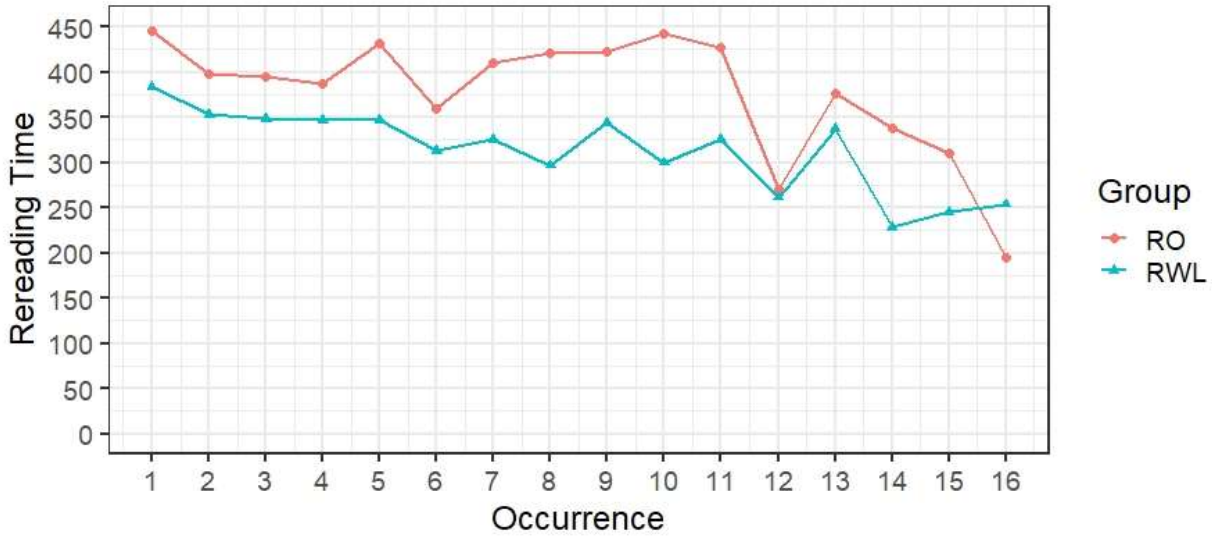
The picture that emerged for the rereading duration was different (Figure 9). Overall, the reading only group had higher rereading durations than the reading while listening group on all occurrences. The initial difference between the two groups was 62 ms (see Table 8). As can be seen in the plot with fitted loess lines in Figure 9, while the reading while listening group indicated a more linear decrease in the rereading times from the first encounter to the last, the rereading patterns for the reading only group varied widely. The slope of decrease was less steep

for the reading only group until around the 6th occurrence. After the 6th occurrence, there was an increase in the rereading times until around 10th occurrence, which was followed with a very steep decrease until the last encounter with the target pseudowords.

Figure 9

Two Groups' Rereading Durations on the Target Words in the Reading Text

a) Plot of means



b) Plot with fitted loess lines and 95% CIs

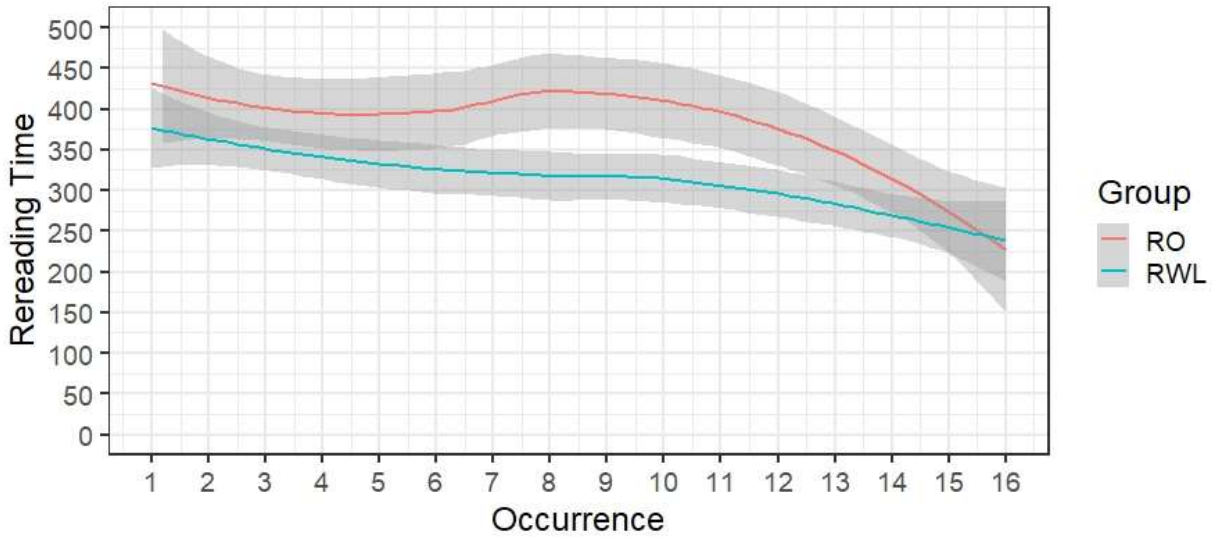


Table 8

Means, Standard Deviations, and 95% CIs of Two Groups' Rereading Times

	Reading Only				Reading While Listening			
Encounter	*n*	*M*	*SD*	*95% CIs*	*n*	*M*	*SD*	*95% CIs*
1	295	445.32	290.16	[412, 479]	240	383.26	244.06	[352, 414]
2	285	397.09	302.26	[362, 432]	261	352.72	219.09	[326, 379]
3	232	394.44	254.44	[362, 427]	225	348.60	203.80	[322, 375]
4	165	386.36	312.14	[338, 434]	152	346.78	234.77	[309, 384]
5	156	431.04	321.51	[380, 482]	125	346.74	204.89	[310, 383]
6	107	358.81	280.03	[305, 412]	114	313.11	161.39	[283, 343]
7	152	410.07	306.02	[361, 459]	123	324.88	213.33	[287, 363]
8	101	420.61	451.00	[332, 510]	122	296.61	152.43	[269, 324]
9	93	421.75	342.33	[351, 492]	100	344.04	214.37	[302, 387]
10	49	441.84	397.32	[328, 556]	48	299.56	225.68	[234, 365]
11	38	427.08	359.50	[309, 545]	31	324.87	145.16	[272, 378]
12	29	269.45	175.72	[203, 336]	31	261.16	138.26	[210, 312]
13	28	376.96	204.29	[298, 456]	24	336.92	191.81	[256, 418]
14	5	337.20	175.20	[120, 555]	4	228.25	104.94	[61, 395]
15	11	310.64	124.27	[227, 394]	6	245.00	92.54	[148, 342]
16	9	194.89	69.93	[141, 249]	9	253.33	165.88	[126, 381]

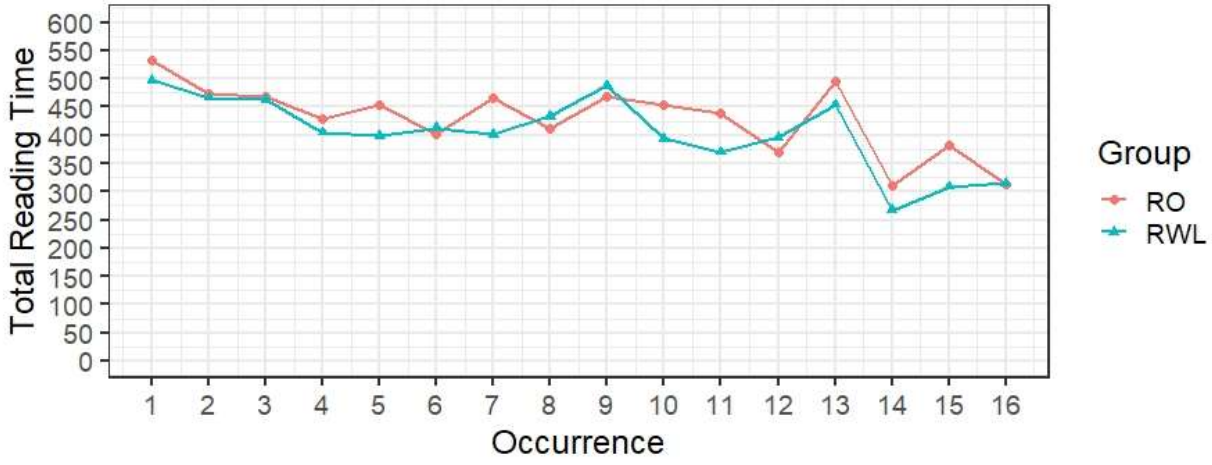
For both groups, total reading times indicated an S-shaped reading pattern (see the plot with fitted loess lines in Figure 10). First, from the first encounter to the 5th encounter, there was an initial negative slope, showing an initial decrease in total reading times. Second, from the 5th encounter to the pattern was more stable with a slight increase around the 9th exposure. Third, there was a steep gradual decrease in the total reading times from tenth exposure to the last. Overall, total reading times were a bit higher for the reading only group with an initial 35 ms

(see Table 9). This small difference between the two groups persisted until the final exposure.

Figure 10

Two Groups' Total Reading Times on the Target Words in the Reading Text

a) Plot of means



b) Plot with fitted loess lines and 95% CIs

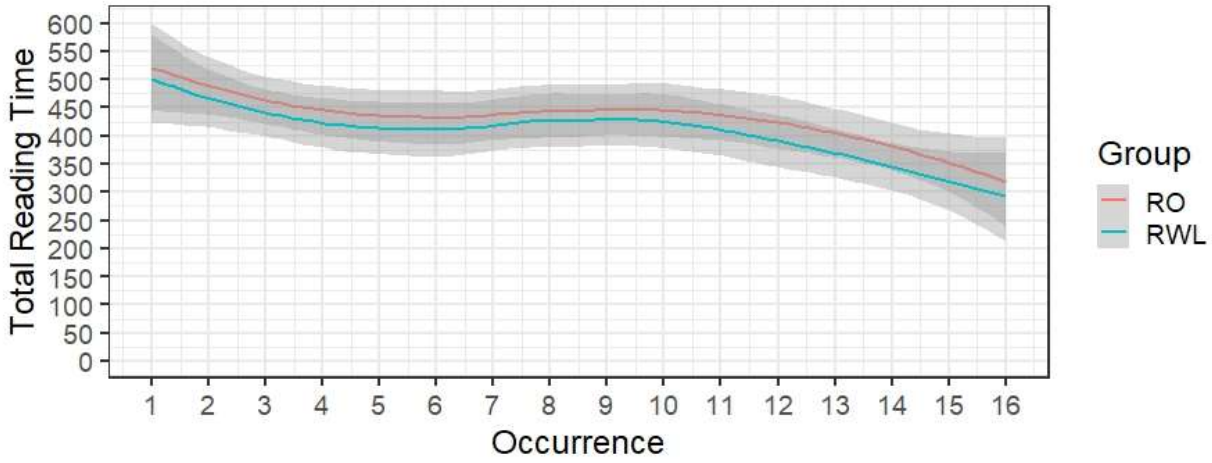


Table 9

Means, Standard Deviations, and 95% CIs of Two Groups' Total Reading Times

	Reading Only				Reading While Listening			
Encounter	*n*	*M*	*SD*	*95% CIs*	*n*	*M*	*SD*	*95% CIs*
1	657	532.63	347.08	[506, 559]	659	497.70	259.45	[478, 518]
2	656	472.20	295.43	[450, 495]	670	466.51	241.92	[448, 485]
3	583	469.29	287.38	[446, 493]	582	462.30	231.11	[443, 481]
4	490	430.10	299.59	[404, 457]	490	404.62	233.54	[384, 425]
5	458	453.89	326.39	[424, 484]	462	398.49	212.90	[379, 418]
6	379	402.79	260.26	[377, 429]	376	412.70	199.33	[392, 433]
7	340	465.99	316.11	[432, 500]	359	400.94	210.46	[379, 423]
8	326	412.15	354.36	[374, 451]	341	432.99	208.81	[411, 455]
9	237	468.87	345.85	[425, 513]	259	488.09	272.47	[455, 521]
10	133	452.46	345.38	[393, 512]	135	394.33	198.39	[361, 428]
11	105	439.38	320.09	[377, 501]	109	370.21	190.33	[334, 406]
12	82	370.33	259.09	[313, 427]	79	396.95	190.14	[354, 440]
13	57	494.51	249.88	[428, 561]	56	454.41	242.27	[390, 519]
14	28	309.39	186.88	[237, 382]	25	267.36	92.314	[229, 305]
15	26	383.23	214.41	[297, 470]	29	308.62	146.68	[253, 364]
16	27	312.41	158.96	[250, 375]	27	314.70	188.47	[240, 384]

Inferential Statistics

After examining the descriptive statistics and the scatter plots, I investigated the statistical significance of any differences in reading patterns of both groups using linear-mixed effects models. I fitted a separate model for each eye-tracking measure, and to find the best fitting model for each measure, I compared the more complex and simpler models using the likelihood ratio tests for nested models and the AIC values. The best fitting model for each eye-

tracking measure along with the simplest models and the maximal models are given in Appendix J. I used a backward stepwise approach in model selection (predictors with non-significant effects were removed from the maximal model one by one until reaching the best fitting model), however, for the sake of space, I only provide information about the values for the simplest, maximal, and best fitting models in Appendix J.

The best fitting model for gaze duration is given in Table 10. For gaze duration, there was a significant interaction between group and cubic effect of exposure, indicating that the two groups showed different reading patterns across different exposures. The interaction between group and linear effect of exposure and group and quadratic effect of exposure were also marginally significant. To understand the nature of these interactions, I ran two follow-up growth curve analyses, where I examined the effects of exposure for each group separately. Results showed that for the reading while listening group the decrease from the first exposure to the last was statistically significant: Exposure Linear (b = -0.23, SE = 0.07, t = -3.32, p = .001, d = 0.46), Exposure Quadratic (b = -0.15, SE = 0.06, t = -2.70, p = .007, d = 0.30), and Exposure Cubic (b = -0.19, SE = 0.04, t = -4.865, p < .001, d = 0.38). For the reading only group, however, none of the exposure terms indicated a statistically significant decrease: Exposure Linear (b = -0.08, SE = 0.07, t = -1.23, p = .22, d = 0.17), Exposure Quadratic (b = -0.03, SE = 0.05, t = -0.60, p = .55, d = 0.06), and Exposure Cubic (b = -0.06, SE = 0.04, t = -1.61, p = .11, d = 0.13). Thus, the gradual decrease in gaze durations from the first exposure to the last was significant for the reading while listening group, but not for the reading only group.

Table 10

The Best Fitting Model for Gaze Duration

Fixed Effects	*b*	*SE*	*95% CI*	*t*	*p*	*Exp (b)*
Intercept	5.59	0.03	5.53 – 5.66	174.75	<0.001	267.74
Group (RWL)	0.03	0.04	-0.05 – 0.11	0.70	0.48	1.03
Exposure (Linear)	-0.05	0.06	-0.18 – 0.07	-0.84	0.40	0.95
Exposure (Quadratic)	-0.03	0.05	-0.13 – 0.07	-0.57	0.57	0.97
Exposure (Cubic)	-0.05	0.04	-0.12 – 0.02	-1.37	0.17	0.95
Group (RWL) * Exposure (Linear)	-0.16	0.08	-0.32 – 0.01	-1.89	0.06	0.85
Group (RWL) * Exposure (Quadratic)	-0.13	0.07	-0.27 – 0.01	-1.82	0.07	0.88
Group (RWL) * Exposure (Cubic)	-0.15	0.05	-0.25 – -0.05	-2.85	0.01	0.86
Covariates						
Vocabulary Size (z score)	-0.11	0.04	-0.18 – -0.04	-2.93	0.01	0.90
Concreteness (z score)	0.07	0.03	0.01 – 0.13	2.16	0.03	1.07
Random Effects	*Variance*	*SD*				
1|Subject	0.03	0.13				
1|Item	0.01	0.07				
Residual	0.21	0.45				
$N_{subject}$	57					
N_{Item}	24					
Observations	8828					
Marginal R^2 / Conditional R^2	0.03 / 0.12					

Notes. Intercept level: Reading only. RWL: Reading while listening

Vocabulary size of the participants and concreteness levels of the target pseudowords were significant predictors of gaze durations in the model. Participants with higher vocabulary

size had lower gaze durations than participants with lower vocabulary size (d = 0.23). Moreover, more concrete target pseudowords received higher gaze duration than less concrete target pseudowords (d = 0.15). Participants' enjoyment and interest levels, and target words' part of speech and length were not significant predictors of gaze duration.

The best fitting model for the regression path duration is presented in Table 11. This model revealed a significant main effect of linear exposure and an interaction between group and cubic exposure. I followed up on the interaction by running growth curve models for each group separately. The analysis for the reading while listening group revealed a statistically significant linear (b = -0.23, SE = 0.08, t = -3.00, p = 0.003, d = 0.40) and cubic effect of exposure (b = -0.17, SE = 0.05, t = -3.69, p <0.001, d = 0.29), indicating a decrease in regression path duration from the first exposure to the last one. For the reading only group the linear effect of exposure was only marginally significant (b = -0.14, SE = 0.08, t = -1.78, p = 0.07, d = 0.23). The quadratic (b = 0.08, SE = 0.07, t = 1.23, p = .22, d = 0.13) and cubic effects of exposure (b = -0.01, SE = 0.05, t = -0.15, p = 0.88, d = 0.02) were not statistically significant.

Part of speech, concreteness, and word length were significant predictors of regression path duration. The pairwise comparisons for the effects of part of speech with Tukey adjustment p < .05 indicated that regression path durations were significantly shorter for nouns than adjectives (b = 0.08, SE = 0.03, z = 2.35, p = 0.05, d = 0.14) and verbs (b = 0.07, SE = 0.03, z = 2.23, p = 0.07, d = 0.13 only marginal significance). There was no statistically significant difference between regression path durations for adjectives and verbs (b = 0.004, SE = 0.03, z = 0.14, p = .99, d = 0.01).

Table 11

The Best Fitting Model for Regression Path Duration

Fixed Effects	*b*	*SE*	*95% CI*	*t*	*p*	*Exp (b)*
Intercept	5.76	0.04	5.68 – 5.84	144.17	<0.001	317.35
Group (RWL)	0.07	0.05	-0.03 – 0.15	1.37	0.17	1.07
Exposure (Linear)	-0.16	0.07	-0.30 – -0.02	-2.29	0.02	0.85
Exposure (Quadratic)	0.05	0.06	-0.06 – 0.17	0.92	0.36	1.05
Exposure (Cubic)	-0.02	0.04	-0.10 – 0.06	-0.53	0.59	0.98
Group (RWL) * Exposure (Linear)	-0.04	0.09	-0.22 – 0.15	-0.39	0.70	0.97
Group (RWL) * Exposure (Quadratic)	-0.05	0.08	-0.21 – 0.11	-0.63	0.53	0.95
Group (RWL) * Exposure (Cubic)	-0.15	0.06	-0.26 – -0.04	-2.57	0.01	0.86
Covariates						
Part of Speech (Verb)	0.07	0.03	0.01 – 0.14	2.23	0.03	1.08
Part of Speech (Adjective)	0.08	0.03	0.01 – 0.15	2.35	0.02	1.08
Concreteness (z score)	0.05	0.03	-0.01 – 0.11	1.78	0.07	1.05
Word Length (z score)	0.06	0.03	0.01 – 0.11	2.23	0.03	1.06
Random Effects	*Variance*	*SD*				
1|Subject	0.03	0.16				
1|Item	0.01	0.06				
Residual	0.27	0.52				
$N_{Subject}$	59					
N_{Item}	24					
Observations	9138					
Marginal R^2 / Conditional R^2	0.02 / 0.12					

Notes. Intercept level: Reading only and noun. RWL: Reading while listening

Moreover, six-letter pseudowords had significantly higher regression path durations than five-letter pseudowords, indicating that even one additional letter influences participants' reading times ($d = 0.11$). Finally, concreteness had a marginally significant effect on regression path durations, with more concrete words having longer regression path durations than less concrete words ($d = 0.09$). Participants' vocabulary size and their enjoyment and interest levels were not statistically significant predictors of regression path durations.

The best fitting model for rereading time (Table 12) indicated different results than the models for gaze duration and regression path duration. The best fitting model for rereading durations did not include an interaction between group and exposure, but there was a statistically significant main effect of exposure on the rereading durations for both linear effect of exposure and cubic effect of exposure. Unlike the larger differences observed in the raw mean scores of the two groups, there was not a statistically significant difference between the two group's rereading times.

Word length was the only statistically significant control variable for the rereading durations. Six-letter words were reread significantly longer than five-letter words ($d = 0.20$). Participants' vocabulary, their enjoyment and interest levels and target words' part of speech and concreteness levels were not statistically significant predictors of rereading durations.

Table 12

The Best Fitting Model for Rereading Time

Fixed Effects	*b*	*SE*	*95% CI*	*t*	*p*	*Exp(b)*
Intercept	5.65	0.04	5.58 – 5.73	148.60	<0.001	284.29
Group (RWL)	-0.06	0.04	-0.14 – 0.02	-1.42	0.16	0.94
Exposure (Linear)	-0.40	0.09	-0.59 – -0.22	-4.33	<0.001	0.67
Exposure (Quadratic)	-0.14	0.08	-0.29 – 0.01	-1.79	0.07	0.87
Exposure (Cubic)	-0.16	0.05	-0.26 – -0.06	-3.06	0.002	0.85
Covariates						
Word Length (z score)	0.11	0.04	0.04 – 0.18	3.08	0.002	1.11
Random Effects	*Variance*	*SD*				
1|Subject	0.02	0.14				
1|Item	0.01	0.07				
Residual	0.29	0.54				
$N_{Subject}$	59					
N_{Item}	24					
Observations	3338					
Marginal R^2 / Conditional R^2	0.03 / 0.10					

Notes. Intercept level: Reading only. RWL: Reading while listening

The best fitting model for total reading time included the main effects of group and exposure but not interactions between these predictors (Table 13). The model revealed a main effect of exposure on total reading times for linear effect of exposure and cubic effect of exposure. In other words, total reading times decreased as participants' exposure to the target pseudowords increased. Group was not a statistically significant predictor of total reading times, indicating that the total reading times were not affected by the presence or absence of audio while reading.

Table 13

The Best Fitting Model for Total Reading Time

Fixed Effects	*b*	*SE*	*95% CI*	*t*	*p*	*Exp (b)*
Intercept	5.84	0.04	5.75 – 5.93	131.04	<0.001	343.78
Group (RWL)	0.02	0.05	-0.08 – 0.12	0.44	0.66	1.02
Exposure (Linear)	-0.30	0.05	-0.40 – -0.20	-5.98	<0.001	0.74
Exposure (Quadratic)	-0.02	0.04	-0.10 – 0.06	-0.50	0.62	0.98
Exposure (Cubic)	-0.19	0.03	-0.25 – -0.13	-6.71	<0.001	0.83
Covariates						
Part of Speech (Verb)	0.09	0.04	0.02 – 0.16	2.43	0.02	1.09
Part of Speech (Adjective)	0.05	0.04	-0.02 – 0.12	1.36	0.18	1.05
Concreteness (z score)	0.07	0.03	0.01 – 0.13	2.35	0.02	1.07
Word Length (z score)	0.06	0.03	0.00 – 0.11	1.99	0.05	1.06
Random Effects	*Variance*	*SD*				
1|Subject	0.04	0.19				
1|Item	0.01	0.06				
Residual	0.24	0.49				
$N_{subject}$	59					
N_{Item}	24					
Observations	9131					
Marginal R^2 / Conditional R^2	0.03 / 0.17					

Notes. Intercept level: Reading only and noun. RWL: Reading while listening

Part of speech, concreteness, and word length had a statistically significant influence on total reading times. Pairwise comparisons with Tukey adjustments $p < .05$ showed that verbs received significantly higher total reading times than nouns ($b = 0.09$, $SE = 0.04$, $z = 2.42$, $p = .04$, $d = 0.17$). There were no statistically significant differences for total reading times

between adjectives and nouns (b = 0.05, SE = 0.04, z = 1.36, p = 0.36, d = 0.09) and adjectives and verbs (b = -0.04, SE = 0.037, z = -1.03, p = 0.56, d = 0.08). Total reading times were longer for more concrete words than less concrete words (d = 0.13). Finally, six-letter words received significantly higher total reading times than five-letter words (d = 0.11). Participants' vocabulary size as well as their enjoyment and interest levels did not affect their total reading times on target pseudowords significantly.

Summary of Results for Eye-Tracking Measures

In Table 14, I summarize the results for the eye-tracking measures, including the research questions and overall findings for each eye-tracking measure.

Table 14

Summary of Results for Eye-Tracking Measures

Research Question	Eye-Tracking Measure	Findings
(1a) How does input modality affect L2 learners' attention to target words while reading?	Gaze duration	Overall longer gaze duration for the reading while listening group (particularly on initial encounters) An S-shaped decrease with a sharper decreasing trend for the reading while listening group • RWL: A statistically significant decrease in gaze durations from the first exposure to the last • RO: No statistically significant decrease
	Regression path duration	Overall longer regression path duration for the reading while listening group (particularly on initial encounters) An S-shaped decrease with sharper decreasing trend for the reading while listening group • RWL: A statistically significant decrease in regression path durations from the first exposure to the last • RO: No statistically significant decrease
	Rereading time	An S-shaped, statistically significant decrease in rereading durations, no interaction between group and exposure

Table 14 (cont'd)

Research Question	Eye-Tracking Measure	Findings
(1a) How does input modality affect L2 learners' attention to target words while reading?	Total reading time	An S-shaped, statistically significant decrease in total reading time, no interaction between group and exposure
(1b) Do learner-related factors and item-related factors predict L2 learners' attention to target words while reading?	Gaze duration	Significant effects of vocabulary size • participants with lower vocabulary size > participants with higher vocabulary size Significant effects of concreteness • more concrete > less concrete
	Regression path duration	Significant effects of part of speech • nouns < verbs • nouns < adjectives • verbs = adjectives Significant effects of concreteness • more concrete > less concrete Significant effects of word length • longer words > shorter words
	Rereading time	Significant effects of word length • longer words > shorter words
	Total reading time	Significant effects of part of speech • nouns < verbs • nouns = adjectives • verbs = adjectives Significant effects of concreteness • more concrete > less concrete Significant effects of word length • longer words > shorter words

Notes. RO: Reading only, RWL: Reading while listening.

CHAPTER 5: RESULTS (RETROSPECTIVE VERBAL REPORTS)

In this chapter, I report the findings about participants' reported awareness levels for target words (research question 2a). I also provide information about the extent these awareness levels are influenced by learner-related and item-related factors (research question 2b).

Distribution of Awareness Levels

Table 15 shows the percentages of reported awareness levels of the participants. Both groups reported some level of awareness (noetic and autonoetic combined) for around 60% percent of the of the target words, whereas they reported no awareness for the remaining 40%. While the reading only group had slightly higher amount of noetic awareness than the reading while listening group, the reading while listening group had slightly higher amount of autonoetic awareness than the reading only group. Overall, these percentages indicated similar distributions of the three awareness levels between the two groups.

Table 15

Total Counts and Percentages for Awareness Levels of Two Groups

	Reading Only		Reading While Listening	
	Count	Percentage	Count	Percentage
No Awareness	280	%40.23	290	%40.28
Noetic Awareness	224	%32.18	220	%30.56
Autonoetic Awareness	192	%27.59	210	% 29.16
Total	696	%100	720	%100

Inferential Statistics for Awareness Data

To investigate the statistical significance of the slight differences between the two groups' awareness levels, I conducted two mixed-effects logistics regressions. In the first regression model, I compared the two groups' reports of aware and unaware cases; in the second model, I compared the groups' reports of noetic and autonoetic awareness cases. See Appendix K for the comparison between the models with only the main predictors, the maximal models with all possible predictors, and the best fitting models.

The results for the first mixed logistic regression model is given in Table 16. The best fitting model for comparing the aware and unaware cases included the main effects of group and main effects of total exposure, word length, and summed total reading time. The model did not demonstrate any statistically significant effect of group on the awareness and unawareness levels of the participants. In other words, participants in both groups were comparable in terms of their reported awareness and unawareness levels for the target pseudowords in the reading text.

The amount of exposure to the target pseudowords in the reading text was a statistically significant predictor of awareness levels. As the exposure frequency to target words in the reading text increased, the chances of participants developing awareness of these words also increased. Moreover, participants were more likely to develop awareness of the longer target words compared to the shorter ones. Finally, summed total reading time was a statistically significant predictor of awareness levels. When participants spent time looking a target word, the chances of them reporting this word as being recalled during the retrospective interview was higher. Participants' vocabulary size, their enjoyment and interest levels, and target words' concreteness levels and parts of speech were not statistically significant predictors of awareness levels.

Table 16

The Best Fitting Model Comparing Awareness and Unawareness Levels

Fixed Effects	*b*	*Exp(b)*	*SE*	*95% CI*	*z*	*p*	*d*
Intercept	0.49	1.62	0.34	1.08 – 2.44	2.34	0.02	0.27
Group (RWL)	0.04	1.04	0.26	0.64 – 1.71	0.17	0.86	0.04
Covariates	*b*	*Exp(b)*	*SE*	*95% CI*	*z*	*p*	*d*
Total Exposure (z score)	1.02	2.77	0.74	1.65 – 4.66	3.84	<0.001	0.56
Word Length (z score)	0.66	1.93	0.48	1.18 – 3.14	2.63	0.01	0.36
Summed Total Reading Time (z score)	0.53	1.70	0.37	1.10 – 2.61	2.40	0.02	0.29

Random Effects	*Variance*	*SD*
1|Subject	0.72	0.85
1|Item	0.25	0.50
$N_{Subject}$	59	
N_{Item}	24	
Observations	1416	
Marginal R^2 / Conditional R^2	0.10 / 0.30	

Notes. Intercept level: Reading only. RWL: Reading while listening

The results of the second regression model where I compared the noetic and autonoetic awareness levels of the two groups are given in Table 17. The best fitting model for this comparison included the main effects of group along with the main effects of vocabulary size, total exposure, and summed total reading time, and an interaction between group and summed total reading time. This model did not show a statistically significant effect of group on the participants' reports of noetic and autonoetic awareness of target words in the reading text. In other words, the amount of noetic and autonoetic awareness cases was similar for the reading

only and reading while listening groups, indicating that both groups had similar quality of awareness for target words.

Table 17

The Best Fitting Model Comparing Noetic and Autonoetic Awareness Levels

Fixed Effects	*b*	*Exp(b)*	*SE*	*95% CI*	*z*	*p*	*d*
Intercept	-0.32	0.72	0.14	0.50 – 1.05	-1.68	0.09	0.18
Group (RWL)	1.18	1.19	0.23	0.82 – 1.74	0.92	0.36	0.65
Covariates	*b*	*Exp(b)*	*SE*	*95% CI*	*z*	*p*	*d*
Vocabulary Size (z score)	0.42	1.52	0.29	1.05 – 2.20	2.24	0.03	0.23
Total Exposure (z score)	0.81	2.24	0.71	1.20 – 4.19	2.53	0.01	0.45
Summed Total Reading Time (z score)	-0.07	0.93	0.20	0.61 – 1.43	-0.32	0.75	0.04
Group (RWL):Summed Total Reading Time (z score)	0.87	2.39	0.83	1.21 – 4.71	2.52	0.01	0.48
Random Effects	*Variance*	*SD*					
1|Subject	0.15	0.39					
1|Item	0.41	0.64					
N Subject	57						
N Item	24						
Observations	813						
Marginal R2 / Conditional R2	0.085 / 0.219						

Notes. Intercept level: Reading only. RWL: Reading while listening

Participants' vocabulary size was a positive, statistically significant predictor in the model. This means that as participants' vocabulary size increased, the chances of them reporting

autonoetic awareness for the target words increased. In other words, participants with higher vocabulary size reported more experience-based awareness for the target words than participants with lower vocabulary size. Moreover, the odds of reporting experienced-based awareness were higher for the words that appeared more times in the reading text than the ones appeared fewer times. Finally, there was a statistically significant interaction between group and summed total reading times. This interaction showed that when the reading while listening group spent more time looking at a word while reading, their chances of reporting autonoetic awareness of that word was higher. Participants' enjoyment and interest levels and target words' concreteness levels and parts of speech were not statistically significant predictors of awareness levels.

Summary of Results for Retrospective Verbal Reports

A summary of the results for the awareness data is given in Table 18.

Table 18

Summary of Results for Retrospective Verbal Reports

Research Question	Findings
(2a) How does input modality affect L2 learners' awareness of target words while reading?	Awareness vs. No Awareness • No statistically significant difference between RWL and RO groups. Noetic Awareness vs. Autonoetic Awareness • No statistically significant difference between RWL and RO groups.
(2b) Do summed total reading times, learner-related factors (vocabulary size, interest level, and enjoyment level), and item-related factors (frequency of occurrence, part of speech, and concreteness) predict L2 learners' awareness of target words while reading?	Awareness vs. No Awareness: Significant effects of total exposure • Increased chances of developing awareness for target words with higher frequency of occurrence Significant effects of word length • Increased chances of developing awareness for longer target words

Table 18 (cont'd)

Research Question	Findings
	Significant effects of summed total reading time • Increased chances of developing awareness for word which received higher summed total reading time while reading Noetic Awareness vs. Autonoetic Awareness: Significant effects of vocabulary size • Increased chances of developing autonoetic awareness for participants with higher vocabulary size Significant effects of total exposure • Increased chances of developing autonoetic awareness for target words with higher frequency of occurrence A significant interaction between group and summed total reading time • RWL: Increased chances of developing awareness for word which received higher summed total reading time while reading

Notes. RO: Reading only, RWL: Reading while listening.

CHAPTER 6: RESULTS (VOCABULARY TESTS)

In this chapter, I report the learning gains on the three traditional vocabulary tests (form recognition, meaning recall, and meaning recognition) (research question 3a) and the extent these gains are influenced by attention levels of participants as well as learner- and item-related predictors (research question 3b). Then, I provide the results of the sentence reading test by focusing on two eye-tracking measures (research 3c) and the effects of learner-related factors on these eye-tracking measures (research question 3d).

Vocabulary Tests

Descriptive Results

Tables 19, 20, and 21 indicate the performance of the two groups on the form recognition, meaning recall, and meaning recognition tests respectively. On the form recognition test, participants in the reading only group recognized the form of around 53% of the target words correctly, while the form recognition percentage was around 56% for the reading while listening group. The mean scores of each group on the delayed tests differed from each other. There was around a 5% decrease in the performance of the reading only group, with a 48% mean accuracy. In contrast, there was a 5% increase in the performance of the reading while listening group, reaching 61% mean accuracy on the delayed test. The two groups' performances were more similar for the meaning recall test. Their performance accuracy was around 5% for the meaning recall test on both testing times. The reading while listening group scored somewhat higher than the reading only group on the meaning recognition test on both testing times (24% and 25% for the RWL and 20% and 22% for the RO), but the differences between scores were very small.

Overall, for both groups the learning gains were higher for form recognition than

meaning recognition. Moreover, their meaning recognition scores were higher than their meaning recall scores. The learning gains for each target pseudoword is given in Appendix M.

Table 19

Performance on the Form Recognition Test (Max. = 24)

	Immediate		Delayed	
	M (SD)	*95% CIs*	*M (SD)*	*95% CIs*
Reading Only (n = 29)	12.79 (4.92)	[10.90, 14.70]	11.72 (5.01)	[9.82, 13.60]
Reading While Listening (n = 30)	13.57 (4.13)	[12.00, 15.10]	14.70 (5.36)	[12.70, 16.70]

Note: Mean scores only for the target items in the form recognition test.

Table 20

Performance on the Meaning Recall Test (Max. = 24)

	Immediate		Delayed	
	M (SD)	*95% CIs*	*M (SD)*	*95% CIs*
Reading Only (n = 29)	1.07 (1.67)	[0.44, 1.70]	1.41 (2.10)	[0.62, 2.21]
Reading While Listening (n = 30)	1.33 (1.63)	[0.73, 1.94]	1.7 (2.02)	[0.95, 2.45]

Table 21

Performance on the Meaning Recognition Test (Max. = 24)

	Immediate		Delayed	
	M (SD)	*95% CIs*	*M (SD)*	*95% CIs*
Reading Only (n = 29)	5.31 (3.25)	[4.07, 6.55]	4.90 (3.06)	[3.73, 6.06]
Reading While Listening (n = 30)	5.93 (3.32)	[4.69, 7.17]	6.00 (3.77)	[4.59, 7.41]

Form Recognition Test

As a preliminary analysis, I first ran a mixed logistic regression to examine if learners in both groups chose more target pseudowords than distractors on the form recognition test. The results of the model are presented in Table 22. The main effect of item type, and the interaction between item type and test time were significant, whereas the main effect of group and the interaction between group and item type were not statistically significant. Pairwise comparisons (results averaged over the levels of test time) showed that the estimated odds of choosing target pseudowords as recognized words in the test were much higher than the odds of choosing distractors in the reading only group (b = 2.46, SE = 0.20, p = <.0001) and reading while listening group (b = -2.13, SE = 0.19, p <.0001). As this preliminary analysis revealed significantly higher correct responses for the target items than distractor items, I continued my data analysis with the main analysis where I compared the effects of input condition on the learning of the form of the target pseudowords only.

Table 22

The Model Comparing Target Words and Distractors on the Form Recognition Test

Fixed Effects	*b*	*Exp(b)*	*SE*	*95% CI*	*z*	*p*	*d*
Intercept	-2.01	0.13	0.03	0.09 – 0.20	-9.52	<0.001	1.11
Group (RWL)	0.45	1.57	0.38	0.98 – 2.51	1.87	0.06	0.25
Item Type (Target)	2.15	8.56	1.85	5.61 – 13.06	9.95	<0.001	1.19
Test Time (delayed)	-0.85	0.43	0.08	0.30 – 0.61	-4.64	<0.001	0.47
Group (RWL):Item Type (Target)	-0.29	0.75	0.14	0.52 – 1.09	-1.51	0.13	0.16
Group (RWL):Test Time (delayed)	0.54	1.72	0.40	1.09 – 2.71	2.33	0.02	0.30

Table 22 (cont'd)

Fixed Effects	*b*	*Exp(b)*	*SE*	*95% CI*	*z*	*p*	*d*
Item Type (Target): Test Time (delayed)	0.64	1.89	0.41	1.23 – 2.89	2.92	0.01	0.35
Group (RWL):Item Type (Target):Test Time (delayed)	-0.10	0.91	0.26	0.52 – 1.58	-0.35	0.73	0.056

Random Effects	*Variance*	*SD*
1|Subject	0.53	0.73
1|Item	0.32	0.57
$N_{Subject}$	59	
N_{Item}	48	
Observations	5664	
Marginal R^2 / Conditional R^2	0.260 / 0.412	

Notes. Intercept: Reading only, distractor items, immediate test. RWL: Reading while listening.

I investigated the effects of reading condition on the form recognition using mixed-effects logistic regressions. First, I investigated the main effects of group and test time and the interaction between group and test time on the learning gains. The model indicated a statistically significant interaction between group and test time on the learning of form of the target pseudowords (b = 0.46, SE = 0.27, p = .006, d = 0.25). Pairwise comparisons showed that the two groups performed similarly on the immediate test (b = -0.16, SE = 0.25, p = .51), but their scores differed on the delayed test (b = -0.63, SE = 0.25, p = 0.01). On the delayed posttest, the odds of obtaining a correct answer by the reading while listening group were 1.88 times the odds of a correct answer by the reading only group. Moreover, there was no statistically significant difference between the immediate and delayed test scores of the reading only group (b = 0.22, SE = 0.12, p = 0.06) whereas the scores of the reading while listening group increased significantly

from the immediate test to the delayed test (b = -0.24, SE = 0.12, p = 0.04).

After investigating the effects of main predictors of group and time on the form recognition scores, I conducted a second mixed-effects logistic regression where I examined the effects of learner- and item-related factors on the form recognition levels along with the effects of group and time. The best fitting model included the main effects of group and test time, the interaction between group and test time, and vocabulary size, total count, word length, and summed total reading time as predictors. See Appendix L for the comparison between the model with only the main predictors, the maximal model with all possible predictors, and the best fitting model.

The results of the best fitting model, given in Table 23, indicated a significant interaction between group and test time. In this model, the interaction between group and time was still statistically significant, and the pairwise comparisons also indicated similar results. According to pairwise comparisons the two groups performed similarly on the immediate test (b = -0.31, SE = 0.25, p = .21), but their scores differed on the delayed test (b = -0.75, SE = 0.25, p = .003). Moreover, there was no statistically significant difference between the immediate and delayed test scores of the reading only group (b = 0.18, SE = 0.12, p = .13) whereas the scores of the reading while listening group increased significantly from the immediate test to the delayed test (b = -0.26, SE = 0.12, p = .03).

Table 23

The Best Fitting Model for the Form Recognition Test

Fixed Effects	*b*	*OR*	*SE*	*95% CIs for OR*	*z*	*p*	*d*
Intercept	0.04	1.04	0.19	0.71 – 1.53	0.22	0.83	0.02
Group (RWL)	0.32	1.37	0.25	0.84 – 2.25	1.27	0.21	0.18
Test time (delayed)	-0.18	0.83	0.12	0.66 – 1.06	-1.51	0.13	0.10
Group (RWL):Test time (delayed)	0.42	1.55	0.20	1.11 – 2.17	2.57	0.01	0.23
Covariates							
Vocabulary Size (z score)	0.52	1.68	0.24	1.06 – 2.68	2.16	0.03	0.29
Total Exposure (z score)	0.56	1.76	0.23	1.13 – 2.74	2.69	0.01	0.31
Word Length (z score)	0.51	1.66	0.18	1.16 – 2.38	2.74	0.01	0.28
Summed Total Reading Time (z score)	0.32	1.38	0.18	0.96 – 1.97	2.59	0.08	0.18
Group (RWL):Summed Total Reading Time (z score)	0.42	1.52	0.20	1.02 – 2.25	2.05	0.04	0.23

Random Effects	*Variance*	*SD*
1|Subject	0.66	0.82
1|Item	0.15	0.38
N_{sub}	57	
N_{item}	24	
Observations	2736	
Marginal R^2 / Conditional R^2	0.09 / 0.27	

Notes. Intercept: Reading only, immediate test. RWL: Reading while listening.

Vocabulary size of the participants, total number of exposures to the target pseudowords, the length of target pseudowords, and summed total reading time were significant predictors of

performance on the form recognition test. The words with higher frequency of occurrence in the reading text were learned better than the words with lower frequency of occurrence. Additionally, there was a statistically significant interaction between group and summed total reading time. When L2 learners in the reading while listening group spent more time looking at the target novel words in the reading text, their chances of learning the written form of these words were higher. As reflected in the *p* value for the main effect of summed total reading time, the same effect was not observed for the reading only group.

Moreover, participants with higher vocabulary size had higher learning gains of the forms of the target pseudowords than participants with lower vocabulary size. Finally, even though all target pseudowords were five- to six-letters long, word length was still a significant predictor of the performance on the form recognition test with six-letter long words being learned significantly better than five-letter long words. Participants' enjoyment and interest levels and the target pseudowords' part of speech and concreteness levels were not significant predictors of the performance on the form recognition test.

Meaning Recall Test

For the meaning recall test, I first ran a model with the main predictors, group and time. This model indicated a statistically significant effect of test time on the scores of both groups ($b = 0.35$, $SE = 0.17$, $p = .05$, $d = 0.19$). Both groups had higher scores on the delayed test than the immediate test which indicates the possible influence of the immediate meaning recognition test on the scores of the delayed meaning recall test. The main effect of group was not statistically significant ($b = 0.33$, $SE = 0.40$, $p = .42$, $d = 0.18$), which meant that the performance of the two groups were comparable on the meaning recall test.

Following this model, I conducted a second mixed-effects logistic regression where I

examined the effects of learner- and item-related factors on the meaning recall levels along with the effects of group and time. The best fitting model for the meaning recall test included the main effects of group and test time along with the effects of target pseudowords' part of speech and their frequency of occurrence in the reading text. The comparisons between the model with the two main predictors, the maximal model with all possible predictors, and the best fitting model are given in Appendix L.

The results of the best fitting model (see Table 24) revealed a marginally significant effect of test time on the performance. This means that when covariates were added to the model, the effect of test time was not statistically significant. The nonsignificance of the group effects stayed the same.

The statistically significant predictors of the meaning recall performance were target pseudowords' part of speech and total exposure. The pairwise comparisons for the effects of part of speech with Tukey adjustment $p < .05$ indicated that the meanings of nouns were recalled significantly better than the meanings of adjectives ($b = 1.62$, $SE = 0.64$, $z = 2.55$, $p = .03$). There was no statistically significant difference between the scores for nouns and verbs ($b = 0.69$, $SE = 0.58$, $z = 1.20$, $p = .45$) and verbs and adjectives ($b = 0.93$, $SE = 0.67$, $z = 1.40$, $p = .34$). Moreover, pseudowords with higher frequency of occurrence in the reading text were recalled significantly better than the words with lower frequency of occurrence. Unlike the results for the form recognition test, summed total reading time was not a statistically significant predictor of the performance on the meaning recall test. Participants' vocabulary size, their enjoyment and interest levels, and the items' concreteness levels were not significant predictors of meaning recall performance either.

Table 24

The Best Fitting Model for the Meaning Recall Test

Fixed Effects	*b*	*OR*	*SE*	*95% CIs for OR*	*z*	*p*	*d*
Intercept	-4.01	0.02	0.01	0.01 – 0.05	-8.19	<0.001	2.22
Group (RWL)	0.33	1.39	0.57	0.62 – 3.11	0.81	0.42	0.18
Test Time (delayed)	0.35	1.42	0.26	0.99 – 2.04	1.89	0.06	0.19
Covariates							
Part of Speech (verb)	-0.69	0.50	0.29	0.16 – 1.55	-1.20	0.23	0.38
Part of Speech (adjective)	-1.62	0.20	0.13	0.06 – 0.69	-2.55	0.01	0.90
Total Exposure (z score)	2.25	9.45	4.84	3.46 – 25.79	4.39	<0.001	1.24

Random Effects	*Variance*	*SD*
1|Subject	1.69	1.30
1|Item	0.98	0.99
$N_{Subject}$	59	
N_{Item}	24	
Observations	2830	
Marginal R^2 / Conditional R^2	0.237 / 0.579	

Notes. Intercept: Reading only, immediate test, noun. RWL: Reading while listening.

Meaning Recognition

The model including only the main effects of group and test time did not indicate any statistically significant effects of group (b = 0.26, SE = 0.25, p = .29) or test time (b = -0.05, SE = 0.10, p = .62) on the scores on the meaning recognition test. In other words, the two groups were comparable in terms of the amount of words whose meaning they recognized correctly, and their scores were similar for both testing times.

As in previous tests, I again conducted a second model including the learner-related and item-related variables along with the main predictors, group and test time. The best fitting model for the meaning recognition test included the main effects of group and test time. The other predictors in the model were vocabulary size, interest levels, part of speech, total exposure, and word length. The model comparisons for the meaning recognition test is given in Appendix L.

The results of the best fitting model, presented in Table 25, also did not show any statistically significant effect of group and test time on the performance of the participants. Several learner-level and item-level variables were significant predictors of performance on the meaning recognition test. First, participants with higher vocabulary size recognized the meaning of target pseudowords significantly better than participants with lower vocabulary size. Moreover, there was a positive association between participants' interest levels for the topic of the reading text and their performance on the meaning recall test. That is, participants who found the reading text more interesting learned the meaning of the target pseudowords better at the recognition level than the participants with lower interest levels. Target words' part of speech was another statistically significant predictor of the meaning recognition levels. The pairwise comparisons with Tukey adjustments applied $p < .05$ indicated a statistically significant difference between recognizing the meaning of nouns and adjectives. The meanings of nouns were recognized significantly better than the meanings of adjectives ($b = 0.82$, $SE = 0.35$, $z = -2.38$, $p = .05$). There were no statistically significant differences between the recognition of nouns and verbs ($b = 0.67$, $SE = 0.34$, $z = 2.00$, $p = .11$) and the recognition of verbs and adjectives ($b = 0.15$, $SE = 0.36$, $z = -0.42$, $p = .91$).

Table 25

The Best Fitting Model for the Meaning Recognition Test

Fixed Effects	*b*	*OR*	*SE*	*95% CIs for OR*	*z*	*p*	*d*
Intercept	-2.11	0.12	0.04	0.07 – 0.22	-6.80	<0.001	1.17
Group (RWL)	0.28	1.32	0.30	0.84 – 2.07	1.22	0.22	0.15
Test Time (delayed)	-0.05	0.95	0.11	0.78 – 1.16	-0.51	0.61	0.03
Covariates							
Vocabulary Size (z score)	0.57	1.77	0.42	1.11 – 2.81	2.42	0.02	0.31
Interest (z score)	0.49	1.63	0.39	1.01 – 2.61	2.02	0.04	0.27
Part of Speech (verb)	-0.67	0.51	0.79	0.26 – 0.99	-2.00	0.05	0.37
Part of Speech (adjective)	-0.82	0.44	0.41	0.22 – 0.87	-2.38	0.02	0.45
Total Exposure (z score)	0.63	1.88	0.54	1.07 – 3.30	2.19	0.03	0.35
Word Length (z score)	-0.53	0.59	0.17	0.33 – 1.04	-1.83	0.07	0.29

Random Effects	*Variance*	*SD*
1|Subject	0.57	0.75
1|Item	0.39	0.62
$N_{Subject}$	57	
N_{Item}	24	
Observations	2736	
Marginal R^2 / Conditional R^2	0.111 / 0.311	

Notes. Intercept: Reading only, immediate test, noun. RWL: Reading while listening.

Total exposure also had a statistically significant effect on recognizing the meaning of the target pseudowords. The meanings of the target words that were repeated more times in the reading text were recognized better than the words with fewer occurrences. Finally, the meanings

of shorter words (i.e., five-letter words) were recognized better than the meanings of longer words (i.e., six-letter words). Participants' enjoyment levels, target words' concreteness, and summed total reading times were not statistically significant predictors of performance on the meaning recognition test.

Sentence-Reading Test

For the sentence reading test, I compared the processing of target pseudowords with those of real words and nonwords using two eye-tracking measures: gaze duration and total reading time. The mean scores, standard deviations, and the 95% confidence intervals for the eye-tracking measures for each group at each testing time are given in Table 26 (for gaze duration) and Table 27 (for total reading time). For the reading only and reading while listening groups, gaze duration and total reading times were the shortest for real words, followed by target pseudowords and nonwords respectively. Overall, the two groups' reading times indicated by each measure were similar.

Table 26

Mean Scores, Standard Deviations, and 95% CIs for the Sentence Reading Test (Gaze Duration)

		RO		RWL	
		M (SD)	*95% CIs*	*M (SD)*	*95% CIs*
Immediate Test	Nonword	362.44 (190.96)	[338, 387]	381.12 (265.10)	[347, 415]
	Pseudoword	346.20 (184.20)	[322, 370]	349.98 (226.23)	[321, 379]
	Real word	284.70 (150.20)	[265, 305]	262.91 (132.21)	[246, 280]
Delayed Test	Nonword	355.99 (193.65)	[331, 381]	375.80 (231.65)	[346, 405]
	Pseudoword	334.65 (188.10)	[310, 359]	331.30 (191.42)	[307, 356]
	Real word	251.92 (115.72)	[237, 267]	260.15 (120.73)	[244, 276]

Table 27

Mean Scores, Standard Deviations, and 95% CIs for the Sentence Reading Test (Total Reading Times)

		RO		RWL	
		M (SD)	*95% CIs*	*M (SD)*	*95% CIs*
Immediate Test	Nonword	719.02 (430.17)	[663, 775]	851.71 (519.17)	[785, 919]
	Pseudoword	625.42 (379.01)	[576, 675]	719.79 (455.15)	[662, 778]
	Real word	431.97 (313.90)	[390, 474]	435.03 (264.91)	[401, 470]
Delayed Test	Nonword	688.13 (428.30)	[632, 744]	817.06 (497.92)	[754, 881]
	Pseudoword	678.55 (532.37)	[609, 748]	735.86 (556.56)	[665, 807]
	Real word	362.71 (278.96)	[326, 399]	411.15 (259.02)	[377, 445]

To investigate whether there were any statistically significant differences between the processing of the three word types by the two groups, I conducted two linear mixed-effects models, one for each eye-tracking measure. To investigate whether the change in reading patterns between the three word types were different for the reading only and reading while listening groups, an interaction between group and word type was included in the models. I also added the main effect of test time to the models to examine whether the gaze duration and total reading times differed from the immediate test to the delayed test.

The results for gaze duration are presented in Table 28. This model revealed a statistically significant main effect of word types on the processing, but there was no statistically significant interaction between group and word types. Moreover, the main effect of test time was not statistically significant, indicating similar gaze durations for both testing times. For the reading only group, the results indicated a statistically significant difference between the gaze durations on pseudowords and real words ($d = 0.42$); however, the difference between the gaze

durations on pseudowords and nonwords were not statistically significant (d = 0.08), indicating that gaze durations on pseudowords were similar to those on nonwords. When I releveled the model to compare directly the gaze durations on the nonwords and real words (i.e., nonword as the reference level for item type), the releveled model indicated a statistically significant difference between nonwords and real words (b = -0.25, SE = 0.04, t = -6.34, p < .001, d = 0.50). Nonwords received significantly longer gaze durations than real words. The results were the same for the model where the reference level for test time was the delayed test. In brief, these results did not indicate any processing advantages for the pseudowords compared to nonwords for the gaze durations of the reading only group for any test times.

To directly compare the gaze durations on the three item types for the reading while listening group, I releveled the model and set reading while listening group as the reference level model for group. The model which included reading while listening as the reference model indicated a statistically significant gaze duration difference between pseudowords and real words (b = 0.21, SE = 0.04, t = 5.42, p < .001, d = 0.42) and pseudowords and nonwords (b = -0.10, SE = 0.04, t = -2.52, p = 0.01, d = 0.20). These results illustrate that gaze durations were significantly longer on nonwords compared to pseudowords, and gaze durations on pseudowords were significantly longer than the ones on real words. When the model was releveled using nonwords as the reference level as item type, there was also a statistically significant difference for gaze durations on nonwords and real words (b = 0.31, SE = .04, t = 7.92, p < .001, d = 0.62). The results were consistent for the model where the reference level for test time was the delayed test. Overall, the results from these releveled models showed a processing advantage for the pseudowords compared to nonwords on the gaze durations of the reading while listening group at both testing times. However, despite this processing advantage, pseudowords still received

longer gaze durations than the real words.

None of the learner-related factors, i.e., vocabulary size, enjoyment levels, or interest levels were statistically significant predictors of participants' gaze durations in the sentence reading test.

Table 28

The Best Fitting Model for Gaze Duration on Sentence Reading Test

Fixed Effects	*b*	*SE*	*95% CIs*	*t*	*p*	*Exp(b)*
Intercept	5.73	0.05	5.64 – 5.82	123.66	<0.001	307.97
Group (RWL)	-0.03	0.06	-0.15 – 0.08	-0.57	0.57	0.97
Item Type (Nonword)	0.04	0.04	-0.04 – 0.12	1.03	0.30	1.04
Item Type (Real word)	-0.21	0.04	-0.29 – -0.13	-5.30	<0.001	0.81
Test Time (delayed)	-0.03	0.02	-0.06 – 0.00	-1.68	0.09	0.97
Group (RWL):Item Type (Nonword)	0.06	0.04	-0.02 – 0.14	1.38	0.17	1.06
Group (RWL):Item Type (Real word)	-0.01	0.04	-0.09 – 0.08	-0.05	0.96	0.99
Covariates						
Vocabulary size (z score)	-0.09	0.05	-0.19 – 0.02	-1.64	0.10	0.91
Random Effects	*Variance*	*SD*				
1|Subject	0.04	0.19				
1|Item	0.01	0.09				
Residual	0.20	0.44				
N_{sub}	57					
Observations	2656					
Marginal R^2 / Conditional R^2	0.07 / 0.23					

Notes. Intercept: Reading only, pseudoword, and immediate test. RWL: Reading while listening.

The results of the model for the total reading time are presented in Table 29. This model indicated statistically significant effects of word type and test time on total reading times, but there was no statistically significant interaction between group and word types. The main effect of test time was statistically significant, indicating that the two groups' total reading times were significantly shorter on the delayed test than the immediate test. For the reading only group, the model showed statistically significant differences for total reading times between pseudowords and real words (d = 0.79) and pseudowords and nonwords (d = 0.18). When I releveled the model by setting nonwords as the reference level for item type, this releveled model indicated a statistically significant total reading times difference between nonwords and real words (b = 0.60, SE = 0.06, t = -10.80, p < .001, d = 0.97). The model where delayed test was the reference level indicated the same results. The results showed a processing advantage for the pseudowords when compared to nonwords on the total reading times of the reading only group, but pseudowords were still read significantly slower than real words.

To directly examined the effects of item type on the total reading times of the reading while listening group, I releveled the model to include reading while listening as the reference level of group. This releveled model also indicated a statistically significant difference for total reading times between pseudowords and nonwords (b = - 0.16, SE = 0.06, t = 2.99, p = .004, d = 0.26) and pseudowords and real words (b = 0.51, SE = 0.06, t = -9.24, p < .001, d = 0.82). Moreover, the effect sizes for these differences were greater for the reading while listening group compared to the effects sizes for the reading only group. When this model was releveled to include nonwords as the reference level for item type, a statistically significant difference between nonwords and real words on total reading times were also observed (b = 0.67, SE = 0.06, t = 12.21, p < .001, d = 1.08). These results indicated a processing advantage for the

pseudowords compared to nonwords for the reading while listening group as well, and similar to the reading only group, despite this processing advantage, pseudowords received significantly greater total reading times than real words in the sentence reading test. None of the learner-related factors were statistically significant predictors of the total reading times on the sentence reading test.

Table 29

The Best Fitting Model for Total Reading Time on Sentence Reading Test

Fixed Effects	*b*	*SE*	*95% CIs*	*t*	*p*	*Exp(b)*
Intercept	6.30	0.07	6.15 – 6.44	86.71	<0.001	544.57
Group (RWL)	0.11	0.09	-0.06 – 0.29	1.27	0.21	1.12
Item Type (Nonword)	0.11	0.06	0.00 – 0.22	2.00	0.05	1.12
Item Type (Real word)	-0.49	0.06	-0.60 – -0.38	-8.79	<0.001	0.61
Test Time (Delayed)	-0.06	0.02	-0.10 – -0.02	-3.18	0.001	0.94
Group (RWL):Item Type (Nonword)	0.05	0.05	-0.04 – 0.15	1.14	0.25	1.05
Group (RWL):Item Type (Real word)	-0.02	0.05	-0.11 – 0.07	-0.45	0.65	0.98

Random Effects	*Variance*	*SD*
1|Subject	0.11	0.32
1|Item	0.02	0.15
Residual	0.26	0.51
N_{sub}	59	
Observations	2750	
Marginal R^2 / Conditional R^2	0.17 / 0.45	

Notes. Intercept: Reading only, pseudoword, and immediate test. RWL: Reading while listening.

Summary of Results for Vocabulary Tests

In Table 30, I summarize the results for the vocabulary tests, including the research questions and overall findings for each test.

Table 30

Summary of Vocabulary Test Results

Research Question	Vocabulary Test	Findings
(3a) How does input modality affect the development of declarative knowledge of target words?	Form recognition test	A significant interaction between group and test time • Immediate test: RWL = RO • Delayed test: RWL > RO
	Meaning recall test	Overall, very low learning gains • Immediate test: RWL = RO • Delayed test: RWL = RO
	Meaning recognition test	Overall, low learning gains • Immediate test: RWL = RO • Delayed test: RWL = RO
(3b) Do summed total reading times, learner-related factors, and item-related factors predict the learning gains on the declarative knowledge tests?	Form recognition test	Significant effects of vocabulary size • participants with higher vocabulary size > participants with lower vocabulary size Significant effects of total number of exposures • words with higher frequency of occurrence > words with lower frequency of occurrence Significant effects of word length • longer words > shorter words Significant interaction between summed total reading time and group • higher summed total reading time resulted in increased learning gains for the RWL
	Meaning recall test	Significant effects of total number of exposures • words with higher frequency of occurrence > words with lower frequency of occurrence Significant effects of part of speech • nouns > adjectives • nouns = verbs • verbs = adjectives

Table 30 (cont'd)

Research Question	Vocabulary Test	Findings
	Meaning recognition test	Significant effects of vocabulary size • participants with higher vocabulary size > participants with lower vocabulary size Significant effects of interest level • participants who reported higher interest levels > participants who reported lower interest levels Significant effects of total number of exposures • words with higher frequency of occurrence > words with lower frequency of occurrence Significant effects of part of speech • nouns > adjectives • nouns = verbs • verbs = adjectives Significant effects of word length • longer words > shorter words
(3c) How does input modality affect the development of nondeclarative knowledge of target words?	Sentence reading test	**Gaze Duration:** RO group: no learning gains • pseudowords > real words • pseudowords = nonwords • nonwords > real words RWL group: learning gains • nonwords > pseudowords > real words **Total Reading Time:** RO group: learning gains • nonwords > pseudowords > real words RWL group: learning gains • nonwords > pseudowords > real words
(3d) Do learner-related factors predict the learning gains on the nondeclarative knowledge test?	Sentence reading test	**Gaze Duration:** No effects of learner-related factors **Total Reading Time:** No effects of learner-related factors

Notes. RO: Reading only, RWL: Reading while listening.

CHAPTER 7: DISCUSSION AND CONCLUSION

In this chapter, I discuss the findings in relation to the research questions. I also provide information regarding the pedagogical implications of the findings for L2 learners. Finally, I provide suggestions for further research before concluding the present dissertation.

Processing of Unimodal and Bimodal Input During Reading

In this study, using eye-tracking, I first aimed to investigate the potential benefits of audio support by comparing the processing of 24 target pseudowords in a reading text by reading only and reading while listening groups. For the reading only group, eye-tracking measures indicated a decrease in reading times from the first encounter with the target words to the last. These reduced reading times in the current study are consistent with the results of the previous vocabulary learning studies showing that repeated exposure can speed up the recognition process of unfamiliar words in a text and enhance learning gains (Godfroid et al., 2018; Elgort et al., 2018; Joseph et al., 2014; Mohamed, 2018; Pellicer-Sánchez, 2016). Moreover, this study extended the findings of previous studies to the processing of novel words from bimodal input as the same decrease in reading times was observed for the reading while listening group as well. This shows that even when audio is present, each additional encounter reduces the time L2 learners spend on novel words in a text. So, increased familiarity with novel words influences the reading behaviors of participants in reading only and reading while listening groups similarly.

Despite the general decreasing reading time patterns in the data, a closer look at the four eye-tracking measures indicated different results for the effects of audio support during reading. The results for gaze duration, which represents the early familiarity check and lexical-access stages in reading, pointed out an effect of the presence of audio in the early processing stages of reading. The audio support in reading while listening condition led to increased gaze durations

on the initial encounters with the target pseudowords and a sharper decrease in gaze durations from 10th encounter and forward. Moreover, this decrease in gaze durations was statistically significant for the reading while listening group. The plots of gaze duration showed a decreasing pattern for the reading only group as well, but the overall decrease was less steep, and the inferential models did not indicate a reliable decrease in gaze durations from the first exposure to the last for this group. The increased gaze durations of participants in the reading while listening condition on initial encounters with the target pseudowords compared to the participants in the reading only group can be an indicator of the attention-directing feature of the auditory input provided to these participants (Long, 2017). Bimodal input allows readers to have access to both written and aural forms of unfamiliar words, which can increase the chances of a word being detected as unfamiliar and make an unfamiliar word more salient for the readers. In other words, hearing the phonological form of an unfamiliar word in the audio can encourage readers to spend more time looking at the orthographic form of the word during their first pass before moving their eyes to the next word in the reading. This additional attention on the initial exposures, in return, can speed up the lexical-access process in the later exposures, which is reflected in the sharper decrease in gaze durations of participants in the reading while listening group.

A second explanation for the gaze duration differences between the two groups can be related to the nature of written and aural input. When reading for comprehension, participants might prefer focusing on the main ideas in a text more while focusing on specific words less (Huang et al., 2022). The possibility of adopting such a reading strategy can be higher especially when the knowledge of a particular word adds little to the overall message in a sentence or when the context preceding the unfamiliar word does not provide much information for the meaning of the word. In such cases, instead of focusing on an unfamiliar word, readers might allocate their

attention to other parts of the reading text to increase their overall comprehension. However, when there is an audio accompanying the written input, audio might moderate the reading speed and readers might fixate more and longer on words (Conklin et al., 2020), especially during initial encounters. Once learners get more familiar with the word form, their gaze durations on unfamiliar words might decrease.

For the regression path duration, which represents the ability to fluently access word meanings and integrate them into the preceding context, the results were somewhat similar. Both groups indicated decreasing regression path durations as the frequency of exposure to target pseudowords increased. The decrease in regression path durations seemed more linear compared to the nonlinear decrease in gaze durations. If regression path duration is considered to show effort in integrating the meaning of a word into the preceding context (Godfroid, 2020), the participants in the reading while listening group can be considered to have showed more effort in processing the meaning of unfamiliar words on the initial encounters than the participants in the reading only group. This finding, again, can be related to the increased attention to unfamiliar words induced by the audio on initial encounters in the reading while listening condition.

For rereading times, there was a decrease from the first encounter with a novel word to the last encounter, indicating that as participants' familiarity with the novel pseudowords increased, their tendency to reread these words decreased. In other words, considering that rereading captures the amount of reanalysis on a word (Godfroid, 2020), participants had reduced needs to reanalyze the novel words as they became more familiar with the word. The participants in the reading only group had higher rereading durations than the participants in the reading while listening group, but this difference between the groups was not statistically significant. Despite these statistically nonsignificant differences indicated in the statistical model

for rereading durations, the statements of the participants in the reading while listening group in the retrospective interviews confirmed that these participants did not have as many opportunities to reread certain parts of the text even when they wanted to. Participants in the reading while listening group reported that although the speed of the audio was suitable for them, to keep up with the audio, they could not reread most of the target pseudowords. As argued by Conklin et al. (2020), reading pace is determined by the speed of the audio in reading while listening. Therefore, the fast speed of audio may prevent learners from regressing back to the previous sentences or words in a screen, decreasing the rereading durations. In other words, to keep up with the pacing of the audio, readers may not be able to go back to the previous words. In this respect, like horses pulling a carriage, in reading while listening conditions, readers are pulled by the audio and forced to keep their pace constant if they do not want to miss the upcoming information in the text, and as will be explained in the following sections, keeping their pace constant is proving to be beneficial for learning. The pace of reading in the reading only condition, on the other hand, is determined by individual participants. Since there is not an audio making them to keep a regular pace, readers can reread the previous contexts in a text multiple times and spend more time reanalyzing particular words.

Moreover, there was a curvilinear decrease in total reading times for both groups. Consistent with the S-shaped reading pattern reported in Godfroid et al. (2018), the total reading times showed an initial decrease until the 7th to 8th exposure. This decrease was followed with a slight increase in reading times from the 8th exposure to the 10th before the second decreasing trend from the 10th exposure onwards. These findings, particularly the initial decrease until the 8th exposure fits nicely with the results from previous studies (Elgort et al., 2018; Pellicer-Sánchez, 2016) which indicated that by the 8th exposure reading times for novel words and

known words become similar. The stable pattern with a slight increase from the 8th exposure to the 10th can be an indicator of readers' increased cognitive efforts to infer the meaning of the target pseudowords and integrate their meaning into context. The stability from the 8th to the 10th exposure can also be an indicator of novel impact of each target word encounter to word learning process (Bisson et al., 2014). That is, each encounter with the target word might promote different processing strategies. In the initial encounters with words, L2 readers might spend more time looking at novel words due to their unfamiliar nature, but as they become more familiar with the form of these words they may allocate less attention to them. However, if the L2 readers keep seeing the same words repeatedly without knowing their meanings, after a few encounters (in this study around 8th encounter), they may start thinking that these words could be important for the overall comprehension of a text and try harder to infer their meanings. These findings regarding the changes in reading patterns, again, corroborate the findings of Godfroid et al. (2018), who reported a similar plateau in reading times after an initial decrease for L1 and L2 speakers of English who read a novel with Dari words. Interestingly, in the current study, this S-shaped reading pattern was observed for both the reading only group and the reading while listening group. This means that even with the audio, participants' eye fixations on novel words decreased over time. In other words, even when the audio in the reading while listening condition forces readers to keep a constant pace while reading, these readers still manage to spend to less time fixating on target words that they encounter in text multiple times over time. In that sense, even the existence of the audio does not guarantee allocation of similar amounts of attention on target words occurring multiple times in a text. Finally, although the total reading times were slightly higher for the reading only group, the difference between the two groups was not statistically significant. This finding is similar to the findings of Conklin et al. (2020) who

investigated the reading patterns of L1 and L2 English speakers while reading a text for comprehension (no unknown words) and reported similar average total reading times for L2 speakers in reading only and reading while listening conditions.

Overall, compared to the mean fixation durations on the real words reported by Conklin et al, (2020) for the reading only (336 ms) and reading while listening conditions (325 ms), the mean fixation durations on target pseudowords were higher for both groups in this study (428 ms for the reading only group and 413 ms for the reading while listening group). These increased mean durations indicate that participants spend more time fixating on novel words than known real words while reading even when the audio in the reading while listening condition forces them to keep a constant pace in reading. To better compare the eye movement patterns while reading real words and novel words, one interesting future direction can be investigating the extent to which readers' eye movements align with the audio while reading novel words. Such an investigation can allow for comparing the eye movement patterns on real words reported by Conklin et al. (2020) with the eye movement patterns while reading previously unknown L2 words.

Finally, the reported awareness levels data also complimented the results from the late eye-tracking measures, showing a lack of additional, beneficial influence of the aural input in the reading while listening condition in increasing readers' awareness levels of the target words. Participants in the reading only and reading while listening groups were found to have similar amounts of noetic and autonoetic awareness of the target words they encountered in the reading text. This means that being exposed to the phonologic forms of the target words in the audio along with encountering their orthographic forms in the text did not increase readers' reported familiarity with the target words or their ability to recall more specific details regarding how they

processed these target words in the reading text. Overall, regardless of the reading condition, all participants had higher amounts of noetic and autonoetic awareness of target words in the current study than the amounts reported in Godfroid and Schmidtke (2013). As will be explained in the following section, these increased amounts of noetic and autonoetic awareness reports can be related to higher frequency of exposure to target words in this study compared to the single target word exposure in the study by Godfroid and Schmidtke (2013). This means that rather than audio support, other item-related factors may play a bigger role in increasing readers' awareness of target novel words.

The Effects of Learner-Related and Item-Related Variables in the Processing of Unknown Words

The reading patterns on each eye-tracking measure and participants' reported awarenesss levels were modulated by different learner- and item-related factors. Firstly, participants' vocabulary size influenced their gaze durations on the target pseudowords. Readers with greater vocabulary knowledge had shorter gaze durations on the target pseudowords, indicating quicker lexical access. In other words, participants with higher vocabulary size can be quicker in initial familiarity check of novel words while reading, which includes determining whether a word is familiar to them or not (i.e., word form processing) and if familiar, accessing the meaning of that word from their mental lexicon before moving their eyes to the next word (Reichle et al., 2013). This finding complements the findings from previous studies which indicated the effects of vocabulary knowledge on the processing of known vocabulary (e.g., Conklin et al., 2020) and indicates that the influence of vocabulary size can be observed even on the processing of novel words. Additionally, vocabulary size was a factor increasing readers' chances of developing autonoetic awareness of target words in a text. If autonoetic awareness is considered to denote "a

higher level, or richer quality, of awareness" (Godfroid & Schmidtke, 2013, p. 198), participants with higher vocabulary size can be more successful in developing richer quality awareness than participants with lower vocabulary size. Participants' enjoyment and interest levels, however, do not seem to play a role on their eye-movements while reading a text with or without audio or their reported awareness levels, showing that individual differences at the affective level may not always be reflected in the cognitive processes of L2 speakers during reading (e.g., Kuperman et al., 2023).

Target pseudowords' part of speech was one of the item-level factors that affected the eye movements of the participants while reading. Participants had shorter regression path durations on nouns than verbs and adjectives and shorter total reading times on nouns than verbs. These results are consistent with the findings of previous studies which showed that processing nouns is easier than processing verbs (e.g., Bultena et al., 2014), which may be due to more stable and specific meaning of nouns compared to more-context dependent meaning of verbs (Gentner, 1981). Shorter durations on nouns can also be related to the predictability of these words from context (Clifton et al., 2007; Kliegl et al., 2004). Inferring the meaning of novel nouns in the text might be easier for readers than that of novel verbs and adjectives, which in turn, might decrease the participants' need to go back to the preceding contexts while reading. Another noteworthy item-level factor influencing the eye movements of the participants was concreteness levels of the target pseudowords. More concrete words received higher gaze durations, regression path durations, and total reading times than less concrete words. This finding is in contrast with studies showing that more concrete words are processed faster than less concrete words (Duñabeitia et al., 2009; Juhasz & Rayner, 2003). However, this contrasting finding in current study can be explained by the use of novel words. Since the target pseudowords were novel

words replacing the real words in the text, participants might have shown more cognitive effort in inferring the meaning of more concrete words than less concrete words. In other words, the difficulty in inferring the meaning of less concrete words from context (Schwanenflugel & Stowe, 1989) might have urged participants to ignore these words and focus on the other known words in the context. Also, the majority of more concrete words were nouns or verbs whose meanings can be more important for overall comprehension of the text. Therefore, readers might have allocated more attention to more concrete words than less concrete ones. Finally, consistent with the findings of previous studies (e.g., Juhasz & Rayner, 2003; Kliegl et al., 2004; Lowell & Morris, 2014), longer words (i.e., six-letter words) received longer fixation durations than shorter words (i.e., five-letter words). Moreover, word length was positively associated with the chances of developing awareness of target novel words in a text, which again can point out to attention-grabbing features of longer words while reading (Juhasz & Rayner, 2003).

Additionally, increased summed total reading times on target words while reading resulted in greater awareness levels in the retrospective verbal reports. When participants fixated longer on target words, their possibility of developing awareness of these words, autonoetic awareness in particular, increased. In other words, developing higher level and richer quality of awareness of target words seems to require attention to those words. These results are in line with the findings from Godfroid and Schmidtke (2013) who reported a positive relationship between attention and awareness levels of L2 learners while reading a text containing novel L2 words.

Vocabulary Knowledge Development from Unimodal and Bimodal Input

The results from the three accuracy-based vocabulary tests indicated an incremental development in the knowledge of the target pseudowords, confirming the findings of previous

research (Godfroid et al., 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez & Schmitt, 2010; Webb, 2007, 2008). Participants had the highest learning gains for form recognition, followed by meaning recognition, and meaning recall. These findings were in line with the results of previous studies showing that recognizing forms and meanings of newly learned words is easier than recalling their meanings (Brown et al., 2008; Godfroid et al., 2018; Pellicer-Sánchez, 2016; Pellicer-Sánchez & Schmitt, 2010) as developing productive aspects of newly learned words is more challenging than developing their receptive aspects.

Regarding the comparison between the reading only and reading while listening groups, this study showed similar gains for both groups on all three tests given right after the reading session on Day 2 (i.e., immediate tests). These findings contradict the findings of several studies which showed increased vocabulary learning gains from reading while listening compared to reading only (e.g., Chen, 2021; Malone, 2018; Teng, 2018). Although the findings in the current study may seem unexpected, a closer comparison of the study designs and participants can provide plausible explanations for these contradictory findings. Firstly, the previous studies conducted the whole experiment in one session. In other words, all participants read or read and listened to the chosen reading texts in one session and received the vocabulary posttests immediately after the reading session. In the current study, the reading session was completed over two days. Exposure to target pseudowords over multiple reading sessions (Qiao & Forster, 2013) and the sleep-related consolidation (Dumay & Gaskell, 2007; Lindsay & Gaskell, 2010) that happened between the two sessions might have enhanced the learning of target pseudowords for both groups, eliminating any differences in performance. The hypothesis regarding the effects of sleep consolidation on the learning of new vocabulary from different input modes could be tested in a more tightly controlled, future study. Secondly, the proficiency levels of the

participants were higher and the speech rate of the audio was faster in the current study than the previous studies. These variations may also account for the difference in the findings.

Moreover, participants' reading behaviors reflected in the eye-tracking measures can also explain the lack of difference between the two groups. As explained in the previous section, while participants in the reading while listening group had higher gaze durations, participants in the reading only group had higher rereading durations on the target pseudowords. These differences in the early and late reading processes might have neutralized any additional benefits of the audio in the reading while listening group. In fact, the total reading times of the two groups were comparable with similar S-shaped reading patterns. When this similarity in total reading times is considered, the lack of difference in learning gains for the immediate tests may not be very surprising.

Interestingly, on the delayed test, there was an increase in form recognition scores of the participants in the reading while listening group whereas the same increase was not observed for the participants in the reading only group. This finding is interesting because it shows that the auditory input in reading while listening can influence the retention levels of the participants positively although this influence may not be observed immediately after the treatment session. This means that being exposed to the phonologic forms of the target words along with their orthographic forms might strengthen the representations of these words in memory. This finding is in line with the findings of the studies indicating the positive effects of exposure to orthographic forms in learning the phonologic forms of novel words (Escudero, 2015; Escudero et al., 2008; Escudero et al., 2014). The results in this study show that this positive relationship between visual and auditory input can be bidirectional. That is, exposure to the phonologic forms of novel words can enhance the learning and retention of their orthographic forms. The increased

scores of the reading while listening group on the delayed test can also be explained by testing effects, i.e., encountering the same words in the immediate meaning recall and meaning recognition tests. However, I believe this explanation is less probable considering the lack of a similar increase for the reading only group. If the increase in form recognition scores in the delayed test was due to the additional exposure to the target items in the other two tests, we would expect to see a similar increase for the participants in the reading only group as well.

Unlike the recognition tests, there was a testing effect on the delayed meaning recall test. Both groups' meaning recall scores were higher in the delayed test than in the immediate test, showing an effect of the immediate meaning recognition test on the delayed meaning recall test. Increased meaning recall levels at the delayed test indicate that additional exposure to target pseudowords and their possible meanings in the meaning recognition test can influence the creation of form-meaning links of target pseudowords in readers' mental lexicon. In other words, when exposed to the meaning of a target words in a test, even when its meaning is accompanied with distractors, can urge L2 learners to search for an appropriate response in their mental lexicon (Lindstromberg, 2020, p . 243) and increase their chances of learning the meaning of that words at retrieval level. However, even with this testing effect, the overall learning gains for both groups did not exceed 7%, indicating a floor effect for this test. The testing effect was not observed for participants' meaning recognition levels. Participants' scores on the delayed meaning recognition test were similar to their scores on the immediate test.

The Effects of Learner-Related and Item-Related Variables on the Learning Gains

Total number of exposures had a positive effect on form recognition, meaning recall, and meaning recognition levels of the participants. On all three tests, participants had higher scores for the words that were repeated multiple times in the reading text, and their scores increased as

the number of exposures increased. Moreover, frequency of exposure had the largest effect sizes as a predictor of learning gains on all three tests (based on the standardized coefficients). These findings add credence to the findings of previous studies that indicated positive effects of exposure frequency on the learning of L2 vocabulary in incidental learning environments (Elgort & Warren, 2014; Godfroid et al., 2018; Malone, 2018; Mohamed, 2018; Tekmen & Daloğlu, 2006; Vidal, 2011; Webb, 2007; Zahar et al., 2001).

Summed total reading time had a positive effect on form recognition, but it did not influence the meaning recognition and recall levels of the participants. These results showed that frequency of exposure had a greater influence on the learning of target pseudowords than the summed total reading times. These results are the opposite of the findings of Godfroid et al. (2018) and Pellicer-Sanchez (2016), who reported a positive effect of summed total reading time in learning of the meanings of target words but not in learning their forms. One reason for contradicting findings, particularly for the learning of meanings of novel words, can be the low learning rates in the meaning tests in the current study. However, this cannot be the only explanation as several other studies (e.g., Elgort et al., 2018; Bisson et al., 2015) also failed to find any significant relationship between reading times and vocabulary scores in a meaning recall test. As argued by Pellicer-Sánchez (2020), these discrepancies in the findings of different studies regarding the relationship between reading times and vocabulary learning gains "demonstrate the complexity of the relationship between reading times and performance measures and question the direct link between amount of attention and learning gains" (p. 142).

Interestingly, the effects of summed total reading time on form recognition was even higher for the participants in the reading while listening group. This shows that resisting the fleeting nature of the audio and spending more time fixating on the target words while reading

can be helpful for the learning of word forms. In other words, as explained above, while simultaneous exposure to the phonologic and orthographic forms of target words in the reading while listening condition is proven to be beneficial for the learning and retention of novel word forms, when readers spend more time looking at the target words despite the ongoing audio, their chances of learning of the target word forms augments even more.

Another item-level factor that affected the learning of target pseudowords, particularly the learning of their meanings, was part of speech. In line with the results of previous studies (Ellis & Beaton, 1993; Godfroid et al., 2018; Van Zeeland & Schmitt, 2013), meanings of the nouns were recognized and recalled better than the meanings of verbs and adjectives. However, the same effect was not observed in recognizing the forms of the target pseudowords. This contrast in the effects of part of speech on learning different aspects of word knowledge can be explained by the differences in processing word forms and meanings while reading. In their cognitive reading model, Khalifa and Weir (2009) states that lexical access to words, including retrieving word class and meaning, can be achieved only after the recognition of word forms. In other words, according to Khalifa and Weir's model, word recognition does not require access to any information regarding the word meaning including information about its part of speech. Therefore, while reading a text with novel words, L2 learners can increase their familiarity with novel word forms before trying to determine their part of speech or meaning. Therefore, part of speech can be a more relevant factor in learning meaning rather than learning form. Overall, the results regarding the effects of part of speech on learning gains show that part of speech can be a factor in creating form-meaning links of novel words, but it may not have a big impact in learning their forms.

Unlike part of speech, target words' concreteness levels did not have an influence on the

learning of form or meaning of the target pseudowords. This finding contrasted with the findings of previous research which reported higher learning gains for more concrete words (Elgort & Warren, 2014; Van Zeeland & Schmitt, 2013). Considering that more concrete words received longer fixations durations (gaze durations, rereading times, and total reading times) than less concrete words, the lack of concreteness effects on learning again shows that total reading times may be not be the best indicator of vocabulary learning gains, at least in the current study. Moreover, the conflicting results with the previous studies show that concreteness may not be a reliable predictor of vocabulary learning in different incidental learning environments, and that more research is needed.

As predicted, participants' vocabulary size influenced the recognition of form and meaning of the target pseudowords. Consistent with the findings of previous research (Elgort & Warren, 2014; Horst et al., 1998; Tekmen & Daloğlu, 2006; Webb & Chang, 2015; Zahar et al., 2001), participants with larger vocabulary size had higher learning gains than participants with lower vocabulary size. The same effects were not observed on the meaning recall test, yet this lack of influence can be due to the very low accuracy scores on the meaning recall test. Therefore, no strong conclusions should be drawn for the effects of vocabulary size on meaning recall levels of participants.

One final noteworthy finding was the positive effect of participants' interest levels on their meaning recognition scores. Participants who found the topic of the reading text more interesting recognized the meaning of target pseudowords better. This finding was in line with the findings of Lee and Pulido (2017) and Elgort and Warren (2014). This positive effect of interest on learning demonstrates that a good reading experience can boost the creation of form-meaning links for novel words but not the recognition of form. In other words, when readers find

the topic of a text interesting, they might want to learn more about the content and possibly try to infer the meanings of unknown words more. However, noticing levels for the novel words in a text can be similar regardless of readers' interest levels.

Processing of Novel Words in New Contexts

The analysis of the eye movements, namely gaze durations and total reading times, on the sentence reading task indicated that participants started learning the orthographic forms of the target pseudowords. The learning effects were more pronounced for the participants in the reading while listening group. Participants in the reading while listening group had significantly shorter gaze durations and total reading times on pseudowords than nonwords. These differences between the processing of target pseudowords and nonwords indicate that the orthographic representations of the target pseudowords were established during reading while listening to the treatment text. However, the gaze durations and total reading times on the real words were even shorter than the ones on target pseudowords, which indicates that the target pseudowords' orthographic representations in the mental lexicon were less robust than those of the real words. Moreover, these processing differences between the target pseudowords and real words were present even for the target pseudowords with the highest frequency of exposure in the reading text. For example, the mean total reading time for the three most frequently occurring pseudowords in the text, *rolley*, *prome*, and *crimb*, was 607.53 (SD = 403.79) whereas the mean total reading time for all real words was 423.17 (SD = 261.99). This indicates that reading while listening can be helpful for creating the orthographic representations of the novel words in the mental lexicon, but establishing robust representations of novel words requires more than 16 encounters, possibly in various contexts.

Similar to the participants in the reading while listening group, participants in the reading

only group also had the shortest gaze durations and total reading times on the real words. However, for the participants in the reading only group, the differences in the gaze durations on nonwords and target pseudowords were not statistically different, indicating a lack of learning effect on the target pseudowords, at least for the quick lexical access. This finding shows that written input alone may not be sufficient for the creation of strong lexical representations of novel words which can be accessed easily during real-time processing. Contrarily, the difference between the total reading times of nonwords and target pseudowords were statistically significant, which indicated that the participants in the reading only group also started learning the orthographic representations of the target pseudowords. This result is consistent with the findings of Elgort et al. (2018) who reported that creating lexical representations of novel words from reading was possible, but the created lexical representations were not as fine-tuned as the representations of already known control words. Interestingly, however, the effect sizes obtained for the reading only group was not as high as the ones obtained for the reading only group. Combined together, the lower effect sizes in the sentence reading test and the changes in the form recognition tests one week after the learning sessions (as explained above) show that access to the written input alone may be less efficient in creating lexical representations of novel words that can be retained over a longer period of time. Instead, the audio in the reading while listening condition can be more beneficial for achieving stronger lexical representations of novel word forms.

When taken together, the results for the sentence reading test show an advantage for the participants in the reading while listening group in developing lexical representations of the target pseudowords and accessing in a real-time activity. Having access to the phonologic forms of novel words through audio can enhance the lexical representations of novel words in mind,

which in turn can increase the chances of retrieving these words more fluently in a real-time activity such as reading. However, as explained in the previous section, these advantages may not always be reflected in untimed, accuracy-based tests of form and meaning. Overall, the results of all vocabulary tests point to a slow and incremental vocabulary knowledge development from both unimodal and bimodal input.

Implications

The findings of this dissertation have important theoretical and pedagogical implications. From a theoretical perspective, the findings shed light on the differences and similarities between the processing of unimodal and bimodal input by L2 speakers of English. This study provided empirical evidence regarding how L2 speakers cognitively engage with novel words in a reading text with or without audio. The early eye-tracking measures indicated an effect of audio in augmenting attention to novel words in early reading processes, i.e., familiarity check and lexical access, confirming Long's (2017) hypothesis regarding the increase in attention thanks to audio support in bimodal input. Despite being statistically insignificant, the late eye-tracking measures, on the other hand, showed the opposite reading behaviors, and unimodal input was found to provide more opportunities for rereading and integrating the novel words into context. Overall, both input modes resulted in an S-shaped change in the reading times from the first encounter with novel words to the last, providing empirical evidence for nonlinearity in vocabulary development in both unimodal and bimodal input modes. Finally, the data from verbal retrospective reports demonstrated evidence that both input modes can be conducive for the development of difference awareness levels for different target words, and one input mode does not transcend the other in achieving greater awareness for target words.

Moreover, the results from the eye-tracking measures and vocabulary tests provide

several pedagogical implications. Firstly, unimodal and bimodal input influenced the processing of novel words differently as reflected in the early and late eye-tracking measures. However, the differences in the processing of unimodal and bimodal input were in milliseconds, so these very small differences in processing may not be directly observed in the performance of L2 learners in language classrooms. In fact, in two of the three accuracy-based vocabulary measures, the two groups had comparable learning scores, indicating equal effectiveness of unimodal and bimodal input in learning the meanings of novel words in incidental learning environments. Additionally, the two input modes were equally effective in inducing greater awareness of target words in a reading text. One difference between the groups was observed in the recognition of the forms of the novel words when tested one week after the learning sessions. The second difference was in the sentence reading test, which indicated an advantage for the reading while listening group and pointed to more robust lexical representations for this group. Based on these results, I would argue that both input modes can be helpful for learning L2 words, so teachers may encourage L2 learners to read different books with or without audio. As an additional advantage, reading with audio can enable L2 learners to learn the phonologic forms of the novel words, which can be difficult from reading alone. Particularly when a text contains target words with unconventional phonologic forms (e.g., words that are pronounced differently than expected based on their orthographic forms), bimodal input can be notably more favorable. L2 learners' overall reading behaviors in the two input modes can be an additional deciding factor for teachers regarding which input mode to use in language classrooms. As reported at the beginning of the discussion chapter, the two groups' mean reading times on the text were similar; however, compared to the reading while listening group, participants in the reading only group showed a greater variability in their mean reading times. While some of the L2 speakers in this group read the text quickly,

other students were more careful in their reading and seemed to pay more attention to details in the text. Participants in the reading while listening group, on the other hand, were very similar in their mean reading times as they had to follow the pace of the audio while reading. If teachers want all of their students to read a text at a similar rate and pay similar amounts of attention to different words in the text, reading while listening can be a better instruction technique. However, if the main aim for reading is comprehension, reading alone can be enough for students to comprehend the text without spending too much class time on one activity.

Secondly, the overall learning gains from both input modes were low, particularly on the tests assessing L2 learners' knowledge of meaning, which reflects the slow and incremental development of vocabulary knowledge in incidental learning environments. These results show that reading for pleasure is not a quick and easy method of learning new vocabulary even when the written input is accompanied with aural input. However, this does not mean that reading for pleasure is completely useless. In fact, the results from the form recognition test and sentence-reading test is still fairly encouraging as these tests showed that vocabulary learning is possible in incidental learning environments, which reinforces the idea that reading for pleasure, with or without audio, can be a good pedagogic activity for L2 learners in or outside the language classrooms on the long term. However, for quicker vocabulary development, particularly in classrooms, learning from reading can be supported with several additional deliberate teaching methods, which can be helpful for increasing L2 learners' attention to and awareness of new L2 words. First, as presented in the results section, higher frequency of exposure to target words leads to increased learning gains, highlighting the importance of repetition on the processing and learning of novel words once more. Considering these beneficial effects of repetition in learning new words and consolidating the learned knowledge, teachers can use different techniques and

activities to increase L2 learners' exposure to target vocabulary to augment their learning gains in language classrooms. Second, teachers can accelerate and maximize vocabulary learning by combining reading only and reading while listening activities with additional activities that can increase learners' awareness of and attention to target words in the text such as providing pre-reading or post-reading instructions to students (e.g., Elgort et al., 2023; Pellicer-Sánchez et al., 2021, 2022) or using different textual enhancement techniques (e.g., Puimège et al., 2023; Sonbul & Schmitt, 2013; Toomer & Elgort, 2019).

Limitations and Future Directions

To my knowledge, this study is one of the first to examine the effects of unimodal and bimodal input on the reading behaviors and learning gains of L2 speakers. Despite its importance, this study was not without limitations. Firstly, the target words in the reading text consisted of pseudowords which replaced high-frequency real words, and participants learned *additional* labels for concepts for which they presumably already have other L2 words in their mental lexicon (e.g., *mittle* vs. *face*). Therefore, vocabulary learning in this study may not represent the learning of a word for a new concept (e.g., tool names in Joseph et al.'s (2014) study) or an L1 concept for which an L2 word is not known yet. How unimodal and bimodal input influences the processing and learning of new L2 concepts should be examined in future studies.

Secondly, the use of pseudowords instead of real words guaranteed that participants had no prior knowledge of the target words and that learning outside of the experiment was not possible. However, the use of pseudowords in a text consisting of mostly high-frequency words might have increased the salience of the target words. Nevertheless, during the interviews, only two participants reported suspecting that the novel words in the text were pseudowords in

English; the remaining participants reported that they assumed these pseudowords were old English words that they were unfamiliar with. In fact, the practice text given at the beginning of the experiment might have helped with the participants' assumption since the practice text stated that *The Time Machine* was published in 1895.

Thirdly, some of the target pseudowords phonetically matched with real words in English (e.g., merky vs. murky); however, the effects of these phonetical relatedness with real English words were not examined in this study. Future research should examine if and how this phonetical resemblance influences the learning from unimodal and bimodal input.

Another limitation of the study can be the use of multiple vocabulary measures. At each testing time, immediate and delayed, participants completed four different vocabulary tests: sentence-reading test, form recognition test, meaning recall test, and meaning recognition test respectively. I used distractor words in the form recognition tests to control for learning from the sentence reading test, and the percentage of chosen distractors in the form recognition test was low, which showed that participants were able to select the words they remember encountering in the reading text. However, exposure to several target pseudowords (eight target pseudowords) along with several nonwords (eight nonwords) in the sentence-reading test could have still affected the scores on the form recognition test, but this effect should have been similar for both groups since they completed the tests in the same order. The effects of the sentence-reading test and form-recognition test on the meaning recall and meaning recognition tests are assumed to be minimal because (1) the target pseudowords were given in neutral sentences in the sentence-reading test and (2) the form-recognition test did not include any context for the target words, making the meaning inference impossible. Moreover, there was only a one-week delay between the learning sessions and delayed tests, so the retention amounts given in this study may not be

applicable for longer term retention. Future studies should examine the retention rates of the novel words in a longer term (e.g., one month after the learning sessions) controlling for possible testing effects.

Another limitation of the study is diversity in the L1 backgrounds of the participants. Although participants with the same L1s were distributed to two groups as equally as possible to make the groups comparable, it is not possible to determine how L1-L2 linguistic distance, a factor affecting L2 reading comprehension (Jeon & Yamashita, 2014), influenced the learning from unimodal and bimodal input. Future studies should investigate the possible influences of L1 backgrounds (e.g., different language families or different writing) on the processing and learning of novel words in unimodal and bimodal input.

Additionally, the main aim of this study was to investigate the differences in reading behaviors and learning gains of participants in reading only and reading while listening groups. Therefore, I did not include familiar English words as control words in the reading text, which could have provided more information about how the frequency of exposure influenced the reading times of target pseudowords compared to control words. Future studies can use a more controlled design to examine the additional possible lexical and contextual factors influencing the processing of target words in a text.

The speech rate for the audio in the current study was based on previous studies and piloting different speech rates with L2 learners with similar profiles to the participants in the main study. However, speech rate can be a factor affecting novel word processing and learning while reading and listening to a text. Therefore, future studies can replicate this study by using different speech rates to examine if and how speech rate can be a factor in processing of and learning from bimodal input.

Conclusion

This study is one of the first studies comparing the processing of novel words in unimodal and bimodal input conditions along with their effectiveness in promoting the learning of novel words in a text. Eye-tracking results for the treatment text indicated different reading behaviors for the reading only and reading while listening groups for early and late reading processes. While participants in the reading only group showed increased attention to novel words during the early reading processes, participants in the reading only group had higher attention to novel words during later stages of reading. Regardless of these differences between groups, an S-shaped, nonlinear decrease in reading times were observed for both groups on all eye-tracking measures from the first exposure with novel words to the last. The results from the retrospective verbal reports also demonstrated similar awareness levels for both groups. The accuracy-based vocabulary tests showed similar learning gains for reading only and reading while listening groups, and the highest learning gains were achieved for the words that participants encountered more in the text, underscoring the importance of repetition on the learning of new vocabulary. The only difference between the two groups was observed for the delayed form recognition test with higher scores for the reading while listening group. This result underlined the importance of access to phonologic forms along with orthographic forms for the recognition of form. The results of the sentence-reading test indicated a learning effect for the target words; participants in the reading while listening group had more robust lexical representations of the target words than the participants in the reading only group.

Overall, this study provided a more comprehensive picture of unimodal and bimodal input by showing how L2 learners cognitively engage with novel words in a text during reading with or without audio. The use of sentence-readings test along with traditional form recognition,

meaning recall, and meaning recognition tests enabled exploring how newly learned words were read in neutral contexts compared to nonwords and high-frequency real words. The present study is hoped to inspire more research examining the cognitive processes underlying learning from different input modes and the development of different vocabulary knowledge aspects and not only the accuracy in recognizing and recalling words in untimed situations.

BIBLIOGRAPHY

Andringa, S. (2020). The emergence of awareness in uninstructed L2 learning: A visual world eye tracking study. *Second Language Research*, *36*(3), 335–357. https://doi.org/10.1177/0267658320915502

Andringa, S., & Rebuschat, P. (2015). New directions in the study of implicit and explicit learning: An introduction. *Studies in Second Language Acquisition*, *37*(2), 185–196. https://doi.org/10.1017/S027226311500008X

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Baayen, H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28. https://doi.org/10.21500/20112084.807

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Barcroft, J. (2004). Second Language Vocabulary Acquisition: A Lexical Input Processing Approach. *Foreign Language Annals*, *37*(2), 200–208. https://doi.org/10.1111/j.1944-9720.2004.tb02193.x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bisson, M., Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). The role of repeated exposure to multimodal input in incidental acquisition of foreign language vocabulary. *Language Learning*, *64*(4), 855–877. https://doi.org/10.1111/lang.12085

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2015). The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. *Quarterly Journal of Experimental Psychology*, *68*(7), 1306–1326. https://doi.org/10.1080/17470218.2014.979211

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*(2), 136–163

Bultena, S., Dijkstra, T., & van Hell, J. G. (2014). Cognate effects in sentence context depend on word class, L2 proficiency, and task. *Quarterly Journal of Experimental Psychology*, *67*(6), 1214–1241. https://doi.org/10.1080/17470218.2013.853090

Carrol, G., & Conklin, K. (2020). Is All Formulaic Language Created Equal? Unpacking the Processing Advantage for Different Types of Formulaic Sequences. *Language and Speech*, *63*(1), 95–122. https://doi.org/10.1177/0023830918823230

Chen, Y. (2021). Comparing incidental vocabulary learning from reading-only and reading-while-listening. *System*, *97*, 102442. https://doi.org/10.1016/j.system.2020.102442

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, *19*, 3127–3131. https://doi.org/10.1002/1097-0258(20001130)19:22<3127::aid-sim784>3.0.co;2-m

Choi, I. (2018). *Learning words under incidental and intentional learning conditions: An eye-tracking study* [Unpublished doctoral dissertation]. Michigan State University

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, *21*(3), 403–426. https://doi.org/10.1177/1362168816653271

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*(1), 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

Cleeremans, A. (2011). The radical plasticity thesis: How the brain learns to be conscious. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00086

Clifton, C. Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–371). Elsevier

Cobb, T. (n.d.). Compleat lexical tutor, [Online application software] Version 8.3. https://www.lextutor.ca

Conklin, K., Alotaibi, S., Pellicer-Sánchez, A., & Vilkaitė-Lozdienė, L. (2020). What eye-tracking tells us about reading-only and reading-while-listening in a first and second language. *Second Language Research*, *36*(3), 257–276. https://doi.org/10.1177/0267658320921496

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research*, *32*(3), 453–467. https://doi.org/10.1177/0267658316637401

Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLOS ONE*, *10*(8), e0134008. https://doi.org/10.1371/journal.pone.0134008

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, *26*(2), 197–223. https://doi.org/10.1007/s10648-013-9246-y

Crossley, S., Kyle, K., & Salsbury, T. (2016). A usage-based investigation of l2 lexical acquisition: The role of input and output. *The Modern Language Journal*, *100*(3), 702–715. https://doi.org/10.1111/modl.12344

Dang, T. N. Y., Lu, C., & Webb, S. (2022). Incidental learning of collocations in an academic lecture through different input modes. *Language Learning*, 1–37. https://doi.org/10.1111/lang.12499

de Groot, A. M. B. (2006). Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. *Language Learning*, *56*(3), 463–506. https://doi.org/10.1111/j.1467-9922.2006.00374.x

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Duñabeitia, J. A., Avilés, A., Afonso, O., Scheepers, C., & Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: Evidence from the visual-world paradigm. *Cognition*, *110*(2), 284–292. https://doi.org/10.1016/j.cognition.2008.11.012

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, *40*(2), 341–366. https://doi.org/10.1017/S0272263117000109

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*(2), 365–414. https://doi.org/10.1111/lang.12052

Elgort, I., Wetering, R. V. D., Arrow, T., & Beyersmann, E. (2023). Previewing novel words before reading affects their processing during reading: An eye-movement study with first and second language readers. *Language Learning*, lang.12579. https://doi.org/10.1111/lang.12579

Ellis, N. (1994). Consciousness in second language learning: Psychological perpectives on the role of conscious processes in vocabulary acquisition. *AILA Review*, *11*, 37–56

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, *43*(4), 559–617. https://doi.org/10.1111/j.1467-1770.1993.tb00627.x

Ender, A. (2016). Implicit and Explicit Cognitive Processes in Incidental Vocabulary Acquisition. *Applied Linguistics*, *37*(4), 536–560. https://doi.org/10.1093/applin/amu051

Erçetin, G. (2010). Effects of topic interest and prior knowledge on text recall and annotation use in reading a hypermedia text in the L2. *ReCALL*, *22*(2), 228–246. https://doi.org/10.1017/S0958344010000091

Escudero, P. (2015). Orthography plays a limited role when learning the phonological forms of new words: The case of Spanish and English learners of novel Dutch words. *Applied Psycholinguistics*, *36*(1), 7–22. https://doi.org/10.1017/S014271641400040X

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*(2), 345–360. https://doi.org/10.1016/j.wocn.2007.11.002

Escudero, P., Simon, E., & Mulak, K. E. (2014). Learning words in a new language: Orthography doesn't always help. *Bilingualism: Language and Cognition*, *17*(2), 384–395. https://doi.org/10.1017/S1366728913000436

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538. https://doi.org/10.1037/a0015808

Gardiner, J. M. (2001). Episodic memory and autonoetic consciousness: A first–person approach. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1413), 1351–1361. https://doi.org/10.1098/rstb.2001.0955

Gass, S. (1999). Incidental vocabulary learning. *Studies in Second Language Acquisition*, *21*(2), 319–333. https://doi.org/10.1017/S0272263199002090

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory*, *4*(2), 161–178

Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A., & Yoon, H.-J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, *21*(3), 563–584. https://doi.org/10.1017/S1366728917000219

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*(3), 483–517. https://doi.org/10.1017/S0272263113000119

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 J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 183–205). University of Hawai'i, National Foreign Language Resource Center

Godfroid, A., & Winke, P. (2015). Investigating implicit and explicit processing using L2 learners' eye-movement data. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 325–348). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.48.14god

Griffiths, R. (1990). Speech Rate and NNS Comprehension: A Preliminary Study in Time-Benefit Analysis*. *Language Learning*, *40*(3), 311–336. https://doi.org/10.1111/j.1467-1770.1990.tb00666.x

Hirotani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, *54*(3), 425–443. https://doi.org/10.1016/j.jml.2005.12.001

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*, 207–223

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension . *Reading in a Foreign Language*, *13*(1), 403–430

Huang, L., Ouyang, J., & Jiang, J. (2022). The relationship of word processing with L2 reading comprehension and working memory: Insights from eye-tracking. *Learning and Individual Differences*, *95*, 102143. https://doi.org/10.1016/j.lindif.2022.102143

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). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470756492.ch12

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Studies in Second Language Acquisition*, *27*(02). https://doi.org/10.1017/S0272263105050084

Hulstijn, J. H. (2013). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2632–2640). Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0530

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *51*(3), 539–558. https://doi.org/10.1111/0023-8333.00164

Indrarathne, B., & Kormos, J. (2017). Attentional processing of input in explicit and implicit conditions: An eye-tracking study. *Studies in Second Language Acquisition*, *39*(3), 401–430. https://doi.org/10.1017/S027226311600019X

Issa, B. I., & Morgan-Short, K. (2019). Effects of external and internal attentional manipulations on second language grammar development: An eye-tracking study. *Studies in Second Language Acquisition*, *41*(2), 389–417. https://doi.org/10.1017/S027226311800013X

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis: L2 reading and its correlates. *Language Learning*, *64*(1), 160–212. https://doi.org/10.1111/lang.12034

Jiménez, L. (2003). Intention, attention, and consciousness in probabilistic sequence learning. In L. Jiménez (Ed.), *Attention and implicit learning* (Vol. 48, pp. 43–68). John Benjamins Publishing Company. https://doi.org/10.1075/aicr.48.06jim

Joseph, H. S. S. L., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition*, *133*(1), 238–248. https://doi.org/10.1016/j.cognition.2014.06.015

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1312–1318. https://doi.org/10.1037/0278-7393.29.6.1312

Jung, J., & Révész, A. (2018). The effects of reading activity characteristics on L2 reading processes and noticing of glossed constructions. *Studies in Second Language Acquisition*, *40*(4), 755–780. https://doi.org/10.1017/S0272263118000165

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

Khalifa, H., & Weir, C. J. (2009). *Examining reading: Research and practice in assessing second language reading.* Cambridge University Press

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 262–284. https://doi.org/10.1080/09541440340000213

Kuperman, V., Siegelman, N., Schroeder, S., Acartürk, C., Alexeeva, S., Amenta, S., Bertram, R., Bonandrini, R., Brysbaert, M., Chernova, D., Da Fonseca, S. M., Dirix, N., Duyck, W., Fella, A., Frost, R., Gattei, C. A., Kalaitzi, A., Lõo, K., Marelli, M., … Usal, K. A. (2023). Text reading in English as a second language: Evidence from the multilingual eye-movements corpus. *Studies in Second Language Acquisition*, *45*(1), 3–37. https://doi.org/10.1017/S0272263121000954

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, *7*(1), 12–18. https://doi.org/10.1016/S1364-6613(02)00013-X

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, *29*(4), 694–716. https://doi.org/10.1093/applin/amn018

Lee, S., & Pulido, D. (2017). The impact of topic interest, L2 proficiency, and gender on EFL incidental vocabulary acquisition through reading. *Language Teaching Research*, *21*(1), 118–135. https://doi.org/10.1177/1362168816637381

Li, C., & Wei, L. (2022). Anxiety, enjoyment, and boredom in language learning amongst junior secondary students in rural China: How do they contribute to L2 achievement? *Studies in Second Language Acquisition*, 1–16. https://doi.org/10.1017/S0272263122000031

Lindsay, S., & Gaskell, M. G. (2010). A complementary systems account of word learning in L1 and L2. *Language Learning*, *60*, 45–63. https://doi.org/10.1111/j.1467-9922.2010.00600.x

Lindstromberg, S. (2020). Intentional L2 vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 240–254). Routledge

Loewen, S. (2020). *Introduction to instructed second language acquisition* (2nd ed.). Routledge

Long, M. M. (2017). Instructed second language acquisition (ISLA). *Instructed Second Language Acquisition*, *1*(1), 7–44. https://doi.org/10.1558/isla.33314

Lowell, R., & Morris, R. K. (2014). Word length effects on novel words: Evidence from eye movements. *Attention, Perception, & Psychophysics*, *76*, 179–189. https://doi.org/10.3758/ s13414-013-0556-4

Malone, J. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition*, *40*(3), 651–675. https://doi.org/10.1017/S0272263117000341

Mirman, D. (2014). *Growth curve analysis and visualization using R*. CRC Press/Taylor & Francis Group

Mohamed, A. A. (2018). Exposure frequency in L2reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*, *40*(2), 269–293. https://doi.org/10.1017/S0272263117000092

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video the role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, *42*(4), 749–773. https://doi.org/10.1017/S0272263119000706

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, *5*, 12–25

Nation, I. S. P. (2012). *Vocabulary Size Test information and specifications.* http://www.victoria.ac.nz/lals/about/staff/paul-nation

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*, 9–13

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*(1), 97–130. https://doi.org/10.1017/S0272263115000224

Pellicer-Sánchez, A. (2020). Expanding English vocabulary knowledge through reading: Insights from eye-tracking studies. *RELC Journal*, *51*(1), 134–146. https://doi.org/10.1177/0033688220906904

Pellicer-Sánchez, A., Conklin, K., & Vilkaitė-Lozdienė, L. (2021). The effect of pre-reading instruction on vocabulary learning: An investigation of L1 and L2 readers' eye movements. *Language Learning*, *71*(1), 162–203. https://doi.org/10.1111/lang.12430

Pellicer-Sánchez, A., Conklin, K., & Vilkaitė-Lozdienė, L. (2022). (Re)examining the benefits of pre-reading instruction for vocabulary learning. *TESOL Quarterly*, *56*(1), 363–375. https://doi.org/10.1002/tesq.3058

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart? *Reading in a Foreign Language*, *22*(1), 31–55

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*(1), 113–151. https://doi.org/10.1111/j.1467-9922.2009.00502.x

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Puimège, E., Montero Perez, M., & Peters, E. (2023). The effects of typographic enhancement on L2 collocation processing and learning from reading: An eye-tracking study. *Applied Linguistics*, 1–24. https://doi.org/10.1093/applin/amad003

Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, *47*(4), 424–438. https://doi.org/10.1080/09571736.2019.1638630

Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *42*(3), 525–549. https://doi.org/10.1017/S027226311900055X

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the e-z reader model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 720–732. https://doi.org/10.1037/0096-1523.30.4.720

Rayner, K., & Pollatsek, A. (2006). Eye-movement control in reading. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed, pp. 613–658). Elsevier

Reber, A. S., Allen, R., & Reber, P. J. (1999). Implicit versus explicit learning. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 475–513). MIT Press

Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2013). Using E-Z Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review*, *33*(2), 110–149. https://doi.org/10.1016/j.dr.2013.03.001

Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using E-Z Reader to simulate eye movements in nonreading tasks: A unified framework for understanding the eye–mind link. *Psychological Review*, *119*(1), 155–185. https://doi.org/10.1037/a0026473

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning*, *45*(2), 283–331. https://doi.org/10.1111/j.1467-1770.1995.tb00441.x

Robinson, P. (2017). Attention and awareness. In J. Cenoz, D. Gorter, & S. May (Eds.), *Language awareness and multilingualism* (3rd ed., pp. 125–134). Springer

Robinson, P., Mackey, A., Gass, S. M., & Schmidt, R. (2012). Attention and awareness in second language acquisition. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 247–267). Routledge

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). University of Hawai'i Press

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

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010, Singapore, December 2-4* (pp. 721–737). National University of Singapore, Centre for Language Studies

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88. https://doi-org/10.1177/026553220101800103

Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly*, *24*(1), 114. https://doi.org/10.2307/748013

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*(1), 121–159. https://doi.org/10.1111/j.1467-9922.2012.00730.x

Spit, S., Andringa, S., Rispens, J., & Aboh, E. O. (2021). Do kindergarteners develop awareness of the statistical regularities they acquire? *Language Learning*, *71*(2), 573–611. https://doi.org/10.1111/lang.12445

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*(4), 747–790. https://doi.org/10.1111/lang.12241

Tekmen, E. A. F., & Daloğlu, A. (2006). An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. *Foreign Language Annals*, *39*(2), 220–243. https://doi.org/10.1111/j.1944-9720.2006.tb02263.x

Teng, F. (2018). Incidental vocabulary acquisition from reading-only and reading-while-listening: A multi-dimensional approach. *Innovation in Language Learning and Teaching*, *12*(3), 274–288. https://doi.org/10.1080/17501229.2016.1203328

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*(2), 183–203. https://doi.org/10.1017/S0272263100012870

Toomer, M., & Elgort, I. (2019). The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt (2013). *Language Learning*, *69*(2), 405–439. https://doi.org/10.1111/lang.12335

Tuzcu, A. (2023). The effects of input mode in learning declarative and nondeclarative knowledge of L2 collocations. *System*, *113*, 103006. https://doi.org/10.1016/j.system.2023.103006

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, *69*(3), 559–599. https://doi.org/10.1111/lang.12343

Van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*(3), 609–624. https://doi.org/10.1016/j.system.2013.07.012

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*(1), 219–258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

Vu, D. V., & Peters, E. (2021). Incidental learning of collocations from meaningful input: A longitudinal study into three reading modes and factors that affect learning. *Studies in Second Language Acquisition*, 1–23. https://doi.org/10.1017/S0272263121000462

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46–65. https://doi.org/10.1093/applin/aml048

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*(2), 232–245

Webb, S., & Chang, A. C.-S. (2015a). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, *37*(4), 651–675. https://doi.org/10.1017/S0272263114000606

Webb, S., & Chang, A. C.-S. (2015b). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*(6), 667–686. https://doi.org/10.1177/1362168814559800

Webb, S., & Chang, A. C.-S. (2020). How does mode of input affect the incidental learning of collocations? *Studies in Second Language Acquisition*, 1–22. https://doi.org/10.1017/S0272263120000297

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*(02). https://doi.org/10.1017/S0272263105050138

Willse, J., T. (2018). *Classical test theory functions. R package version 2.3.3.* https://CRAN.R-project.org/package=CTT

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness . *Canadian Modern Language Review*, *57*, 541–572

APPENDIX A: PARTICIPANTS' LANGUAGE BACKGROUNDS

Table 31

The Distribution of Participants' Language Backgrounds

Reading Only		Reading while listening	
Language	Number	Language	Number
Bangla	1	Chinese	2
Bengali	1	Filipino	1
Chinese	2	French	1
Greek	1	Greek	1
Hindi	3	Gujarati	1
Kannada	1	Hindi	4
Kazakh	4	Kazakh	4
Kyrgyz	1	Korean	1
Marathi	1	Malayalam	1
Marwadi	1	Odia	2
Spanish	6	Persian	1
Thai	1	Portuguese	1
Turkish	6	Russian	1
Urdu	1	Sinhalese	1
Uzbek	1	Spanish	6
		Turkish	3
		Vietnamese	1

APPENDIX B: THE READING TEXT

------- shows the end of each screen in the eye-tracking experiment.

Chapter 1 The Time Traveler

The Time Traveler (it will be convenient to call him this) was talking to us about geometry. His grey eyes shone and his usually pale **mittle** was red and excited. The fire burned brightly and there was that relaxed after-dinner feeling when different notions run freely.

'You must listen carefully. I shall have to destroy one or two ideas that almost everyone accepts — for example, the geometry that they taught you at school. Of course, you know that a mathematical line, a line with no thickness, doesn't really exist. They taught you that? A mathematical model, which only has length, width and thickness, doesn't really exist either. It's just an idea.'

'That's all right,' said the Psychologist,

'But if you make that model out of a material,' said Filly, red-haired man who liked an argument, 'it exists. All real things exist.'

'Most people **prome** that this is true. But wait a moment. Imagine a thing that doesn't last for any time. Can it have a real existence?' Filly looked pensive. 'Obviously,' the Time Traveler said, 'a real body must have length, width and thickness (dimensions of space) —and also exist in time. But we usually forget the fourth of these.'

'That,' said a very young man, 'is very true indeed, very right.'

'Well, I can tell you that I have been at work on this geometry of four dimensions for some time. Some of my results are interesting. Here is a record of the weather. This line shows the changes in temperature. Yesterday it was quite high, last night it fell, then this morning it rose again. Surely that line is not in any of the dimensions of space that we generally experience**.** It is along the time-dimension.'

'But,' said the Medical Man, looking hard at the fire, 'if time is really only a fourth dimension of space, why can't we move about in it as we move in the other dimensions?'

The Time Traveler smiled. 'Are you so sure we can move in space freely? We can go right and left, backwards and forwards freely enough. But up and down? That isn't so easy.'

'Well, we can move a little up and down,' said the Medical Man. 'But we can't move at all in time. We are always in the present moment.'

'That is at the center of my **croom** discovery. Why can a modern man not hope that one day he might travel in time?'

'It doesn't make sense,' said Filby.

'Possibly not' said the Time Traveler. 'But now you begin to see the reason for my work on the geometry of four dimensions. Long ago I had an idea for a machine that can travel in any

direction of space and time, as drivers want.'

Filby started to laugh.

'But I have proved this by experiment,' said the Time Traveler.

'It would be very useful for the historian,' the Psychologist suggested. 'He could travel back and see how things really happened!'

'Then there is the future,' said the Very Young Man. 'Just **prome** about it! You could put all your money in the bank, leave it to grow and hurry on ahead!'

'To discover a society,' I said, 'that doesn't use money.'

'Of all the crazy ideas!' began the Psychologist.

'It seemed so to me, and I never talked about it until—'

'An experiment!' I cried. 'You are going to prove that?'

'Let's see what you can do,' said the Psychologist, 'though I **prome** it's all rubbish.'

The Time Traveler smiled at us. Then, with his hands deep inside his trouser pockets, he walked slowly out of the room and we heard him going down to the laboratory.

The Psychologist looked at us. 'I wonder what he's got?'

'A trick probably,' said the Medical Man, and Filby tried to tell us about a trick he had seen once, but before he had really started his **plear** the Time Traveler came back.

He held something in his hand. It was made of shiny metal and was not much larger than a **merky** clock. It seemed interesting and familiar, but I couldn't **tickel** what it was. It was something that I haven't seen before. And now I must be exact, because unless you believe his words it is impossible to grasp what happened next.

He took one of the tables in the room and put it in front of the fire. On this he placed the machine. Then he placed a **hilder** next to it and sat down. The only other object on the table was a **merky** lamp, the light of which fell on the model.

I sat in a low **hilder** nearest the fire and I pulled this forwards so I was almost between the Time Traveler and the fire. Filby sat behind him, looking over his shoulder.

--- The Medical Man watched him from the right; the Psychologist from the left. The Very Young Man stood behind the Psychologist. We were all wide awake. I cannot believe that a trick was played on us under these conditions.

'This little thing,' said the Time Traveler, resting his elbows on the table and pressing his hands together above the machine, 'is only a model. It is my plan for a machine to travel through time. You will notice that it looks a little rough, and this bar has an odd shining appearance — it looks quite unreal.'

--- He pointed to this part with his finger. 'Also, here is one little white lever, and here is another. These levers are very **fanish** for this machine; without them the machine cannot work.'

The Medical Man got out of his **hilder** and looked closely into the thing. 'It's beautifully made,' he said.

‘It took two years to make,’ said the Time Traveler. Then, when we had all had a close look, he said, ‘Now I want you to **crimb** that this lever sends the machine flying into the future, and this other one sends it into the past. Soon, I’m going to press the lever and the machine will disappear into future time. Have a good look at the thing.’

`Look at the table too, and satisfy yourself that there can be no tricks. I don't want to waste this model and then be told I'm dishonest.’

There was a minute's pause perhaps. The Psychologist opened his mouth to **maive** to me but closed it again. Then the Time Traveler put out his finger towards the lever.

`No.’ he said suddenly, pulling his finger away again. ‘Lend me your hand.’ And turning to the Psychologist, he took that person's hand in his own and told him to put out his first finger and touch the lever.

The Psychologist did what he was told. So the Psychologist himself sent the model time machine on its endless **demuse** in front of us. We all saw the lever turn. I am completely certain there was no trick. There was a breath of wind and the lamp flame jumped. The machine suddenly turned round, was seen like a ghost for a second and was gone — disappeared! Except for the lamp, the table was empty.

Everyone was silent for a minute. Then the Psychologist recovered from his surprise and looked under the table.

The Time Traveler laughed cheerfully. ‘Well?’ he said.

We all stared.

‘My friend,’ said the Medical Man quietly, ‘are you serious about this? Do you really believe that machine has travelled in time?’

‘Certainly,’ said the Time Traveler. ‘And I have a big machine nearly finished in there’ — he pointed to the laboratory — ‘and when that is put together I intend to go on a **demuse** myself.’

‘You mean to say that that machine has travelled into the future?’ said Filby.

‘Into the future or the past — I'm not completely sure which.’

After some time the Psychologist said, ‘It has gone into the past if it has gone anywhere.’

‘Why?’ said the Time Traveler.

‘Because it hasn't moved in space for sure**,** and if it travelled into the future it would still be here all this time. It would have to travel through the time that is passing as we stand here.’

‘But,’ I said, ‘if it travelled into the past, why wasn't it here when we first came into this room, and last Thursday when we were here — and the Thursday before that?’

‘Let's be fair — these are serious questions,’ said Filby, turning towards the Time Traveler.

‘I can **halker** that very easily,’ the Time Traveler said to the Psychologist. ‘It’s there but but we just cannot see it with our eyes.’

‘Of course, you are certainly right’ said the Psychologist. ‘That's simple enough. Why didn't I **prome** of it? We can't see it, in the same way that we can't see a bullet flying through the air.

If it is travelling through time fifty times or a hundred times faster than we are, we can see only one-fiftieth or one-hundredth of it.’

We sat and stared at the empty table for a minute or two. Then the Time Traveler asked us about our opinions.

‘It sounds believable enough tonight,’ said the Medical Man, ‘but it will seem different in the morning.’

‘Would you like to see the Time Machine itself?’ asked the Time Traveler. And then, taking the lamp in his hand, he led the way to the laboratory.

I remember with clarity how we all followed him, and how in the laboratory we saw a larger copy of the little machine. It was almost complete, but two bars lay unfinished on the table and I picked one up for a better look.

‘Now listen,’ said the Medical Man, ‘are you really serious?’

‘In that machine,’ said the Time Traveler, holding the lamp high, ‘I intend to travel in time. Is that **tartle** now? I was never more serious in my life.’

None of us knew what to say. I looked at Filby over the shoulder of the Medical Man and he smiled at me.

Chapter 2 The Traveler Returns

I **prome** at that time none of us really believed in the Time Machine. The fact is, the Time Traveler was one of those men who are too clever to be believed. You never felt that you knew everything about him. You always felt that something was hidden, that he was playing a trick on you. If Filby showed us the model and described things in the Time Traveler’s words, we would believe him more easily. We would easily **crimb** his reasons - because anyone could **crimb** Filby. But the Time Traveler had a strong imagination and we didn't really believe him.

The next Thursday I went to Richmond again and, arriving late, found four or five men already in the sitting room.

------The Medical Man was standing in front of the fire with a sheet of paper in one hand and his watch in the other. I looked around for the Time Traveler.

‘I cannot believe it's half-past seven already,’ said the Medical Man. ‘I suppose we'd better start to **sorge** now?’

‘Where's our host?’ I asked.

‘You have just come? It's rather odd. He has been delayed. He asks me in this note to start dinner at seven if he's not back. He says he will **halker** when he comes.’

‘It seems a pity to let the dinner spoil,’ said the editor of a popular daily paper.

Everybody was hungry, so the Medical Man rang the bell.

Only the Psychologist, the Medical Man and myself had attended the first dinner. The other men were the Editor, a journalist and another - a quiet, shy man with a beard and oversized glasses - who I didn't know from previous meetings. This was a very interesting **merse** of people.

There was some discussion at the dinner table about the Time Traveler's absence and I suggested time travelling, in a half-joking way. After hearing this, the Editor asked us to **halker** that to him and the Psychologist gave a very dull description of the 'clever trick' we had seen a week before.

He was in the middle of this when the door opened slowly and without noise. I was facing it and saw him first. 'Well!' I said. 'At last!'

The door opened wider and the Time Traveler stood in front of us. I gave a cry of surprise.

'Oh, my friend! What's the matter?' cried the Medical Man, who saw him next.

The others turned towards the door.

He looked very strange. His coat was dusty and dirty, his hair untidy and, it seemed to me, greyer - either with dust or because its color had gone.

His **mittle** was very pale and his chin had a cut on it. For a moment he stopped at the door; the light seemed too strong for his eyes. Then he came into the room. He walked slowly, with a bad limp. He did not say a word, but came painfully to the table and moved a hand towards the wine.

The Editor filled a glass and pushed it towards him. He drank it and it seemed to do him good because he looked round the table and smiled a little.

'What have you been doing?' said the Medical Man.

The Time Traveler did not seem to hear. 'Don't let me worry you,' he said in a tired voice. 'I'm all right.' He stopped, held out his glass for more, and drank it down. 'This wine is very good,' he said. His eyes grew brighter, and a faint color came to his **mittle** slowly. Then he spoke again. 'I'm going to **slair** my hands and change my clothes. Then I'll come down and **halker** things . . . Save me some of that meat. I'm hungry.' The Editor began a question.

'I'll tell you soon,' said the Time Traveler. 'I'll be all right in a minute.'

He put down his glass and walked towards the door to the stairs. Standing up in my place, I saw his feet as he went out. He had nothing on them except a pair of socks with holes in them. They were covered with dried blood. Then the door closed behind him. For a minute, perhaps, my **spoll** was empty. But this did not last long.

'Strange Behavior of a Famous Scientist,' I heard the Editor say.

'What's happened to him?' said the Journalist. 'I don't **crimb** anything.' I imagined the Time Traveler walking painfully upstairs. I don't **prome** anyone else had noticed his limp.

The Medical Man recovered from his surprise first, and rang the bell for a hot plate.

------ The Editor picked up his knife and fork and the Silent Man did the same. The dinner started

again. Conversation was slow for a minute or two because we were so surprised. Then the Editor said, ‘Does our friend have another job, or just a strong imagination?’

‘I feel sure it’s this business of the Time Machine,’ I said, and continued the Psychologist's **plear** of our earlier meeting. The **croom** guests were very surprised and the Editor said, ‘What is this time travelling? A man couldn't cover himself with dust by doing something impossible, could he?’

The Journalist, too, refused to believe it, and started to make a joke of the whole thing. ‘Our Special Reporter in the Day after Tomorrow reports,’ he was saying — or shouting — when the Time Traveler came back. He was dressed in ordinary evening clothes and nothing except his tired look reminded me of the change that had shocked me.

‘Well,’ said the Editor, laughing, ‘these men say you have been travelling into the middle of next week.’

The Time Traveler sat down without a word. He smiled quietly, in his usual way, ‘Where's my meat?’ he said. ‘How nice it is to stick a fork into meat again.’

‘Tell us your **plear** now!’ cried the Editor.

‘Later,’ said the Time Traveler. ‘I want something to **sorge** first. I won't say a word until I get some food into my stomach. Thanks. And the salt.’

‘One word,’ I said. ‘Have you been time travelling?’

‘Yes,’ said the Time Traveler, with his mouth full.

‘I'd give a pound a line for the **plear** in your own words,’ said the Editor. The Time Traveler quickly pushed his glass towards the Silent Man, who was staring at his tired **mittle** fixedly. He jumped a little, then poured him some wine. The rest of the dinner was uncomfortable. The Journalist tried to relax us by telling funny jokes. The Medical Man smoked a cigarette and watched the Time Traveler closely. The Silent Man seemed nervous, and drank a lot of wine.

At last the Time Traveler pushed his plate away and looked round at us. ‘I suppose I must **thrine** for making you wait,’ he said. ‘I was so hungry. I've had a most interesting time.’ He put out his hand for a cigarette. ‘But come into the smoking room. The **plear** is too long to tell over dirty plates.’ And he led the way.

‘You have told these men about the machine already, right?’ he said to me, sitting back in his **hilder**. He was looking at the three **croom** guests.

‘But the thing's just a trick,’ said the Editor.

‘Yes, it must be a trick,’ said the Silent Man.

‘I can't argue tonight. I can tell you the **plear** now, but I can't argue. I will,’ he continued, ‘tell you the **plear** of what has happened to me, if you like, but you mustn't interrupt. Most of it will sound like lies, but it is true — every word of it. I was in the laboratory earlier, and since

then . . . I have lived eight days ... days like no human being ever lived before!
------‘I am very tired, but I won’t sleep until I have told this thing to you. But no interruptions! Is it agreed?’

We all agreed and the Time Traveler began his **plear** as I have written it down. He sat back in his **hilder** at first and spoke slowly. Afterwards he got more excited.

As I write it down I feel the limits of pen and ink, and my own limits. You will read, I expect, with enough attention, but you cannot see the Time Traveler's white, honest **mittle** in the bright circle of the little lamp, or hear his voice. Most of us listeners were in shadow.

------At first each looked at the others. After a time, we stopped doing that and looked only at the Time Traveler’s **mittle** with interest. He looked back at us for a few minutes, as if he wanted to increase our curiosity, and started to tell us his adventure.

Chapter 3 Forwards in Time

‘I told some of you last Thursday how the Time Machine works, and showed you the actual thing itself, incomplete in the laboratory. It is there now, a little damaged by travel, but not in bad condition. I expected to finish it on Friday, but when I had put most of it together, I found that one piece was too short. I had to make this again, and the thing wasn't complete until this morning. So, the first of all Time Machines began its **demuse** at ten o'clock today.’

‘I checked everything, and got into the **hilder** afterwards. I felt a little frightened, but interested in what was going to happen next.

------I took the starting lever in one hand and the stopping one in the other. Then I pressed the first and almost immediately the second. I felt that I was falling but, looking around, I saw the laboratory exactly as before. Had anything happened? For a moment I considered that my **spoll** had tricked me. Then I noticed the clock. A moment before, it had showed a minute or two past ten. Now it was nearly half-past three.

‘I took a breath, held the starting lever with both hands and pushed it harder. The laboratory became vague and went dark. Mrs. Watchett, my cook, came in and walked, without seeing me, towards the garden door.

------I suppose it took her a minute or two to cross the room, but she seemed to move at high speed. I pressed the lever over to its furthest position.

The night came, and in another moment came tomorrow. The laboratory grew vague. Tomorrow night became black, then day again, night again, day again — faster and faster. A low and changing sound filled my ears, and my **spoll** became confused.

‘As my speed increased, night followed day faster and faster. The faint picture of the laboratory seemed soon to move away from me. I saw the sun jumping quickly across the sky, once every minute, each minute being a day.

------I supposed that the laboratory had been destroyed and I had come into the open air. The quick changes of darkness and light were very painful to my eyes. Then, in the short dark times, I saw the moon turning quickly through her quarters from **croom** to full.

‘Soon, as I continued, still increasing speed, the change from night to day became one

continuous greyness. The sky turned a wonderful deep blue. The jumping sun became a line of fire, the moon a faint line that changed in width.

'The land was difficult to see properly. I was still on the hillside where this house now stands. I saw trees growing and changing. They changed from **tifled** to brown and back to **tifled** again, grew tall, died and fell. I saw buildings rise up, then disappear like dreams. The speed dials on the machine went round faster and faster. The line of the sun moved up and down, from summer to winter, in a minute or less. Minute by minute white snow disappeared, and was followed by the spring.

'The unpleasant feelings at the beginning now changed into a kind of crazy excitement. I noticed a strange movement of the machine from side to side. I couldn't **halker** what was happening. Also, my **spoll** was too confused to pay any attention to it. So with a kind of madness growing in me, I threw myself into the future.

At first I was excited. I didn't **prome** of stopping. But then a **croom** feeling grew in me slowly— a sense of fear mixed with the need to know what happened.

'What strange changes had happened to people?

------What wonderful improvements to our simple way of life might appear when I looked into that **rolley** more closely? I saw large and wonderful buildings growing in front of me, bigger than ours. I saw a stronger **tifled** color move up the hillside, and stay there without any interruption by snow. Although I was travelling so quickly, the **rolley** still seemed beautiful, and so I wanted to stop the machine.

'My biggest fear was that there would already be something in the space when I, or the machine, stopped. While I travelled at high speed through time, this didn't matter much — I seemed to move like a gas through other things.

------ But when I stopped, I would put myself into whatever lay in my way. Such close contact with the other thing might cause a **vandy** explosion. I had considered this possibility again and again while I was making the machine, but then I had cheerfully accepted it as the necessary **fleak** that a man must confront. I wasn't as cheerful now, when I couldn't escape it.

'The strangeness of everything, the movement of the machine and the feeling of continual falling had made me very nervous. I told myself that I could never stop. Then, becoming suddenly angry, I decided to stop immediately.

------Like a fool in a hurry, I pulled over the lever. The machine turned over and I was thrown through the air.

'There was the sound of thunder in my ears. For a moment I forgot what was happening, then I found myself sitting on soft grass in front of the machine. Heavy rain was falling. Everything still seemed grey, but soon I noticed that the confusion in my ears was gone. I looked around me. I was on a **merky** lawn, surrounded by bushes. Their purple flowers were dropping under the beating of the heavy rain.

------In a moment I was wet to the skin. "A fine welcome," I considered, "to a man who has

travelled so many years to see you."

'Soon I stood up and started looking around me. Although the rain was very heavy, I could still see a **vandy** figure cut, perhaps, out of white stone. But the rest of the **rolley** was vague.

'As the rain became lighter, I saw the white figure better. It was very large — a tree touched its shoulder.

------It was shaped a little like a sphinx with spread wings, and seemed to be flying. The pedestal seemed to be made of metal, and had turned **tifled** with age. I stood looking at the figure for some time.

When, at last, I took my eyes from it for a moment, I saw that the rain was stopping and the sky was growing lighter.

'Then I suddenly realized the **fleak** I was in. What might appear when the rain stopped? What might people be like? Had they perhaps changed into something inhuman and very strong?
------I might seem like an old wild animal, but more frightening because I looked like them — a horrible creature to be speedily killed.

'Already I saw the shapes of buildings, and very slowly, a wooded hillside started to grow **tartle** through the dying storm. I turned quickly to the Time Machine and tried hard to turn it the right way up. As I did so, the grey rain suddenly stopped and the sun shone through the clouds. My fear grew stronger and I fought hard with the machine. It moved under my attack and turned over. It hit my chin violently. One hand on the **hilder** trying to stay still, the other on the lever, I stood breathing heavily, ready to climb inside it again.

'But now I had a way of escaping, my confidence recovered. I looked with more interest and less fear at this **rolley** of the future. In a round opening, high up in the wall of the nearest building, I saw a **merse** of figures wearing soft robes. They had seen me and turned towards me.

'Then I heard voices coming nearer. Through the bushes I saw the heads and shoulders of running men. One of these appeared on a path leading straight to the lawn where I stood.

------He was quite thin, just over a meter high, wearing only a long purple shirt tied at the waist with a leather belt. Noticing that, I realized for the first time how warm the air was.

'He seemed to be very beautiful, but he also had a **fanky** body with very thin and frail arms and legs. At the sight of him, my confidence returned. I took my hands from the machine.

Chapter 4 The People of the Future

'In another moment we were standing together, I and this **fanky** creature from the future. He came straight up to me and laughed into my eyes. I noticed immediately that he had no fear in him. Then he turned to the two others who were following him. He spoke to them in a strange and very sweet-sounding language.

'There were more coming, and soon a little **merse** of perhaps eight or ten of these beautiful people were around me. One of them spoke to me. I don't know why, but I felt that my voice was too strong and deep for them.

------So I shook my head and, pointing to my ears, shook it again. He came a step forwards,

stopped and then touched my hand. Then I felt other soft little hands on my back and shoulders. They wanted to make sure that I was real.

‘There was nothing at all frightening in this. In fact, these pretty little people had a relaxed gentleness that made me confident. And also, they looked so **fanky** to me**.** In fact, I could imagine myself throwing the whole **merse** of them to the ground. I felt relieved after all these thoughts.

‘But I made a sudden movement to warn them when I saw their little pink hands touching the Time Machine. Fortunately, then, when it wasn't too late, I remembered the **fleak** that I had forgotten. Reaching over the bars of the machine, I took out the little levers that would make it move. I put these in my pocket. Then I turned again to the little people to see how I could communicate.

‘Looking closer at them, I saw some strange differences in their sweet prettiness.

------They all had the same wavy hair, and this came to a sharp end at the neck and below the ears. They did not have any beard, and their ears were as **merky** as children’s ears. Their mouths were also **merky** with bright red, rather thin lips. Their little chins came to a point and their eyes were large and gentle.

‘Because they didn't try to **maive** to me, but just stood smiling and looking at each other, I began the conversation. I pointed to the Time Machine and to myself. Then, after contemplating for a moment how to describe time, I pointed to the sun.

------At once a pretty little figure dressed in purple and white did the same, and then made the sound of thunder.

‘For a moment I was very surprised with this behavior, though the meaning of his movement was **tartle** enough. The question had come into my **spoll** suddenly: were these people fools? They were so different from modern people in our times. You couldn't really **crimb** how I felt. I had always expected that people living about 800,000 years in the future would have much more knowledge than us in science, art — everything.

‘But one of them had asked me a very simple question that could be answered easily. This question showed him to be on the level of **mantil** of one of our five-year old children. He had asked me, in fact, if I had come from the sun in a thunderstorm!

‘A feeling of sadness came into my **spoll** suddenly. For a moment I felt that I had built the Time Machine for no reason at all.

‘I said yes, pointed to the sun, and made a sound like thunder. This was so real that it frightened them. They all shook their **fanky** arms and legs and stood back a step or two and bent their heads down. Then one came laughing towards me, carrying some beautiful flowers which were **croom** to me. He put these around my neck.

‘The idea made them all happy. Soon they were running around for flowers and throwing them on me until I was almost covered with them.

------You cannot imagine what wonderful flowers countless years of work had produced.

‘Then someone suggested that their **croom** toy should be shown to others in the nearest building, and so I was led past the sphinx made of white stone, which had seemed to watch me all the time with a smile at my surprise. As I went with them, the memory of my hopes for a future full of highly smart people came to my poor confused **spoll** and made me smile.

‘The building had a very large entrance.

------I was worried about the growing crowd of little people, and the shadows beyond the big open doors. Around me I saw many bushes and flowers. It was **tartle** that no gardener was looking after them, but they still looked beautiful. The Time Machine was left on the lawn.

‘Several more brightly-dressed people met me in the doorway and we walked through into a large hall. The roof was in shadow and the windows, partly made of colored glass, let in a soft light. The floor was made of large pieces of a very hard white metal, lower in places where people had evidently walked across it for hundreds of years.

‘Along the length of the room were many tables made of shiny stone, perhaps half a meter above the floor, and on these were piles of fruit. Some I identified as larger apples and oranges, but mostly they were strange.

‘The people with me sat down around a table and made signs for me to do the same. They immediately began to **sorge** the fruit with their hands. I was happy to follow their example because I felt thirsty and hungry. As I did so, I took some time to look around the hall and noticed that the glass windows were broken in many places and the curtains were thick with dust. The general effect, though, was very attractive.

‘There were, perhaps, a couple of hundred people in the hall, and most of them were watching me with interest, their little eyes shining over the fruit in front of them. All of them were wearing the same soft but strong material on them. It was quite **tartle** that at that moment all they wanted was to **sorge** and not talk.

‘I later learned that fruit was all that they had. These people of the future didn't **sorge** meat. While I was with them, I could only **sorge** fruit too. In fact, I discovered later that horses, cows and sheep, and dogs, had disappeared from Earth. But the fruits were very pleasant.

‘When I had filled my stomach, I tried to learn some words in the weird language of these **croom** people. The fruits seemed an easy thing to start with, and holding one of these up, I began using questioning sounds and movements. I had difficulty making them **crimb** what I was trying to say. At first they stared in surprise and laughed, but soon a fair-haired little female seemed to realize what I wanted and repeated a name.

‘They had to talk for some time to **halker** things to each other, and when I first tried to make the sounds of their language they were very amused.

------I felt like a teacher among children, but soon I at least knew a number of names for things and even how to say **sorge** in their language.

‘It was slow work, though, and the little people soon got tired and wanted to get away from my questions, so I decided to let them give short lessons when they wanted to. And they were very short lessons because I have never met people who are lazier or more easily tired. They used to come to me with happy cries of surprise, like children, but like children they soon stopped examining me and went away to find another toy.

‘When the dinner ended, I noted the disappearance of almost all the creatures who had surrounded me at first. It is odd, too, how quickly I stopped caring about these little people. I was continually meeting more of them. They followed me a little distance, talked and laughed around me, smiled in a friendly way, then left me alone.

Chapter 5 Life in the Future

‘The evening was calm as I came out of the **vandy** hall, and the land was lit by the color of the sun as it went down. The big building was on the side of a wide river valley, but the Thames had moved a kilometer or two from its present position. I decided to climb to the top of a hill from where I could see more of our **rolley** in the year 802,701. That was the date the little dials of my machine had showed.

‘I started walking towards the hill. As I walked slowly, I looked for anything that could **halker** the bad condition of things. A little way up the hill, for example, was a **vandy** pile of stones held together by pieces of metal. These were the ruins of a building, although I couldn't imagine what its use had been.

‘Looking round, I realized that there were no **merky** houses. Here and there among the trees and bushes were palace-like buildings, but the single house, and possibly even the family, had disappeared.

‘And then came another notion. I looked at the **merky** figures who were following me. I saw that all had the same type of clothes, the same wavy hair and the same girlish arms and legs.

‘It may seem odd, perhaps, that I hadn't noticed this before. But everything was so strange. Now, I saw the fact with lucidity. These people of the future were all very similar in clothes, and in all other ways the differences between men and women had almost disappeared. And the children seemed to my eyes to be just tiny adults.

‘Seeing how safely and comfortably these people lived, I felt that this close similarity of the sexes was unsurprising. If there are enough people, it becomes a problem rather than an advantage to have a lot of children. If violence comes only rarely and children are safe, there is less need for men to be strong and protect their families.

--------This, I must remind you, was my feeling at the time. Later, I discovered how wrong I was.

‘I continued, and because I could walk better than the people of the future, I found myself alone for the first time. At the top of the hill I found a **hilder** of a yellow metal. I didn't **tickel** it

from before. I sat down on it and looked at the wide view of our **rolley** under the sunset of that long day. It was as beautiful as I have ever seen. The west was burning gold, mixed with some purple and red.

--------Below was the valley of the Thames, in which the river lay like a line of shining metal.

'As I watched, I began to try to **crimb** the things I had seen. (Afterwards I realized I had only learned half the truth.) It seemed to me that people were now past their best. The sunset made me **prome** about the sunset of our people. For the first time I began to **crimb** an odd result of the social changes we are trying to make at the moment. Strength comes because we need to be strong; vulnerability comes when we feel safe. The work of improving the conditions of life, of making life safer and safer, had continued until nothing more could be done. The result was what I saw!

'The science of our time has attacked only a few human diseases, but it moves forwards. Farming today is still at an early stage. We improve our plants and animals very slowly — a **croom** and better apple, a prettier and larger flower, a cow that gives more milk. One day the whole **rolley** will be better organized, and better.

'I knew that this change had been made, and made well, in the space of time across which my machine had jumped. The air was free of unpleasant insects; the earth was free of useless plants. Everywhere there were fruits and sweet and pleasant flowers. Beautiful birds flew here and there. And I saw no diseases during my stay.

'Social changes, too, had been made. I saw people living in fine buildings, beautifully dressed, but I hadn't yet found them doing any work. There were no signs of economic activity. The shop, the advertisement, buying and selling — all of these things are so **fanish** to us, and all of them were gone. It was natural in the evening that I had the idea of a social heaven.

'But this change in conditions has to produce changes in people. What is the cause of human **mantil** and energy? Difficulties make people strong and clever and help them to work together.

--------And the family, with its protective love and selfishness, is there for the care of children. The love of parents helps to keep the young out of **fleak** all the time. Now, there was neither **fleak** nor need for protection.

'I recalled the shortness of the people, their low **mantil** and those big ruined buildings. It strengthened my belief that humans, who had always fought against nature, had finally won — because after the fight comes quietness.

--------People had been strong, energetic and smart in the past, and had used this energy to change their living conditions. And now they too had changed because of the **croom** conditions.

'No doubt the beauty of the buildings was the result of the last waves of the now purposeless energy of people. After that, they began to lead quieter lives. Even artistic activity would finally disappear — had almost disappeared in the time I saw. The people liked to cover themselves in flowers, to dance and to sing in the sunlight. That was all they did.

‘As I stood there in the growing dark, I believed that I had been aware of the whole secret of these pleasant people. Possibly their population control had worked too well, and their numbers had fallen instead of staying the same. That would easily **halker** the empty ruins. My assumptions were very simple and believable enough — as most wrong ideas are!

Chapter 6 Lost in Time

‘As I stood there considering this too perfect success of humans, the full moon came up in the north-east. The little figures stopped moving around below me and the night began to feel cold. I decided to go down and find a place to sleep.

‘I looked for the building I knew. Then my eye moved to the white sphinx on the pedestal. There were the bushes and there was the **merky** lawn. I looked at it again. A strange doubt made me feel cold. "No," I said to myself, "that isn't the lawn.”

‘But it was the lawn, because the white sphinx was towards it. Can you imagine how I felt as I realized this? But you can't. The Time Machine had gone!

‘At once I realized the possibility of being left helpless in this **rolley** of the future and losing my own time. I ran with **vandy** jumps down the hillside. Once I fell and cut my **mittle** slightly. I did nothing to stop the blood, but jumped up and continued running. All the time I was saying to myself, "They have just pushed it under the bushes out of the way."

‘But I knew that I was wrong. I suppose in only ten minutes I covered the whole distance to the **merky** lawn, three kilometers perhaps. I shouted but nobody answered. Nobody seemed to be moving in that moonlit **rolley** but me.

‘When I reached the lawn, I found that my worst fears were true. The Time Machine was nowhere to be seen. I felt **fanky** and cold. I ran round the lawn quickly, checking every corner, then stopped suddenly. Above me was the white sphinx. It seemed to smile with pleasure at my problems.

‘It is possible that the little people had put the machine in a safe place for me, but I didn't feel that they were either strong enough or caring enough to move it. This is what worried me, the feeling of a **croom** power that had moved the machine. But where could it be?

‘I **prome** I went a little mad. I remember running violently in and out of the moonlit bushes all around the sphinx and frightening a white pure animal which I didn't **tickel** from before. Then, crying and shouting, I went down to the building of stone.

--------The big hall was dark, silent and empty. I lit a match and continued past the dusty curtains.

‘There I found a second **vandy** hall, where about twenty of the little people were sleeping. I have no doubt they found my second appearance strange, as I came suddenly out of the quiet darkness with mad noises and the sudden light of a match. Perhaps they had forgotten about matches, "Where is my Time Machine?" I began, shaking them with my hands.

‘This behavior was very strange to them. Some laughed, but most looked very frightened.

--------When I saw them standing round me, I realized that it was foolish to try and frighten them. judging by their daylight behavior, I believed that fear must be forgotten.

‘I threw down the match and, knocking one of the people over as I went, I ran across the big dining-hall again, out under the moonlight. I heard cries of terror and their little feet running this way and that. I don't remember everything I did as the moon moved slowly tip the sky. I know that I ran here and there screaming, then lay on the ground near the sphinx and cried. After that I slept, and when I woke up again it was light.

‘I sat up in the freshness of the morning, trying to remember how I had got there. Then things became **tartle** slowly. I realized the wild stupidity of my madness overnight and I could reason with myself. "Suppose the worst,” I said. "Suppose the machine is really lost — perhaps destroyed? I should be calm and patient, learn the ways of the people, learn what has happened and how to get materials and tools — then, in the end, perhaps, I can make another machine." That would be my only hope, but better than giving up. And it was a beautiful and interesting **rolley** worth to explore.

‘But probably the machine had only been taken away.

‘I needed to be calm, find its hiding-place and get it back by force and cleverness. I stood up and looked around me, wondering where I could **slair** my body. I felt tired and dirty and rather surprised by my emotional state the night before.

‘I made a careful examination of the ground around the little lawn. I wasted some time in useless questions, asked, as well as I could, to the little people that passed. They all failed to **crimb** what I meant. Some said nothing; others believed it was a joke and laughed at me.

‘The grass told me more. I found a line in it.

There were other signs around, with strange narrow footprints. Thus made me look again at the pedestal. It was made, as I **prome** I have said, of metal. It was highly decorated with metal panels on either side.

‘I went and knocked at these. The pedestal was hollow. There was no way to pull to open the panels, but perhaps if they were doors they opened from inside. One thing was **tartle** enough: it wasn't difficult to work out that the Time Machine was inside that pedestal. But how had it got there?

‘I saw the heads of two people dressed in orange coming through the bushes towards me. They came and, pointing to the pedestal, I tried to make them **crimb** my wish to open it. But at my first move to do this they behaved very oddly. I don't know how to describe their expressions to you. They looked insulted.

‘I tried a sweet-looking man in white next, with exactly the same result. He made me feel ashamed of myself. But as you know already, I wanted the Time Machine and I tried him again.

As he turned away, like the others, I lost my temper.

-------In three steps I was after him, took him by the loose part of his robe round the neck and began pulling him towards the pedestal. I saw the fear on his **mittle** after a few steps, said a few words to **thrine** and let him go.

‘But I wasn't beaten yet. I hit the metal panels with my hands. I felt like I heard something move inside — to be exact, I felt like I heard a sound like a laugh — but perhaps I was mistaken. Then I got a big stone from the river and hit the metal until had flattened part of the decoration. The little people could hear the noise a kilometer way in all directions, but they did nothing.

‘I saw a crowd of them on the hillside, looking at me in a frightened way. At last, hot and tired, I sat down to watch the place. But I was too impatient to watch for long. I could work at a problem for years, but I was unable to wait, inactive, for twenty-four hours.

‘I got up after a time and began walking aimlessly through the bushes towards the hill again. "Patience," I said to myself. "If you want your machine again, you must leave that pedestal alone. If they intend to take your machine away, it won't help if you destroy their metal panels.

-------If they don't, you will get it back when you can ask for it.

"Accept this **rolley** and learn its ways. Watch it and be careful of guessing its meaning too quickly. In the end you will find an answer to it all." Then the humor of the situation came into my **spoll** suddenly: I had spent in study and work to get into the future age, and now I was impatient to get out of it. I had put myself into the most hopeless situation a man could ever imagine. I couldn't help laughing at myself.

Chapter 7 Ghosts

‘Going through the big palace, it seemed to me that the little people were staying away from me. Perhaps it was my imagination, or because I had hit the metal panels. I was careful, though, to show no worry and not try to catch any of them, and after a day or two the situation got back to normal.

‘I decided to not **prome** about my Time Machine and the mystery of the metal doors as much as possible.

-------I hoped that in the end, growing knowledge would lead me back to them in a natural way. But you can **crimb** why I stayed within a circle of a few kilometers around my point of arrival.

‘As far as I could see, all the **rolley** seemed to be like the Thames valley. From every hill I saw the same large numbers of fine buildings, all very different in material and style, and the same kinds of trees and bushes. I soon noticed, though, a number of wells in the ground. Several of these, it seemed to me, were very deep.

-------One lay by a path up the hill, which I had followed during my first walk. Like the others, it had a top made of metal, interestingly decorated and protected by a little roof from the rain.

‘Sitting by the side of these wells, and looking down into the darkness, I could see no sign of

water or any reflection when I lit a match. But in all of them I heard a certain sound like the beating of a big engine. I also discovered, from the flames of my matches, that air was going down into them.

-------I threw a piece of paper down into one and, instead of falling slowly, it was at once pulled quickly out of sight. I couldn't imagine what these wells were for.

‘And I must say now that I learned very little about many parts of the life of these people. Let me describe my difficulties. I went into several big palaces, but they were just living places, **vandy** dining-halls and sleeping apartments. I could find no machines of any kind, but these people were dressed in fine cloth that didn't seem very old, and their shoes, though undecorated, were very well made.

‘But the people didn't seem to make things themselves. There were no shops, no factories, no signs that they brought things in from other places. All they did was to play gently, swim in the river, fall in love, constantly **sorge** fruit and sleep. I couldn't see how or where things were produced.

‘But something had taken the Time Machine into the pedestal. Why? I couldn't imagine.

-------Suppose you found something written in English, with here and there some words that were completely unfamiliar to you. Well, on the third day of my visit, that was how I felt about the **rolley** of 802,701.

‘That day I made a friend — a kind of friend. As I was watching some of the little people playing in a shallow part of the river, one of them was suddenly pulled away by the water. The river there could run quite quickly, but not too quickly for a swimmer of normal ability. None tried to help the one that was in such **fleak** even after seeing her drowning. This should give you an idea, therefore, of these people’s lack of strength.

‘When I realized this, I quickly took off my clothes and, walking into the water at a place lower down, I caught her and brought her safely to land.

‘She soon began to feel better and I saw that she was all right before I left her. I had such a low opinion of her people by then that I didn't expect any thanks from her. I was wrong about that, though.

‘This happened in the morning. In the afternoon I met my woman as I was returning from a long walk, and she greeted me with cries of happiness and gave me some flowers.

Perhaps because I had been very lonely I did my best to show I was happy with the gift. We were soon sitting together and deep in a conversation, mainly of smiles.

‘The woman's friendliness affected me exactly as a child's would. We passed each other flowers and she kissed my hands. I did the same to hers. Then I tried to talk, and found that her name was Weena. That was the beginning of a strange friendship, which continued for a week and ended ... as I will tell you!

‘She was exactly like a child. She wanted to be with me always. She wanted to follow me everywhere, and on my next **demuse** around the area I walked fast and tried to leave her behind. She gave up at last, calling after me rather sadly. But the problems of the **rolley** had to be solved and I hadn't, I said to myself, come into the future to start a relationship.

‘She was, though, a very **vandy** comfort. When it was too late, only when it was too late, I realized how badly she felt when I left her, and what she meant to me.

-------By seeming fond of me, and by showing in her weird way that she cared for me, the little person soon gave my returns to the place of the white sphinx almost the feeling of coming home. I used to watch for her when I came over the hill.

‘From her, too, I learned that fear had not yet left the **rolley** unlike my assumptions. She was fearless enough in the daylight, but she hated the dark shadows.

-------To her, darkness was the one thing to be frightened of. It was a very strong emotion, a different one from what most men in our times experience, and I started to **prome** and watch.

‘Among many things, I discovered that these people got together in the **vandy** houses after dark and slept together. I never found one outside, alone. And if I entered the room without a light, I made them very afraid. But I was such a fool that I missed the lesson of that fear, and although it made Weena unhappy I slept away from the others.

‘It worried her a lot, but in the end her feelings for me won. For five of the nights of our friendship, including the last night of all, she slept with her head on my arm. But my **plear** is running away from me as I **maive** of her.

‘On the night before I met her I woke up early in the morning. I had slept badly, dreaming that I was under water, and that fish were touching my **mittle** viciously. I woke suddenly and with the odd feeling that a greyish animal had just rushed out of the room.

‘I tried to go to sleep again, but I felt uncomfortable. It was that grey hour when things are just appearing from the darkness, but are still unreal. I got up, went down into the **vandy** hall and out onto the stones in front of the palace. I wanted to go to the **tifled** hillside and watch the sun come up.

‘The moon was going down, and the dying moonlight and the first light of day were mixed in a pale half-light. The bushes were inky black, the ground a dark grey, and up on the hillside I believed I could see ghosts.

-------Three times I saw white figures and twice I believed I saw a single white animal running quickly on two legs.

‘Near the ruins I saw a **merse** of them carrying a dark body. They moved quickly and it seemed that they disappeared among the bushes. You must **crimb** that the light was still faint. I was experiencing that cold, uncertain, early-morning feeling you may be familiar with, and I doubted my eyes.

‘As the eastern sky grew brighter, and the light of day brought stronger colors to the **rolley**

again, I watched the hillside closely. But I saw no more white figures. I speculated about them all morning — or at least until I had to get Weena out of the river. I connected them in some way with the white animal I had touched in my first mad search for the Time Machine. It was more pleasant to **prome** about Weena, but these ghosts took much stronger control of my **spoll** soon enough. Who were they and what they were doing there?

APPENDIX C: PREDICTABILITY OF PSEUDOWORDS FROM THEIR FIRST OCCURRENCE IN THE TEXT

The percentage of correct answers to each item was calculated by dividing the number of correct answers by the number of total answers. All semantically, syntactically, and contextually appropriate words were regarded as correct answers. For example, for the word fanish, "important", "vital", and "critical" were accepted as correct answers.

Participants also rated the ease of guessing on a 10-point scale (1 = very easy to guess, 10 = very easy difficult).

Table 32

The Mean Predictability Scores of the Pseudowords from Their First Occurrence in the Test

Pseudoword	Meaning	Percentage (mean)	Rating (mean)
crimb (n = 10)	understand	70	4.29
croom (n = 10)	new	70	3.85
demuse (n = 3)	journey	66	4
fanish (n = 10)	important	100	3
fanky (n = 10)	weak	55	6.4
fleak (n = 10)	danger	55	3.2
halker (n = 3)	explain	33	6
hilder (n = 10)	seat	70	3.86
maive (n = 3)	speak	66	4.5
mantil (n = 10)	intelligence	33	2.66
merky (n = 10)	small	50	5.8
merse (n = 10)	group	100	2.66
mittle (n = 10)	face	100	2
plear (n = 10)	story	50	3.4
prome (n = 10)	think	80	2.25
rolley (n = 3)	world	100	5.33
slair (n = 10)	wash	100	1.66
sorge (n = 10)	eat	44	3.75
spoll (n = 10)	mind	44	4.75

Table 32 (cont'd)

Pseudoword	Meaning	Percentage (mean)	Rating (mean)
tartle (n = 3)	clear	33	7
thrine (n = 10)	apologize	100	2.22
tickel (n = 10)	recognize	90	2.11
tifled (n = 10)	green	88	3.88
vandy (n = 3)	big	100	4.66

APPENDIX D: COMPREHENSION TEST

Instructions in the Experiment

You have reached the end of this chapter. Now, you will see 4 true/false questions about the chapter you read. Please read each sentence and decide whether the information in the sentence is true or false.

After you have decided your answer, please press the Left or Right Arrow Key that corresponds with your choice.

Press "SPACE" to go to the next screen.

Questions

1	The Time Traveler sends the machine to the future. (Ch 1)	**True**	False	I don't know.
2	The time machine has three levers. (Ch 1)	True	**False**	I don't know.
3	Most people usually forget the dimension of length. (Ch 1)	True	**False**	I don't know.
4	The Psychologist thinks that a time machine would be useful for mathematicians. (Ch 1)	True	**False**	I don't know.
5	The Medical Man attended both dinners in the Time Traveler's house. (Ch 2)	**True**	False	I don't know.
6	The Time Traveler was wearing no pants when he arrived his home. (Ch 2)	True	**False**	I don't know.
7	The guests had started having dinner before the Time Traveler arrived. (Ch 2)	**True**	False	I don't know.
8	The Time Traveler was gone for two weeks. (Ch 2)	True	**False**	I don't know.
9	The Time Traveler first knew that his machine was working when he looked at the clock. (Ch 3)	**True**	False	I don't know.
10	The Time Traveler's greatest fear was that his machine might have stopped working. (Ch 3)	True	**False**	I don't know.
11	When the Time Traveler stopped the time machine, the weather was sunny. (Ch 3)	True	**False**	I don't know.
12	The Time Traveler felt confident after seeing the short man. (Ch 3)	**True**	False	I don't know.
13	The people in the future were very hard working. (Ch 4)	True	**False**	I don't know.

14	The Time Traveler put the levers in his pocket before leaving his time machine. (Ch 4)	**True**	False	I don't know.
15	In the future, the dogs had disappeared from Earth. (Ch 4)	**True**	False	I don't know.
16	The Time Traveler felt sad when he met the people of the future. (Ch 4)	**True**	False	I don't know.
17	The Time Traveler climbed a hill and saw ruins of a building. (Ch 5)	**True**	False	I don't know.
18	People of the future were living in small houses with their families. (Ch 5)	True	**False**	I don't know.
19	There were no unpleasant insects or plants in the future. (Ch 5)	**True**	False	I don't know.
20	People of the future were very smart. (Ch 5)	True	**False**	I don't know.
21	When the Time Traveler came down from the hill, his time machine was gone. (Ch 6)	**True**	False	I don't know.
22	People of the future did not like going out at night. (Ch 6)	**True**	False	I don't know.
23	The lines on the grass showed that the time machine was taken to the underground. (Ch 6)	True	**False**	I don't know.
24	The Time Traveler managed to learn who took his time machine. (Ch 6)	True	**False**	I don't know.
25	When the little woman was drowning, all people around tried to help her. (Ch 7)	True	**False**	I don't know.
26	The Time Traveler was thinking about his friends all the time. (Ch 7)	True	**False**	I don't know.
27	The little woman, Weena, and the Time Traveler became friends. (Ch 7)	**True**	False	I don't know.
28	The Time Traveler thought that he saw ghosts carrying a body. (Ch 7)	**True**	False	I don't know.

APPENDIX E: VOCABULARY TESTS

E.1. Sentence Reading Test

Instructions in the Experiment (1)

You will read 52 English sentences. After you have read each sentence, press "SPACE" to see the next sentence.
Some sentences are followed by a simple true/false question. The true/false question is always about the last sentence you have read. So read all sentences carefully.
After you have decided your answer, please press the Left or Right arrow key that corresponds with your choice.
Press "SPACE" to continue.

Instructions in the Experiment (2)

A drift check happens before every sentence. So, look at the center of the dot on the left side for the sentence to appear.
The first 4 sentences will be practice sentences and will not count towards your score.
Press "SPACE" to continue.

Stimuli Lists in the Sentence Reading Test

List 1

ID	Sentence	Question	Answer	Item Type
1	Kara saw a hilder next to the window.	The event takes place near a window.	Left	pseudoword
2	They always value respect more than anything.	The event described is about opinions.	Right	word
3	Jason listened to Mary's snash before leaving the building.	The event takes place in a building.	Left	nonword
4	They were aware of the fleak around them.	The event includes more than one person.	Right	pseudoword
5	They want to save the region they live in.	The action described is about saving.	Left	word
6	How the human tremp works is still unknown.	The information described is known by many people.	Right	nonword
7	Alicia has a normal mittle and very long arms.	The person described has long arms.	Left	pseudoword
8	There was a big train in front of the school.	The event takes place in front of a hospital.	Right	word
9	Dylan has waited for this fundle for a long time.	The person's name is Dylan.	Left	nonword

10	Adam started to sorge after they arrived.	The event described takes place before people arrive.	Right	pseudoword
11	I would like to join but I do not have time.	The description is about a person who has no time.	Left	word
12	Britney and Rob always sloker to each other.	The action described includes three people.	Right	nonword
13	He tried to prome while walking in the park.	The action described takes place in a park.	Left	pseudoword
14	Julia needs to realize her responsibilities as a student.	The action described is about being a doctor.	Right	word
15	I need someone to grench my car for me.	The event described is about a car.	Left	nonword
16	Chloe could not tickel the house after a month.	The event described is about a library.	Right	pseudoword
17	Julia could not sleep for months after the accident.	The event described happens after an accident.	Left	word
18	Many animals live in toyal lands in North America.	The event takes place in Western Europe.	Right	nonword
19	Jenna is a very fanish person in Germany.	The event takes place in Germany.	Left	pseudoword
20	I do not think blue paint is suitable for this place.	The event described is about an animal.	Right	word
21	Brian was carrying a tabble ball in his hands.	The event described is about a ball.	Left	nonword
22	Carlo was very fanky after the event.	The event described is about two people.	Right	pseudoword
23	The singer had a sweet voice and good taste in music.	The description is about a singer.	Left	word
24	The room was full of mooky tools and colorful books.	The description is about a class.	Right	nonword

List 2

ID	Sentence	Question	Answer	Item Type
1	Kara saw a plant next to the window.	The event takes place near a window.	Left	word
2	They always value coster more than anything.	The event described is about opinions.	Right	nonword
3	Jason listened to Mary's plear before leaving the building.	The event takes place in a building.	Left	pseudoword
4	They were aware of the animals around them.	The event includes more than one person.	Right	word
5	They want to save the farcel they live in.	The action described is about saving.	Left	nonword

6	How the human spoll works is still unknown.	The information described is known by many people.	Right	pseudoword
7	Alicia has a normal mouth and very long arms.	The person described has long arms.	Left	word
8	There was a big ladin in front of the school.	The event takes place in front of a hospital.	Right	nonword
9	Dylan has waited for this demuse for a long time.	The person's name is Dylan.	Left	pseudoword
10	Adam started to dance after they arrived.	The event described takes place before people arrive.	Right	word
11	I would like to lobble but I don’t have time.	The description is about a person who has no time.	Left	nonword
12	Britney and Rob always thrine to each other.	The action described includes three people.	Right	pseudoword
13	He tried to relax while walking in the park.	The action described takes place in a park.	Left	word
14	Julia needs to luner her responsibilities as a student.	The action described is about being a doctor.	Right	nonword
15	I need someone to slair my car for me.	The event described is about a car.	Left	pseudoword
16	Chloe could not finish the house after a month.	The event described is about a library.	Right	word
17	Julia could not nince for months after the accident.	The event described happens after an accident.	Left	nonword
18	Many animals live in vandy lands in North America.	The event takes place in Western Europe.	Right	pseudoword
19	Jenna is a very famous person in Germany.	The event takes place in Germany.	Left	word
20	I do not think borale paint is suitable for this place.	The event described is about an animal.	Right	nonword
21	Brian was carrying a merky ball in his hands.	The event described is about a ball.	Left	pseudoword
22	Carlo was very angry after the event.	The event described is about two people.	Right	word
23	The singer had a pellow voice and good taste in music.	The description is about a singer.	Left	nonword
24	The room was full of croom tools and colorful books.	The description is about a class.	Right	pseudoword

List 3

ID	Sentence	Question	Answer	Item Type
1	Kara saw a biddle next to the window.	The event takes place near a window.	Left	nonword
2	They always value mantil more than anything.	The event described is about opinions.	Right	pseudoword
3	Jason listened to Mary's secret before leaving the building.	The event takes place in a building.	Left	word
4	They were aware of the lurge around them.	The event includes more than one person.	Right	nonword
5	They want to save the rolley they live in.	The action described is about saving.	Left	pseudoword
6	How the human brain works is still unknown.	The information described is known by many people.	Right	word
7	Alicia has a normal slamp and very long arms.	The person described has long arms.	Left	nonword
8	There was a big merse in front of the school.	The event takes place in front of a hospital.	Right	pseudoword
9	Dylan has waited for this present for a long time.	The person's name is Dylan.	Left	word
10	Adam started to draze after they arrived.	The event described takes place before people arrive.	Right	nonword
11	I would like to halker but I don’t have time.	The description is about a person who has no time.	Left	pseudoword
12	Britney and Rob always smile to each other.	The action described includes three people.	Right	word
13	He tried to brune while walking in the park.	The action described takes place in a park.	Left	nonword
14	Julia needs to crimb her responsibilities as a student.	The action described is about being a doctor.	Right	pseudoword
15	I need someone to clean my car for me.	The event described is about a car.	Left	word
16	Chloe could not sumper the house after a month.	The event described is about a library.	Right	nonword
17	Julia could not maive for months after the accident	The event described happens after an accident.	Left	pseudoword
18	Many animals live in wild lands in North America.	The event takes place in Western Europe.	Right	word
19	Jenna is a very rimped person in Germany.	The event takes place in Germany.	Left	nonword
20	I do not think tifled paint is suitable for this place.	The event described is about an animal.	Right	pseudoword
21	Brian was carrying a brown ball in his hands.	The event described is about a ball.	Left	word

22	Carlo was very bairy after the event.	The event described is about two people.	Right	nonword
23	The singer had a tartle voice and good taste in music.	The description is about a singer.	Left	pseudoword
24	The room was full of weird tools and colorful books.	The description is about a class.	Right	word

Fillers and Practice Sentences

ID	Sentence	Question	Answer	Item Type
25	He got the recipe from his mother last year.	The action described is about a recipe.	Left	filler
26	Mary was joking about going on hiking that day.	The event described is about going swimming.	Right	filler
27	I submitted my assignment before I went out.	The action takes place before going out.	Left	filler
28	Laura went to the store to buy a bracelet.	The event takes place at a library.	Right	filler
29	The woman found some old globes in the store.	The event takes place in a store.	Left	filler
30	This package is too large to put in that drawer.	The description is about a building.	Right	filler
31	The company will go bankrupt in five years.	The event is about a company.	Left	filler
32	The dog learned how to imitate a cat.	The action described is about a lion.	Right	filler
33	Megan wanted a new car for her birthday.	The person's name is Megan.	Left	filler
34	Gale goes running every weekend since his childhood.	The event is about going climbing.	Right	filler
35	Jason shook hands with me when we met.	The person's name is Jason.	Left	filler
36	There is a museum on the south part of the city.	The description is about a person.	Right	filler
37	There are two polar regions on Earth, the Arctic and the Antarctic.	The description is about the Earth.	Left	filler
38	Daniel bought a purple scooter for her daughter.	The person's name is Caleb.	Right	filler
39	I was reading a book when John knocked on the door of my room.	The event takes place in a room.	Left	filler
40	The baby was awake all day yesterday.	The event described is about an adult.	Right	filler
41	The nights were slowly getting	The description is about the	Left	filler

	colder and colder.	nights.		
42	Lisa had a pet rabbit throughout her childhood.	The person's name is Juliet.	Right	filler
43	She needed to collect four more dollars for the fundraiser.	The event described is not complete.	Left	filler
44	The man saw a car in the distance.	The description is about a shool bus.	Right	filler
45	The building was destroyed during the hurricane yesterday.	The event is about a building.	Left	filler
46	Jennifer criticized Alison for being late for the appointment.	The event is about being early for an appointment.	Right	filler
47	Margaret was writing a letter when I entered the room.	The event takes place in a room.	Left	filler
48	The first coffee beans were collected in Eastern Africa.	The event takes place is Western Asia.	Right	filler
49	The news of earthquake in Turkey made everybody sad.	The event takes place is Turkey.	Left	practice
50	The women bought a lamper for herself.	The event is about a man.	Right	practice
51	Eating too much late at night gives me nightmares.	The event is about eating too much.	Left	practice
52	He wanted to go to the restaurant but he had to work.	The event is about going to a school.	Right	practice

E.2. Form Recognition Posttest

Choose the words you saw in the book (The Time Machine). Do your best not to guess!

☐ tremp	☐ luner	☐ brune
☐ thrine	☐ demuse	☐ tartle
☐ hilder	☐ farcel	☐ croom
☐ biddle	☐ halker	☐ coster
☐ snash	☐ ladin	☐ lurge
☐ mantil	☐ lobble	☐ slamp
☐ plear	☐ prome	☐ slair
☐ draze	☐ tickel	☐ grench
☐ fleak	☐ bairy	☐ sorge
☐ rolley	☐ pellow	☐ sumper
☐ fundle	☐ maive	☐ nince
☐ sloker	☐ vandy	☐ tifled
☐ spoll	☐ borale	☐ toyal
☐ tabble	☐ fanish	☐ crimb
☐ mittle	☐ merky	☐ rimped
☐ merse	☐ fanky	☐ mooky

E.3. Meaning Recall Test

Please have a look at the words and write their English definitions or synonyms. Then choose a percentage for your confidence level.

If you do not know the meaning of the word, please choose "I don't know".

	Word	**English definition or synonym**	**I don't know**	**How confident are you with your answer?** 0 - 50% = not very confident 60-70% = somewhat confident 70-80% = confident 80-90% = very confident 100% = absolutely certain
1	hilder		☐	
2	prome		☐	
3	plear		☐	
4	fleak		☐	
5	halker		☐	
6	croom		☐	
7	tickel		☐	
8	merse		☐	
9	tifled		☐	
10	fanky		☐	
11	rolley		☐	
12	thrine		☐	
13	mantil		☐	
14	crimb		☐	
15	slair		☐	
16	mittle		☐	
17	maive		☐	
18	vandy		☐	
19	fanish		☐	

20	demuse		☐	
21	merky		☐	
22	sorge		☐	
23	tartle		☐	
24	spoll		☐	

E.4. Meaning Recognition Test

Look at the following target words and circle the word that has the closest meaning to the target word. Then choose a percentage for your confidence level.

If you do not know the meaning of the word, please choose "I don't know".

	Word	1	2	3	4	I don't know.	How confident are you with your answer? 0 - 50% = not very confident 60-70% = somewhat confident 70-80% = confident 80-90% = very confident 100% = absolutely certain
1	**mantil**	☐ intelligence	☐ material	☐ existence	☐ trick	☐	
2	**tifled**	☐ dusty	☐ green	☐ daily	☐ ordinary	☐	
3	**fanky**	☐ patient	☐ weak	☐ happy	☐ dirty	☐	
4	**demuse**	☐ journey	☐ dream	☐ window	☐ arm	☐	
5	**rolley**	☐ interruption	☐ damage	☐ world	☐ reason	☐	
6	**spoll**	☐ mind	☐ heart	☐ secret	☐ shoulder	☐	
7	**mittle**	☐ face	☐ wallet	☐ laboratory	☐ wine	☐	
8	**merse**	☐ age	☐ person	☐ group	☐ sleep	☐	
9	**sorge**	☐ gossip	☐ talk	☐ dream	☐ eat	☐	
10	**hilder**	☐ finger	☐ seat	☐ door	☐ space	☐	
11	**thrine**	☐ apologize	☐ argue	☐ smoke	☐ drink	☐	
12	**halker**	☐ look	☐ notice	☐ admire	☐ explain	☐	
13	**crimb**	☐ walk	☐ hold	☐ open	☐ understand	☐	
14	**merky**	☐ complex	☐ fast	☐ small	☐ late	☐	
15	**tickel**	☐ recognize	☐ smile	☐ fight	☐ continue	☐	
16	**vandy**	☐ wild	☐ big	☐ deep	☐ scary	☐	
17	**prome**	☐ hear	☐ think	☐ recover	☐ decorate	☐	
18	**fanish**	☐ surprising	☐ close	☐ important	☐ full	☐	
19	**plear**	☐ stage	☐ flower	☐ story	☐ region	☐	
20	**fleak**	☐ danger	☐ confidence	☐ cloud	☐ money	☐	
21	**maive**	☐ speak	☐ complete	☐ determine	☐ remember	☐	
22	**croom**	☐ fair	☐ new	☐ famous	☐ shiny	☐	
23	**tartle**	☐ pretty	☐ linen	☐ clear	☐ strong	☐	
24	**slair**	☐ feel	☐ laugh	☐ imagine	☐ wash	☐	

APPENDIX F: CHARACTERISTICS OF CRITICAL WORDS IN THE SENTENCE READING TEST

Table 33

Characteristics of Real Words in the Sentence Reading Test

Words	Frequency (BNC/COCA)	Length	Orthographic Neighbors	Bigram Freq.	Mean RT
angry	52980	5	0	4,613.25	559.97
animals	124835	7	0	3,793.17	643.906
blue	102172	4	6	1,405.00	573.647
brain	100235	5	7	5,898.00	583.875
brown	42423	5	8	1,802.00	653.265
clean	49764	5	3	4,522.75	582.829
dance	45680	5	3	3,414.00	530.323
famous	53575	6	0	2,290.00	649.912
finish	111094	6	0	5,530.60	617.594
join	170302	4	7	5,233.00	615.676
mouth	86406	5	5	2,485.25	537.912
plant	118323	5	5	4,575.50	576.788
present	37433	7	1	6,667.33	654.125
realize	161762	7	0	4,283.83	658.788
region	105382	6	1	4,586.60	598.484
relax	32948	5	1	4,098.50	617.033
respect	59226	7	0	4,997.50	595.781
secret	50730	6	0	3,838.20	696.97
sleep	95600	5	6	2,689.50	553.629
smile	73137	5	3	3,352.00	528.242
sweet	67903	5	7	1,545.75	555.824
train	50975	5	7	6,423.50	581.257
weird	43433	5	1	1,163.00	674.758
wild	61605	4	9	1,560.67	576

Table 34

Characteristics of Nonwords in the Sentence Reading Test

Words	Length	Orthographic Neighbors	Bigram Freq.	Mean RT
mooky	5	4	553	835.963
bairy	5	5	784.75	824.516
lurge	5	4	863	874.68
draze	5	4	1,028.50	854.565
slamp	5	5	1,068.00	866.2
biddle	6	4	1,079.20	849.231
snash	5	5	1,085.25	827.913
toyal	5	4	1,170.75	863.926
lobble	6	5	1,319.80	840.1
brune	5	5	1,342.50	867.56
pellow	6	5	1,433.80	859.966
farcel	6	3	1,438.20	825.355
tabble	6	4	1,530.80	855.269
fundle	6	2	1,551.00	841.727
sloker	6	3	1,837.60	823.062
tremp	5	2	2,029.75	864.593
sumper	6	4	2,038.60	824.909
rimped	6	4	2,157.20	868.8
grench	6	4	2,404.20	869.5
borale	6	2	2,621.60	829.406
luner	5	5	2,718.50	848.789
nince	5	5	2,860.75	823.3
ladin	5	2	2,960.50	847.762
coster	6	5	3,488.00	850.714

APPENDIX G: BACKGROUND QUESTIONNAIRE

BACKGROUNG QUESTIONNAIRE **Participant Number: …….**

PERSONAL INFORMATION

1. Gender: ……..
2. Age: ……..
3. Do you have any visual problem?

   Choose one.
4. If YES, how do you correct it?

   Choose one.

   If other, please specify: …….
5. Do you have any reading or learning disabilities (e.g., dyslexia)?

   Choose one.

   If YES, please explain (optional): …….
6. Year in college:

   ☐Freshman ☐Sophomore ☐Junior ☐Senior ☐MA/Ph.D.
7. Major field of study: ……..

LINGUISTIC INFORMATION

8. What is your native language?

   ……..
9. Do you speak any other languages?

   Choose one.
10. If YES, what languages do you speak?

    ……..
11. How many hours a day do you use English?

    ……..
12. How many hours a day do you use your native language?

    ……..
13. What language do you feel is your dominant language?

    ……..
14. How long have you been studying English?

…… (years)

15. How long have you been living in the U.S.?

…… (years) ……(months)

ENGLISH PROFICIENCY

16. Please rate on a scale of 1-6 your overall English proficiency.

beginner ☐	pre-intermediate ☐	intermediate ☐	upper-intermediate ☐	advanced ☐	native-like ☐

17. Please rate on a scale of 1-6 your current ability on English reading.

beginner ☐	pre-intermediate ☐	intermediate ☐	upper-intermediate ☐	advanced ☐	native-like ☐

18. Please rate on a scale of 1-6 your current ability on English writing.

beginner ☐	pre-intermediate ☐	intermediate ☐	upper-intermediate ☐	advanced ☐	native-like ☐

19. Please rate on a scale of 1-6 your current ability on English listening.

beginner ☐	pre-intermediate ☐	intermediate ☐	upper-intermediate ☐	advanced ☐	native-like ☐

20. Please rate on a scale of 1-6 your current ability on English speaking.

beginner ☐	pre-intermediate ☐	intermediate ☐	upper-intermediate ☐	advanced ☐	native-like ☐

APPENDIX H: QUESTIONS IN RETROSPECTIVE INTERVIEWS

Item-based questions: For these questions, I showed target pseudowords (and six distractors) in a randomized order to participants one by one, and they answered questions for each item separately.

1. Do you remember if you selected this word in the previous test (i.e., form recognition test)?

2. Do you recall reading this word in the book?
    a. If yes, do you remember any of the context(s) where this word appeared in the book? Do you remember what was going on in the story?

    b. Did you do anything to learn this word or keep it on your mind? Or did you just keep reading even though this word was an unfamiliar word for you?

    c. If you tried to learn the word, what kind of strategies did you use?

General questions:

3. What was your general strategy to deal with unfamiliar words in the book?

4. Did you expect to receive any vocabulary tests after reading the book?

5. What did you think of reading and listening at the same time? Did you like it or not? (only for the participants in the reading while listening group)

APPENDIX I: ENJOYMENT AND INTEREST QUESTIONNAIRE

Enjoyment and interest questionnaire

- *How much did you enjoy reading the book (The Time Machine)?*
  Please write a number out of 100. (0 being not enjoyed at all).

  …....

- *How interested were you in reading the book? In other words, was the book (The Time Machine) interesting for you?*
  Please write a number out of 100. (0 being not interesting at all).

  …....

APPENDIX J: MODEL COMPARISONS FOR EACH EYE-TRACKING MEASURE

Table 35

Model Comparisons for Each Eye-Tracking Measure

Measure	Model		R^2 *(Mar.)*	R^2 *(Cond.)*	*AIC*
Gaze Duration	Simplest	log(gaze duration) ~ group + exposure^1 + exposure^2 + exposure^3 + group: exposure^1 + group: exposure^2 + group: exposure^3 + (1|Subject) + (1|Item)	0.008	0.107	11957
	Maximal	log(gaze duration) ~ group + exposure^1 + exposure^2 + exposure^3 + group: exposure^1 + group: exposure^2 + group: exposure^3 + VST + enjoyment + interest + part of speech + concreteness + word length + (1|Subject) + (1|Item)	0.026	0.112	11953
	Best fit	log(gaze duration) ~ group + exposure^1 + exposure^2 + exposure^3 + group: exposure^1 + group: exposure^2 + group: exposure^3 + VST + concreteness + (1|Subject) + (1|Item)	0.026	0.122	11948
Regression Path Duration	Simplest	log(regression path duration) ~ group + exposure^1 + exposure^2 + exposure^3 + group: exposure^1 + group: exposure^2 + group: exposure^3 + (1|Subject) + (1|Item)	0.011	0.095	15348
	Maximal	log(regression path duration) ~ group + exposure^1 + exposure^2 + exposure^3 + group: exposure^1 + group: exposure^2 + group: exposure^3 + VST + enjoyment + interest + part of speech + concreteness + word length + (1|Subject) + (1|Item)	0.019	0.100	15349
	Best fit	log(regression path duration) ~ group + exposure^1 + exposure^2 + exposure^3 + group: exposure^1 + group: exposure^2 + group: exposure^3 + part of speech + concreteness + word length + (1|Subject) + (1|Item)	0.021	0.118	15345
Rereading Time	Simplest	log(rereading time) ~ group + exposure^1 + exposure^2 + exposure^3 + group: exposure^1 + group: exposure^2 + group: exposure^3 + (1|Subject) + (1|Item)	0.014	0.098	5453.9
	Maximal	log(rereading time) ~ group + exposure^1 + exposure^2 + exposure^3 + group: exposure^1 + group: exposure^2 + group: exposure^3 + VST + enjoyment + interest + part of speech + concreteness + word length + (1|Subject) + (1|Item)	0.026	0.093	5453.3
	Best fit	log(rereading time) ~ group + exposure^1 + exposure^2 + exposure^3 + word length + (1|Subject) + (1|Item)	0.025	0.103	5443.6
Total reading time	Simplest	log(total reading time) ~ group + exposure^1 + exposure^2 + exposure^3 + group: exposure^1 + group: exposure^2 + group: exposure^3 + (1|Subject) + (1|Item)	0.017	0.1512264	14188.44
	Maximal	log(total reading time) ~ group + exposure^1 + exposure^2 + exposure^3 + group: exposure^1 + group: exposure^2 + group: exposure^3 + VST + enjoyment + interest + part of speech + concreteness + word length + (1|Subject) + (1|Item)	0.027	0.149	13669.09
	Best fit	log(total reading time) ~ group + exposure^1 + exposure^2 + exposure^3 + part of speech + concreteness + word length + (1|Subject) + (1|Item)	0.024	0.143	13639.04

APPENDIX K: MODEL COMPARISONS FOR AWARENESS LEVELS

Table 36

Model Comparisons for Awareness vs. No Awareness

	Model	R^2 *(Marginal)*	R^2 *(Conditional)*	*AIC*
The model with the two main predictors	awareness ~ group + (1|Subject) + (1|Item)	0.00	0.28	1681.9
The maximal model	awareness ~ group + VST + enjoyment + interest + part of speech + concreteness + word length + total count + summed total reading time + group:summed total reading time + (1|Subject) + (1|Item)	0.11	0.30	1674.5
The best fit model	awareness ~ group + word length + total count + summed total reading time + (1|Subject) + (1|Item)	0.09	0.30	1665.3

Table 37

Model Comparisons for Noetic Awareness vs. Autonoetic Awareness

	Model	R^2 *(Marginal)*	R^2 *(Conditional)*	*AIC*
The model with the two main predictors	awareness ~ group + (1|Subject) + (1|Item)	0.00	0.21	1063.9
The maximal model	awareness ~ group + VST + enjoyment + interest + part of speech + concreteness + word length + total count + summed total reading time + group:summed total reading time + (1|Subject) + (1|Item)	0.10	0.22	1059.7
The best fit model	awareness ~ group + VST + word length + total count + summed total reading time + group:summed total reading time + (1|Subject) + (1|Item)	0.08	0.22	1050.9

APPENDIX L: MODEL COMPARISONS FOR VOCABULARY TESTS

Table 38

Model Comparisons for the Form Recognition Test

	Model	R^2 *(Marginal)*	R^2 *(Conditional)*	*AIC*
The model with the two main predictors	accuracy ~ group + test_time + group:test_time + (1|Subject) + (1|Item)	0.013	0.26	3389.6
The maximal model	accuracy ~ group + test_time + group:test_time + VST + enjoyment + interest + part of speech + concreteness + word length + total count + summed total reading time + group:summed total reading time + (1|Subject) + (1|Item)	0.089	0.265	3375.5
The best fit model	accuracy ~ group + test_time + group:test_time + VST + word length + total count + summed total reading time + group:summed total reading time + (1|Subject) + (1|Item)	0.084	0.265	3367.9

Table 39

Model Comparisons for the Meaning Recall Test

	Model	R^2 *(Marginal)*	R^2 *(Conditional)*	*AIC*
The model with the two main predictors	accuracy ~ group + test time + group:test time + (1|Subject) + (1|Item)	0.008	0.570	951.80
The maximal model	accuracy ~ group + test time + group:test time + VST + enjoyment + interest + part of speech + concreteness + word length + total count + summed total reading time + group:summed total reading time + (1|Subject) + (1|Item)	0.298	0.598	918.37
The best fit model	accuracy ~ group + test time + part of speech + total count _ (1|Subject) + (1|Item)	0.237	0.579	910.38

Table 40

Model Comparisons for the Meaning Recognition Test

	Model	R^2 *(Marginal)*	R^2 *(Conditional)*	*AIC*
The model with the two main predictors	accuracy ~ group + test time + group:test time + (1|Subject) + (1|Item)	0.003	0.308	2601.2
The maximal model	accuracy ~ group + test time + group:test time + VST + enjoyment + interest + part of speech + concreteness + word length + total count + summed total reading time + group:summed total reading time + (1|Subject) + (1|Item)	0.112	0.312	2595.5
The best fit model	accuracy ~ group + test time + VST + interest + part of speech + word length + total count + (1|Subject) + (1|Item)	0.111	0.311	2588.2

APPENDIX M: LEARNING GAINS FOR EACH TARGET PSEUDOWORD ON THREE VOCABULARY TESTS

Table 41

Learning Gains for Each Target Pseudoword on Three Vocabulary Tests

	Form Recognition		Meaning Recall		Meaning Recognition	
	Mean	SD	Mean	SD	Mean	SD
crimb	**0.75**	**0.43**	**0.09**	**0.29**	0.25	0.44
croom	**0.64**	**0.48**	0.01	0.09	0.13	0.34
demuse	**0.62**	**0.49**	**0.06**	**0.24**	**0.33**	**0.47**
fanish	0.42	0.49	0.00	0.00	0.10	0.31
fanky	0.42	0.49	0.02	0.13	**0.36**	**0.48**
fleak	0.47	0.50	**0.05**	**0.22**	**0.35**	**0.48**
halker	**0.63**	**0.49**	0.02	0.13	0.16	0.36
hilder	**0.64**	**0.48**	0.03	0.18	0.17	0.38
maive	0.53	0.50	0.01	0.09	0.17	0.38
mantil	**0.66**	**0.48**	0.03	0.16	0.19	0.39
merky	**0.69**	**0.47**	**0.04**	**0.20**	**0.38**	**0.49**
merse	0.52	0.50	0.03	0.16	**0.31**	**0.47**
mittle	**0.71**	**0.45**	**0.23**	**0.42**	**0.51**	**0.50**
plear	0.42	0.49	0.04	0.20	0.18	0.39
prome	**0.62**	**0.49**	**0.12**	**0.32**	**0.38**	**0.49**
rolley	**0.80**	**0.40**	**0.19**	**0.39**	**0.37**	**0.49**
slair	0.25	0.44	0.01	0.09	0.09	0.28
sorge	0.53	0.50	**0.22**	**0.42**	**0.41**	**0.49**
spoll	0.45	0.50	**0.08**	**0.27**	**0.27**	**0.45**
tartle	0.50	0.50	0.00	0.00	0.05	0.22
thrine	0.31	0.47	0.00	0.00	0.07	0.25
tickel	0.60	0.49	0.00	0.00	0.15	0.36
tifled	0.49	0.50	0.00	0.00	0.06	0.24
vandy	0.55	0.50	**0.12**	**0.32**	0.21	0.41

Notes. The first ten pseudowords with highest learning gains for each test are highlighted in bold.