THE ROLES OF CONTEXT AND REPETITION IN INCIDENTAL VOCABULARY
ACQUISITION FROM L2 READING: AN EYE MOVEMENT STUDY

By

Ayman Ahmed Abdelsamie Mohamed

A DISSERTATION

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

Second Language Studies – Doctor of Philosophy

2015

ABSTRACT

THE ROLES OF CONTEXT AND REPETITION IN INCIDENTAL VOCABULARY
ACQUISITION FROM L2 READING: AN EYE MOVEMENT STUDY

By

Ayman Ahmed Abdelsamie Mohamed

Research on extensive reading has provided ample evidence on the role of repetition in lexical learning and called for further research on the role of context in vocabulary acquisition from L2 reading (e.g. Chen, & Truscott, 2010; Horst, 2005; Waring, & Nation, 2004; Webb, 2007, 2008). On the other hand, eye movement studies on reading behavior documented cognitive effects of repetition and context quality on lexical processing and associated vocabulary learning with processing patterns in the light of the eye-mind link hypothesis (Rayner, 1998, 2009). The present study aimed at bringing together methods from both strands to investigate incidental vocabulary acquisition and track the cognitive roles of repetition and context predictability in the development of different aspects of vocabulary knowledge.

Forty-two upper-intermediate and advanced second language learners of English read a stage 1 graded reader, '*Goodbye Mr. Hollywood*', on a desk-mounted eye tracker screen followed by comprehension questions and vocabulary posttests. Target vocabulary consisted of 20 pseudo words and 20 known words with a range of repetition from 1 to 30. Eye-movement data showed that readers spent more time on pseudo words than on familiar words and that fixation times decreased across encounters with more attention given to target words on early encounters. Context predictability decreased total times spent on target words particularly on late encounters. Readers scored highest in form recognition followed by meaning recognition and finally meaning recall. Repeated exposure supported form recognition while context predictability supported meaning recognition and recall. Moment-by-moment lexical processing

showed that first fixations predicted form recognition while gaze durations predicted meaning recall. Total times spent on each encounter was positively associated with learning success in all vocabulary measures. When aggregating fixation times by vocabulary items, it was found that the amount of attention, as reflected in total reading times on each pseudo word across all its encounters, positively predicted learning outcomes above and beyond total exposure and item predictability, which highlights an important role of readers' individual attention and their optimal use of input to infer and retain meaning from context. Results of the study add a cognitive dimension to the concept of engagement in lexical learning and provide implications on the process of incidental learning from extensive reading and classroom teaching tasks.

To my beloved family
And to my beloved country, Egypt

ACKNOWLEDGMENTS

This dissertation is the culmination of long years of study and research in applied linguistics. Since I started college majoring in English as a foreign language, my life has been revolving around linguistics and my ultimate goal was to become a university professor. The journey for the PhD degree has been quite long and full of different kinds of academic, social and cultural experiences. From the moment I stepped into MSU campus, a whole new world opened for me and my family. We have gone through a lot, establishing a new life in a diverse community. We have fully enjoyed it to the extent that we cannot realize that the journey is coming to an end and that we should close this chapter and move on to the next step.

Many hands have supported my steps towards achieving my dream. The first kind hand was of my father. Although he was reluctant to letting me go away from him, he did not hesitate to financially support me for my first travel tickets. His priority was to see me a professor regardless of the suffering he had to go through all these years without me and away from his little grandsons. My mother shared similar feelings but remained patient over the years. I owe my progress in life and work to them for their care and support all the way.

At the academic level, I cannot express in words how I am grateful to my advisor, Dr. Aline Godfroid who brilliantly shaped my thoughts and developed my research abilities over the years. When Dr. Godfroid first saw me, I was still struggling in my second year trying to find my way. With constructive and intensive feedback, she challenged my abilities and managed to have me produce the best I could. I truly appreciate all the time and effort she spent with all my papers and messy drafts. She fostered my random thoughts about this project until it became something real. It was only through her continuous feedback and follow up that I was able to get my first

I also express my best wishes to all SLS peers who made life easier when we came together and shared our thoughts and challenges. I appreciate the good times, conversations and conference journeys I shared with Scott, Dominik, Jens, Roman and all the old and new members of the SLS family. I particularly extend my gratitude for Sehoon Jung with whom I shared many days of so-called 'study union' although I have frequently broken my promises with him, yielding to my family responsibilities. I will always cherish our memories together and truly appreciate his friendship.

In the background of all this, some people stayed in the shade patiently adapting to my hectic work style. These were my wife Samaa and my two kids Ahmed and Omar. At times we went through a lot of tension that threatened our relationship but Samaa was so patient and accepted the challenge. She suffered through my absence from home and my absent mind at home. As I am writing these lines, it happened to be her birthday. I wish her wonderful days ahead of us and I am sure she will be the happiest person when she realizes that her effort and patience were not in vain. I extend my best wishes to my son Ahmed, who is finishing his 3rd grade at school and little Omar, 3 years and a half, almost the age I spent in my PhD research papers. I can simply say that I missed my kids so much.

I am proud to be the first and only Egyptian Spartan in the SLS department to date. Egypt is a part of my personality and a part of who I am. The start of my PhD journey coincided with the first spark of extraordinary events and sudden changes in Egypt, which I only followed at a distance. As I am wrapping up my degree, I hold high hopes for my country to rise up and gain the fruits of its revolution. A final message that I have for everyone who knew me: please forgive my shortcomings, I am not a perfect person in any way. I believe that we will meet again for surely it is a small world.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Second language research has shown that reading plays an important role in the development of learners' vocabulary knowledge beyond what language classes and textbooks can offer (Coady and Huckin, 1997; Grabe and Stoller, 1997, 2002; Hill and Laufer, 2003 ; Horst, 2005; Huckin, Haynes and Coady, 1993 ; Huckin and Coady, 1999; Kweon & Kim, 2008; Matsouka & Hirsh, 2010; Lupescu and Day, 1993 ; Nagy, Anderson and Herman, 1987; Nation, 2001, 2006; Schmitt, 2008, 2010; Zimmerman, 1997). This type of learning has mainly been characterized as incidental because it occurs in the context of a meaning-oriented task with no intentional emphasis on vocabulary (e.g., Fraser, 1999; Paribakht and Wesche, 1999; Pulido, 2007; Waring and Nation, 2004; Watanabe, 1997). Various factors have been posited to facilitate incidental acquisition from written input such as type of task (Brown,Waring and Donkaewbua, 2008 ; Cho and Krashen, 1994 ; Hulstijn, 1992 ; Hulstijn, Hollander, and Greidanus, 1996 ; Hulstijn and Trompetter, 1998 ; Knight, 1994), repeated exposure (Horst, Cobb and Meara, 1998 ; Pigada and Schmitt, 2006 ; Rott, 1999 ; Webb, 2007) or context properties (Haastrup, 1989 ; Joe, 2010 ; Nagy, 1987 ; Nassaji, 2003; Webb, 2008 ; Zahar, Cobb and Spada, 2001).

Although incidental learning has been challenged as slow and inefficient in terms of acquisition and retention (e.g., Laufer, 2003, 2005; Macaro, 2003; Read, 2004), many researchers and teachers believe it is an essential supplement for learners to expand their vocabulary independently (see Schmitt, 2008, 2010). The argument is that learners face a lexical coverage challenge, given that a knowledge of 8000-9000 word families is required to achieve adequate comprehension of an authentic English text (Hirsh and Nation, 1992; Hu and Nation, 2000 ; Nation, 2001, 2006 ; Nation and Wang, 1999 ; Waring and Nation, 2004; Webb, 2010). Therefore, ESL programs usually incorporate an extensive reading component in their curricula

taking advantage of graded readers which are linguistically and lexically adjusted to learners' levels of competence and can support a smooth transition to unsimplified reading material and bridge the lexical coverage gap (Horst, 2005; Uden, Schmitt & Schmitt, 2014). In fact, extensive reading has been valued for its role in developing reading speed and fluency, reinforcing existing lexical knowledge and providing incidental learning opportunities for less frequent vocabulary (e.g., Cho and Krashen, 1994; Day and Bamford, 1998; Elley, 1991; Grabe and Stoller, 2002; Parry, 1991).

The potential of extensive reading to enhance vocabulary knowledge has been widely investigated (e.g., Cho and Krashen, 1994; Day, Omura and Hiramatsu, 1991; Horst, 2005; Hulstijn, Hollander and Geridanus, 1996; Pitts, White and Krashen, 1989; Saragi, Nation and Meister, 1978). A significant role was shown for repeated exposure (Horst, Cobb & Meara, 1998; Pellicer-Sanchez and Schmitt, 2010; Waring and Takaki, 2003; Webb, 2005) while less conclusive results were reported for the role of context quality and lexical inference on vocabulary learning outcomes (Fraser, 1999; Haastrup, 2008; Hu, 2013; Joe, 2010; Nassaji, 2003; Webb, 2008; Zahar, Cobb, and Spada, 2001).

Most studies on incidental learning from reading were paper-and-pencil based with outcomes measured through posttests or self-report. In an attempt to explain the trend of results in vocabulary studies, Schmitt (2008, 2010) emphasized the role of engagement with target vocabulary which, in his view, can be triggered by different factors including repeated exposure, increased noticing and increased time spent on target words. Expanding the concept of engagement, Hulstijn and Laufer (2001) relied on insights from Schmidt's (1990) noticing hypothesis and Craik and Lockhart's (1972) depth of processing hypothesis to introduce the involvement load hypothesis as a motivational cognitive construct for interpreting and predicting

2

the findings of vocabulary learning studies from a cognitive perspective. However, the hypothesis and its experimental replications (Laufer & Hulstijn, 2001; Kim, 2008; Keating, 2008; Mohamed, in press; Yaqubi, Rayati, & Gorgi, 2010), while informative, could not account for all facets of vocabulary learning because their findings only applied to controlled vocabulary-focused tasks and not to a natural reading setting.

The concepts of engagement and noticing, as cognitive processes underlying lexical learning, have been retrospectively discussed in studies that used think-aloud protocols or interviews (e.g. Fraser, 1999; Haarstup, 1991; Rott, 2005) as well as within the involvement load framework but they were not empirically measured. Because it was difficult to measure these cognitive processes offline, vocabulary researchers took advantage of the eye tracking technique as one advanced psycholinguistic method that has been posited to capture real time processing. This method can be adopted in L2 reading studies to track moment-by-moment processing of input based on the assumption that eye movements reflect an accurate representation of ongoing cognitive processes in a learner's mind. This assumption was coined 'the eye-mind link' which proposes a connection between overt and covert attention (see Rayner, 1998, 2009 for a review).

Reading behavior studies investigated the processing of short sentences and paragraphs in terms of repeated exposure (e.g. Hyönä & Niemi, 1990; Raney & Rayner, 1995; Rayner, Raney, & Pollatsek, 1995) and context predictability (Altarriba, Kroll, Scholl, & Rayner, 1996; Ashby, Rayner, & Clifton, 2005; Balota, Pollatsek, & Rayner, 1985; Clifton, Staub & Rayner, 2007; Ehrlich & Rayner, 1982; Juhasz & Pollatsek, 2011; Kliegl, Grabner, Rolfs & Engbert, 2004; Liversedge & Rayner, 2011; Rayner & Well, 1996; Rayner & Clifton, 2005; Wochna & Juhasz, 2013). However, none of these studies specifically looked at learning opportunities as related to lexical processing. In this regard, William and Morris (2004) and Brusinghan and Folk (2012)

found a systematic relationship between online processing patterns and retention of novel word meanings in reading comprehension. Godfroid, Boers, and Housen (2013) explained this association in terms of attention to novel words, maintaining that fixation times reflected the amount of attention to lexical items and predicted their subsequent recognition.

The current picture of incidental learning from reading thus points to two distinct strands of research. Mainstream vocabulary studies have shown strong evidence on the positive role of repetition yet mixed results on context effects, which may be due to the inconsistency of context rating methods adopted in these studies. On the other hand, eye movement studies have widely examined context predictability effects on reading times using standardized norming procedures but they did not investigate repeated exposure as much, and very few attempts were made to link processing with acquisition. The interaction between repetition and context quality was not directly investigated in these studies either. Finally, eye movement research in this area was based on sentence or paragraph reading, which makes their results less generalizable for longer text or in an extensive reading setting.

In the current study, I attempt to bring together the two strands of research by borrowing methods from both extensive reading research and eye movement studies to provide a picture of lexical processing in natural reading of novels and obtain a real time record of incidental learning of vocabulary from L2 reading. The initial hypothesis that motivates the current study is that exposure to novel lexical items during leisure reading invites some *attention* to form and meaning, which may be reflected in processing *time* and provide opportunities for incidental *intake* and *retention*. These opportunities are likely to be mediated by a hypothesized interaction between exposure frequency and context predictability.

4

To address the research questions, I implement eye-tracking methodology to investigate the online aspects of incidental vocabulary learning from an English graded reader, *Goodbye Mr. Hollywood*, which is a stage 1 short novel made available through Oxford University Press. The main goal of the study is to track the cognitive effects of repeated exposure and context predictability on English learners' reading patterns, and study whether the eye-movement reading measures can predict the development of different components of vocabulary knowledge, including form and meaning recognition and meaning recall.

This dissertation is organized into five chapters. In chapter 1, I review areas of the literature on vocabulary acquisition from reading and relevant eye movement studies. Chapter 2 describes the design, procedures, materials and research questions of the current study, and chapter 3 reports the results of these empirical questions. In chapter 4, I discuss the findings in light of the research questions. Finally, in chapter 5, I summarize the findings of the study, discuss pedagogical implications, address limitations, and make recommendations for future research.

CHAPTER 1: REVIEW OF THE LITERATURE

1.1 Incidental vocabulary learning

Incidental vocabulary acquisition is defined as the process of learning new words from meaningful input or meaning-based activities such as reading, listening, or interaction that has no particular focus on lexical items (Paribakht & Wesche, 1999; Richards & Schmidt, 2002). Earlier conceptualizations of incidental learning varied in how they distinguished it from intentional learning. Ellis (1994, 1999) distinguished them under two types of attention, arguing that in incidental learning the learner's primary attention is placed on meaning while allowing a secondary attention to be directed to form. Similarly, Hulstijn (2001, 2003) maintained that both types of learning must involve attention to varying degrees but the difference is that in incidental learning one does not intend to commit input to memory. Gass (1999) took a more conservative view, stating that incidental learning is more likely to be subconscious and less likely to involve deliberate attention or an active role from the learner. Bruton, Garcia Lopez and Esquiliche Mesa (2011) argued that what is characterized as incidental can be in some fundamental sense 'intentional' at least from the learner's perspective. Because paper-and-pencil studies could not track the existence or the amount of attention, they adopted a methodological distinction, derived from psychology, that learning outcomes were deemed incidental when learners were not expecting to be tested on the input they received (Hulstijn, 2001, 2003).

Several factors were hypothesized to encourage incidental learning including input factors such as word properties, salience and repetition, or individual factors such as proficiency, vocabulary size, increased attention and time devoted to target input, learner's first language, learning strategies and background knowledge (see Schmitt, 2008 for a review). Although the incidental learning rate was described as lower than that of intentional learning, it is now widely

acknowledged in language pedagogy that both modes of learning complement each other in the process of learners' incremental vocabulary development.

1.2 Empirical research on incidental learning

Early studies on incidental vocabulary acquisition were inspired by the interaction hypothesis (Long, 1985, 1996) which stated that communication and negotiation of meaning is a vehicle for language development (see Gass, Behney, & Plonsky, 2013 for a review). Numerous studies found support for this hypothesis in language development in general, particularly question formation (e.g., Gass & Varonis, 1994; Polio & Gass, 1998; Swain & Lapkin 1998; Mackey & Philp 1998). Following similar designs, it was also found that incidental vocabulary acquisition occurred as a byproduct of negotiation and output within interaction and speaking tasks (Ellis, Tanaka & Yamazaki, 1994; Ellis & He, 1999; de la Fuente, 2002; Brown, Sagers & LaPorte, 1999). Listening tasks were found to be conducive to vocabulary learning yet with lower rates than interaction tasks (Brown, Waring, & Donkaebua, 2008; Elley, 1989; Smidt & Hegelheimer, 2004; Vidal, 2010). Some classroom research reported learning outcomes from spontaneous class interaction and teaching activities (e.g., Dobinson, 2001; Mohamed, 2012). Horst (2010) contributed to this line of research with a corpus-based study that indicated many opportunities for incidental intake from teacher-talk and classroom communication.

Text-based tasks were more frequently investigated in vocabulary studies. Research in this area promoted engagement in reading tasks, either by manipulating word presentation and saliency in text or administering different tasks with varying degrees of complexity. For example, learners who inferred the meanings of certain words by having to choose from options provided retained words better than another group who were only provided the meanings of target words in a gloss (Hulstijn, 1992). Looking up meanings in a dictionary was a more

7

effective task than encountering meaning in marginal glosses (Hulstijn, Hollander, & Greidanus, 1996). Reading followed by vocabulary-focused exercises yielded better retention than reading with inferring meaning from context (Paribakht & Wesche, 1997). Reading combined with dictionary usage was more beneficial than reading only (Cho & Krashen, 1994; Knight, 1994; Luppescu & Day, 1993). Using words in a composition was more effective than only encountering words in reading comprehension (Hulstijn & Trompetter, 1998). To find a general interpretation of the common findings in vocabulary studies, Schmitt (2008) referred to engagement with lexical items as a key factor in vocabulary learning. Engagement, in his view, can be fostered by many factors, including, but not limited to, frequency of exposure, increased attention to target words, and increased time spent on the target items. In line with this claim, Watanabe (1997) and Peters, Hulstijn, Sercu and Lutjeharms (2009) found that the text input which affords increased processing due to contextual, lexical or semantic enhancement is more likely to yield more vocabulary gains (see Rott, Williams & Cameron, 2002; Rott & Williams, 2003).

Beyond paper-and-pencil results, some researchers have presented cognitive interpretations for vocabulary learning outcomes. Studies that used think-aloud protocols or interviews might have been the first to probe into the cognitive processes underlying lexical acquisition (e.g. Fraser, 1999; Haastrup, 1991; Paribakht and Wesche, 1999; Rott, 2005). In an attempt to drive the theory-building process, Laufer and Hulstijn (2001) introduced the *involvement load hypothesis* to account for the pattern of results observed in previous literature. The hypothesis was based on an analysis of the cognitive and motivational involvement imposed by any given L2-vocabulary task. Involvement, a cognitive-motivational construct, was defined as the combined effects of need, search and evaluation. Tasks that induce higher involvement

were hypothesized to produce higher vocabulary gains. The hypothesis received empirical support from several studies (e.g. Huang, Willson & Eslami, 2012; Hulstijn & Laufer, 2001; Keating, 2008; Kim, 2008). It also generated further research questions, for example Jing and Jianbin (2009) validated it in listening comprehension tasks while Eckreth and Tavakoli (2012) investigated a combination of involvement and repetition factors on vocabulary learning. Some counterevidence was reported regarding the re-evaluation of the components of the hypothesis regarding input vs. output-based tasks (Flose, 2006; Yaqubi, Rayati & Gorgi, 2010) and the role of individuals' accuracy in task performance on learning outcomes (Mohamed, in press). In general, the hypothesis can explain a good amount of the variance in incidental learning studies, yet it is not directly applicable to natural reading setting or leisure reading, which is proclaimed to have a significant role in learners' vocabulary development.

1.3 Vocabulary learning from L2 reading

Teachers and researchers generally agree that leisure reading beyond class material is a recommended path for lexical development above and beyond the most frequent vocabulary bands. However, when learners are directed to extensive reading of authentic text, they usually face a lexical coverage challenge. Nation (2001, 2006) calculated that the percentage of known words in a text should range between 95% and 98% in order for learners to obtain a sufficient comprehension level. It was thus calculated that authentic novels require at least a vocabulary size of 8000 to 9000 word families for adequate comprehension and new vocabulary intake (Hu & Nation, 2000; Nation & Wang, 1999; Waring & Nation, 2004).

Because it can take several years for L2 learners to reach higher levels of vocabulary size, extensive reading programs have taken advantage of simplified graded readers that are systemically adjusted to different levels. One important advantage of these readers is that they

can provide spaced repeated exposures to new and low frequent vocabulary and reinforce partially known words, which is an ideal setting for incremental vocabulary development.

1.3.1 Extensive reading and L2 vocabulary

Grabe and Stoller (2002) defined extensive reading as reading that exposes learners to "large quantities of material within their linguistic competence" (p.259). Proponents of extensive reading reported its value in increasing reading fluency, reading comprehension, and speed of access to frequent words as well as providing opportunities to meet new words, infer new meanings and build larger mental lexicons (Day & Bamford, 1998; Elley, 1991; Horst, 2005; Lai, 1993; Parry, 1991). One important benefit of extensive reading was reported by Uden, Schmitt, and Schmitt (2014) who found evidence that graded readers can support a smooth transition to authentic novel reading.

Several studies investigated the potential of lexical gains from graded readers and authentic novels. The classic study of Saragi, Nation and Meister (1978) used the novel *A Clockwork Orange* (1962) by Anthony Burgess. It was of particular interest because it included Russian slang words, referred to as *Nadsat*, which were targeted in reading experiments. They found that native English speakers were able to learn an average of 76 % of 90 Russian slang words used in the novel. Pitts, White and Krashen (1989) used one chapter of the same novel with second language readers and found modest rates of learning, about 6.4 % to 8.1 % of 30 target Russian words. Day, Omura and Hiramatsu (1991) reported that Japanese EFL learners learned an average of 3 words out of 17 target words encountered in a simplified short story, *The Mystery of the African Mask*. Horst, Cobb and Meara (1998) had learners read a simplified version of *The Mayor of Casterbridge*, and reported that learners could pick up an average of 5 words out of the 45 target words. Horst (2005) showed that readers picked up around 51 % of the

10

target words from selected extracts of graded readers. A common factor among all these studies was frequency of exposure in that learning chances increased as learners encountered target words more times in the text.

In addition to word meaning, the acquisition of other aspects of lexical knowledge was also investigated in extensive reading. Waring and Takaki (2003) used the 400 headwords graded reader *A Little Princess*. They found that learners scored higher in meaning recognition of the target words than productive translation and that scores in both tests dropped sharply after three months. Pigada and Schmitt (2006) found that a French learner showed considerable improvement in word spelling but a lesser command of meaning and grammatical knowledge after one month of extensive reading especially as exposures with target words increased. Webb (2005, 2007) reported that vocabulary encounters in reading or writing positively reinforced spelling, associations, syntax, grammatical functions, and form-meaning mapping. He found that the group that encountered the target words more than 10 times showed a better grasp of different aspects of word knowledge than other groups who received fewer exposures. Pellicer-Sanchez & Schmitt (2010) investigated vocabulary learning outcome from an authentic novel *Things Fall Apart*, and found that meaning recognition reached 84 % after ten exposures while meaning recall was still around 55 %.

Taken together, all previous studies suggest that reading yields different outcomes for different aspects of word knowledge, with more substantial gains in meaning recognition compared to other lexical aspects. What is also common is that all these studies point to the effect of repeated exposure; specifically, an average of 8 to 10 repetitions was shown to be appropriate for the development of receptive knowledge of vocabulary with relatively low gains in productive knowledge (Schmitt, 2010). Finally, the amount and quality of learning

demonstrated in previous research indicate that incidental learning from reading is possible but retention is not durable unless a learner receives further exposure within a reasonable time span. Schmitt (2008) suggests supplementing extensive reading with an explicit teaching component or activities to enhance engagement and maximize the benefit of exposures. Table 1 summarizes the findings of extensive reading studies and highlights the roles of exposure and context in vocabulary learning.

1.3.2 Contextual richness and vocabulary learning

A basic assumption in learning vocabulary from reading is that learners will use their linguistic resources and lexical inferencing to derive meanings from context and thus be able to retain some knowledge of words if they get repeated over time (Fraser, 1999; Paribakht & Wesche, 1999). Some research indicates that guessing from context is unreliable in learning vocabulary (Laufer, 2005; Nassaji, 2003). In fact, two opposing views were presented in this regard. Schouten van-Parreren (1989) argued that informative contexts support guessing ability, which in turn may transfer to learning. On the other hand, Mondria and Wit-de Boer (1991) argued that rich context can aid comprehension but it diverts attention from the lexical level and that even correct guessing does not guarantee retention. Mondria (2003) found that meaning inference was time consuming and less efficient than other explicit methods of retention. In the same line, Hu and Nassaji (2012) found that ease of guessing affected word retention negatively.

Empirical research on context effect reported inconclusive results. Schwanenflugel, Stahl and McFalls (1997) found no evidence for the role of contextual support in vocabulary development of elementary school children. Zahar, Cobb and Spada (2001) found no clear association between the learning outcome and the quality of contexts in which lexical items occurred. Instead, they suggested that variable contexts are favorable for effective inferencing

12

and retention and that unclear contexts can be ideal for triggering more attention at the lexical level, which sets the scene for meaning retention. Similarly, Haastrup (1989) argued that meeting words in less informative contexts invites more cognitive engagement and thus increases chances of meaning recall in subsequent contexts. Webb (2008) investigated context quality and the effect of repeated exposure in a controlled reading study. He found that while repetition supported form recognition, the quality of context was associated more with meaning recognition. This may indicate that a rich context aided guessing and retention to a certain degree. Joe (2010) found that encountering target words repeatedly in a wide range of tasks is more conducive to vocabulary retention than contextual richness. Hu (2013) found a similar conclusion in that repeated exposure affected knowledge of form while contextual richness was more beneficial to form-meaning connections and grammatical functions.

One possible reason for the somehow mixed results regarding context effects may be related to the way context predictability has been operationalized. Many studies adopted the classification of contexts provided by Beck, McKeown and McCaslin (1983) which categorizes contexts into misdrective, nondirective, general and directive (Zahar et al., 2001; Hu, 2013). Schwanenflugel, Stahl and McFalls (1997) rated contexts from 1 (low transperency) to 4 (high transparenecy). Webb (2008) had two native speakers rate the conetxts from 1 (misleading) to 4 (high chance of lexical inference). An alternative method of measuring predictability, derived from psycholinguistics, is through a modified cloze procedure where native spakers' percentage of agreement in predicting the missing word determines the degree of predictability. Schwanenflugen and LaCount (1988), based on previous literature, defined a high constraint cutoff at 78% or above and low constraint at 68% and below. Table 1 summarizes the roles of exposure and context in vocabulary learning from reading studies.

1.4 Interim summary

Up to this point, I have reviewed how early research defined and operationalized incidental learning and tested it in different modalities: speaking, listening, reading, interaction and text-based tasks. I then discussed how extensive reading programs made use of graded readers to provide incidental learning opportunities for ESL students. As Table 1 indicates, extensive reading research showed significant gains of vocabualry in different aspects of form recognition, meaning recognition and recall. Generally, more than 10 exposures was the recommended threshold for substantial word knowledge gains. Similarly, research on text-based tasks showed a significant role of reading in vocabulary learning with established strong effects of frequency of exposure. Fewer studies looked at the role of context in extensive reading setting. The effect of context was generally unclear or correlated more with meaning recognition rather than the knowlede of word form. Adequate measures for context predictability can shed more light on the pattern of reported learning outcomes from extensive reading research.

In the following section, I argue that insights from online processing can add to our understanding of the lexical factors that determine vocabulary acquisition from context. I outline how psycholinguistic approaches investigated the same factors from a cognitive perspective, particularly through eye tracking.

Table 1

The role of exposure and predictability in vocabulary learning

Study	population	Reading material	Vocabulary gains	Effect of exposure	Effect of predictability
Saragi, Nation & Meister (1978)	20 native English speakers	The authentic novel *A Clockwork Orange*	76 % of 90 Russian slang words	Minimum 10 exposures for learning	exposure mitigated by context
Pitts, White & Krashen (1989)	51 ESL learners	*A Clockwork Orange*	6.4 % - 8%	Not tested	Not tested
Day, Omura & Hiramatsu (1991)	292 Japanese EFL learners	*The Mystery of the African Mask*	17 % of 17 target words	Not tested	Not tested
Horst, Cobb, & Meara (1998)	34 EFL learners	*The Mayor of Casterbridge*	20% of 23 target words	Strong effect	Not tested
Zahar, Cobb, & Spada (2001)	144 ESL students	*The Golden Fleece*	2.3 of 30 words	Strong effect	Subordinate to exposure

Table 1 (cont'd)

Study	Population	Reading material	Vocabulary gains	Effect of exposure	Effect of predictability
Waring & Takaki (2003)	15 Japanese EFL	*A Little Princess* (25 target words)	15 word forms; 20 word meanings	+18 exposures	Not tested
Horst (2005)	21 ESL students	Several graded readers	17 words	Not tested	Not tested
Pigada & Schmitt (2006)	One French learner	Four graded readers	8-23 %	20+ exposures	Not tested
Webb (2007)	121 Japanese EFL	10 paragraphs	Multiple aspects	10 + exposures	Not tested
Webb (2008)	50 Japanese EFL	30 sentences	Multiple aspects	Effective for form	for meaning
Sanchez & Schmitt (2010)	20 Spanish EFL	*Things Fall Apart*	84% meaning	+10 exposures	Not tested
Joe (2010)	One Turkish ESL	Class material	77 % of 20 words	Strong effect	Not significant
Hu & Nassaji (2012)	11 ESL learners	Academic text	Significant gains	Not tested	Negative effect
Hu (2013)	One ESL learner	Graded readers	Significant gains	Effective for form	Effective for meaning

1.5 Cognitive perspectives on lexical learning

The previous review points to a possible interaction between exposure frequency and contextual richness that may be responsible for different attention patterns from readers and thus variable learning outcomes. Based on Schmidt's (1990) noticing hypothesis, vocabulary researchers assume that readers need to notice novel words in context based on text properties or lexical features, and that this pattern of noticing would determine the nature of learning outcomes. However, it is difficult to test this assumption offline because retrospective measures that have been used to track noticing such as note taking, underlining or think-aloud protocols can be less sensitive in capturing moment-by-moment processing of context. Godfroid et al (2013) reviewed these measures, concluding that a more precise and complete account of cognitive processing during reading can be fulfilled by the eye tracking technique, which can provide a more sensitive measure of the amount and locus of attention during processing.

1.5.1 Eye tracking

Eye tracking is defined as the online recording of learners' eye movement behavior, which is described in terms of fixation times (how longer readers look at interest areas) and saccades (the movement of the eyes from one point to the next) (Godfroid, 2012). Reviews of eye tracking research show that eye movements provide an accurate representation of the cognitive processes in the reader's mind. This assumption was coined the 'eye-mind' link, which proposes a connection between overt and covert attention (Rayner, 1998, 2009). In reading research, many variables were tested such as word properties, such as frequency, predictability, familiarity and other context variables in order to examine their effects on reading behavior as measured by eye tracking.

1.5.2 Eye movement models

A large amount of research used recordings of eye movements to explore the psychological processes that control the reading behavior of adult skilled readers (see Rayner, 1998, 2009 for a review). Several computational models were developed to explain the characteristics of reading behavior based on the assumption that there is a strong relationship between lexical encoding and eye fixation measures (Liversedge, Gilchrist, & Everling, 2011; Van Gompel, Fischer, Murray, & Hill, 2007). These models were categorized into serial-attention and parallel-attention models. Serial attention models assume that attention is allocated sequentially to support lexical processing of one word at a time and that lexical processing causes the eye to move from one word to the next (e.g. Reader model: Just, & Carpenter, 1980; EMMA model: Salvucci, 2001; E-Z Reader model, Reichle, Rayner, & Pollatsek, 2003).In parallel attention models, processing is shared to neighboring words due to their specific characteristics (e.g. SWIFT model: Engbert et al., 2002).

Although no model was claimed to account for the whole picture, the E-Z reader model was found to be the most comprehensive in linking lexical recognition process to eye fixations because it provided assumptions as necessary to account for sophisticated observations in reading behavior (Liversedge, Gilchrist, & Everling, 2011). Simulations of eye movements in reading studies showed that the E-Z reader assumptions and the serial attention hypothesis is sufficient to account for reading behavior in alphabetic and non-alphabetic languages (see Pollatsek, Reichle & Rayner, 2006; Rayner, Ashby, Pollatsek, & Richle, 2004 for a full review).

A key assumption of this model is that lexical factors influence when the eyes move in that an early stage called familiarity check triggers the eyes to move to the next word, while later stage of full lexical access causes covert attention to shift to the next word. The mean time spent

on lexical items is the time required for familiarity check, which is influenced by item frequency of occurrence and within sentence predictability. If the next word is highly frequent or predictable, it will most probably be skipped, being processed entirely para-foveally, in which case a familiarity check stage is initiated for the following word to proceed with reading. In the light of this model, it was found that specific early eye movement measures like gaze duration can exclusively reflect a familiarity check stage in lexical processing (Juhasz & Pollatsek, 2011). An important assumption of this model is that the durations of both familiarity check and lexical access are highly sensitive to lexical factors such as word frequency, word familiarity, repeated exposure, lexical ambiguity, age of acquisition, context predictability, morphology and plausibility (Clifton, Staub & Rayner, 2007).

1.5.3 Eye movement research in reading

Many eye movement studies have looked at native and nonnative speakers' processing of written input and responding to different lexical and contextual features. Hyönä and Niemi (1990) used the repeated reading paradigm with Finnish readers. The readers' fixation times decreased consistently from first to third encounter with target sentences, and the number of their progressive fixations and regressions also decreased. Similarly, Raney and Rayner (1995) investigated the effects of repeated exposure on native-English speaker's second reading performance. They found that individuals had shorter reading times, made fewer fixations, and had longer saccades during the second reading of the same text. Moreover, shorter fixation durations were associated with high frequency words, suggesting independent effects of word frequency and repetition on reading times. Rayner, Raney, and Pollatsek (1995) found similar results regarding the effect of three repetitions of lexical items in a given text, and they also found frequency effects after the first two repetitions, but no further differences occurred after

that, which indicated that word frequency was mitigated by repetition. Recently, Joseph, Wonnacott, & Nation (2014) found significant decreases in reading times as a function of repated exposures and shorter reading times for novel words that were presented earlier in the text than later items. Early presneted words were remembered more accurately in an offline post test. In their results, they advocated an important effect of age of aqcuisition on lexical processing and learning.

Regarding lexical processing of context, eye movement studies have consistently shown that high context predictability is associated with shorter fixations and more skipping than low predictability contexts. Ehrlich and Rayner (1981) presented passages with target words the meaning of which was constrained (i.e., predicted) by the preceding context. They found that readers fixated more on target words in low-constraint contexts words in given paragraphs. Moreover, the readers tended to be less sensitive to misspellings of the target words in high-constraint contexts. Similarly, Rayner and Well (1996) found that readers fixated more on target words when they occurred in low-constraining contexts than in medium or high-constraint contexts. The probability of skipping was higher in high-constraining contexts compared to other context conditions. Kliegel, Grabner, Rolfs and Engbert (2004) also reported that high predictability increased skipping rates and it was associated more with second pass reading. Rayner, Ashby, Pollatsek and Reichle (2004) found that skipping was affected by predictability more significantly in high frequency target words. In contrast, Ashby, Rayner and Clifton (2005) found that lexical frequency and predictability independently affected reading times and patterns of processing. They also found qualitative differences between groups of readers as skilled readers were more sensitive to predictability and more consistent in word recognition patterns than average readers.

Few studies have investigated a potential association between online processing patterns and learning new words. Chaffin, Morris, and Seely (2001) found that the familiarity of target words and context quality (informative or neutral) determined the amount of time readers spent on the target words in that learners fixated the most on novel words encountered in neutral contexts. Williams and Morris (2004) examined the effect of word familiarity in reading comprehension and word recognition. They found that readers spent more processing time on novel words than familiar words, and that there was a systematic relationship between online processing patterns (i.e. reading times), and retention of new word meanings. Brusnighan and Folk (2012) conducted a self-paced reading study on incidental vocabulary learning. They found that readers spent more time processing sentences that contained novel compound words, and that they were able to retain new word meanings from a single exposure. They made a case for a strong relationship between increased processing times and accuracy in vocabulary retention measures. They stated that skilled readers spend extra time on difficult items to establish form-meaning connections, which results in memory traces which are available for later recall. On the level of context, they found that opaque contexts triggered higher rereading times and slower processing than transparent contexts, which was considered an ideal situation for meaning inference and retention of target words.

One recent study that specifically targeted vocabulary in second language reading was conducted by Godfroid et al. (2013). They operationalized attention to novel pseudo words as a quantitative variable reflected in the participants' eye fixation times during reading. Twenty-eight advanced EFL learners read 12 paragraphs in English with target areas that consisted of known words, pseudo words or a combination of both. Results showed that readers fixated longer on pseudo words than on known words, regardless of whether these pseudo words were

combined with appositive cues. There was a significant association between the total fixation time on pseudo words and subsequent recognition of these words in a surprise posttest.

Taken together, the eye tracking studies widely investigated both exposure and predictability from a processing perspective yet focused less on learning opportunities as function of processing patterns. Table 2 summarizes the findings of these studies and their implications on the effects of exposure and context.

1.6 General summary

The previous review hints at how the recent trends in applied linguistics can explain findings from earlier studies of vocabulary acquisition from reading. Tables 1 and 2 summarize the current picture of vocabulary learning form extensive reading and eye tracking research traditions. Several studies has validated the potential of extensive reading to foster vocabulary learning in general and support the development of different components of word knowledge. Within this research tradition, many studies has found evidence for the importance of repeated exposure in vocabulary development but less attention was given to the role of context in the process of incidental learning.

On the other hand, Table 2 shows that eye movement research in reading has given a considerable focus on context and repetition from a processing perspective with only few attempts to associate online processing with incidental vocabulary acquisition. While extensive reading research primarily investigated second language reading in authentic or simplified lengthy texts, eye movement studies relied more on customized sentences or short paragraphs read by native speakers.

Table 2

The roles of exposure and predictability in eye tracking studies

Study	Population	Reading material	Effect of exposure	Effect of predictability	Vocabulary
Hyönä and Niemi (1990)	11 Finnish speakers	Text of 371 words (read 3 times)	Decreased fixation times and longer saccades	Not tested	Not tested
Rayner, & Raney (1995)	28 English speakers	16 short passages Read 2 times	Decreased fixation times and longer saccades	Not tested	Not tested
Ehrlich and Rayner (1981)	24 English speakers	48 paragraphs	Not tested	Shorter fixations and more skipping	Not tested
Joseph, Wonnacott, & Nation (2014)	37 college students	16 sentences read 5 times	Decreased reading times	Not tested	Retention of earlier
Rayner and Well (1996)	18 English speakers	36 sentences	Not tested	Shorter fixations	Not tested

Table 2 (cont'd)

Study	Population	Reading material	Effect of exposure	Predictability	Vocabulary
Kliegel, Grabner, Rolfs and Engbert (2004)	50 native German	144 German sentences	Not tested	More skipping and longer second reading times	Not tested
Rayner, Ashby, Pollatsek and Reichle (2004)	44 Native English	32 English sentences	Not tested	More skipping in high frequency words	Not tested
Ashby, Rayner and Clifton (2005)	44 Native English	24 English sentences	Not tested	Shorter reading times for skilled readers	Not tested
Williams & Morris (2004)	24 Native English	48 English sentences	Not tested	Not tested	Tested
Brusnighan and Folk (2012)	56 Native English	English sentences	Not tested	Better retention	Tested
Godfroid et al. (2013)	21 ESL learners	12 paragraphs	Not tested	Not tested	Tested

1.7 Goals of the study

The goal of the present study is to bring together methods from extensive reading tradition and eye movement research to investigate the online aspects of incidental vocabulary learning from L2 reading with a focus on the effects of repeated exposure and context predictability on processing and vocabulary intake. Looking at reading patterns of novel words in context can help us interpret the concepts of engagement and noticing more precisely and associate them with learning outcomes. Tracking moment-by-moment interaction with the text can provide a cognitive picture of the factors that increase or decrease attention to target words, as reflected in online fixation measures.

I aim to investigate the role of different fixation measures in predicting vocabulary intake; in other words, how repeated exposure and context cues provide opportunities for readers to combines bits of information about novel words over successive encounters. To investigate the holistic effects of attention and exposure, I also aim to test the role of summed online measures, the total times readers spent on individual target words, in predicting the variance in vocabulary outcomes, and whether these processing aspects override or support the roles of total exposures and predictability of novel words in L2 reading environment.

CHAPTER 2: CURRENT STUDY

2.1 Research questions

The current study is guided by the following research questions:

(1) How do learners of English in the study process novel lexical items in silent reading relative to known control items? And how do repeated encounters and predictability influence lexical processing of pseudo words compared to control words in the text?

(2) What are the text-based effects of repeated exposure and predictability of novel target words in an L2 English text on the acquisition of receptive and productive knowledge of form and meaning of target words in vocabulary posttests?

(3) To what extent do moment-by-moment eye fixation times on successive encounters with target words predict the learning gains of L2 readers in the vocabulary knowledge posttest?

(4) To what extent do summed reading times of target words predict successful form and meaning gains in vocabulary posttests? And how do online predictors compare to text-based effects of exposure and predictability on vocabulary learning from reading?

2.2 Participants

The participants in this study were 42 advanced second language learners of English (22 females and 20 males) ranging in age from 19 to 35 ($M = 22$, $SD = 4.2$). Thirty participants were undergraduate international students with diverse majors who were also enrolled in advanced ESL reading and writing classes, and 12 participants were graduate students mostly majoring in scientific and engineering fields. Participants represented different language backgrounds including Chinese ($N=13$), Arabic ($N=4$), Spanish ($N=5$), Portuguese ($N=5$), Japanese ($N=5$), African languages ($N=5$), Hindi ($N=2$), in addition to single representations for Korean, Polish and Russian. Proficiency levels were determined based on self-reports of recent TOEFL IBT

scores that ranged from 79 to 100 (M =89, SD =7.3). The minimum of 79 in TOEFL is the cut-off required for undergraduate studies at MSU. Their vocabulary sizes, measured at the 5k level using Meara's (1992) vocabulary size test, yielded an average of 3908 ($SD = 659$). Detailed information about participants' background and levels are provided in Appendix A.

2.3 Material

2.3.1 Background questionnaire

A one-page language background questionnaire was prepared to collect basic information about participants' native languages, majors of study, English learning years and other languages spoken or used. Participants were also asked to provide the most recent TOEFL IBT or any other proficiency test score they received in addition to self-rated proficiency on a scale from 1 to 9 in the areas of reading, writing, vocabulary and overall proficiency. A sample of the questionnaire is shown in Appendix B.

2.3.2 Vocabulary size test

To confirm that students' vocabulary levels matched the selected reading material, a yes-no vocabulary size measure, adapted from Meara (1992), was planned to be administered prior to the experimental session for each participant. The test comprised 5 levels targeting the first 5,000 most frequent words according to Nation (2001). Each level contained 60 words (40 real words and 20 non-words). The score on each level is calculated based on the estimation of hits (real words checked as known) against false alarms (non-words checked as known). A participant's vocabulary size at 5k is estimated as the sum of scores across the five levels multiplied by 10. This particular test was selected for its quick administration besides the fact that the experimental reading material will be targeting the 5k level in lexical coverage. Examples of this test can be found online at (http://www.lextutor.ca/tests/).

2.3.3 Reading material

In selecting an appropriate reading material for the study, it was essential to have a text that is appropriate to learners' level of English-in terms of language structure and lexical coverage. It was equally important to provide an amount of lexical richness with a good spread of vocabulary items showing variable repetition patterns. A third criterion for the text was to maintain a reasonable length while taking into consideration the practical issues of implementing the eye tracking methodology in reading. After screening several resources of modern short novels, it was found that graded readers would be more relevant for the purposes of the study because they were no less authentic and they would easily satisfy the requirements for length and controlled lexical features.

The search for a graded reader involved consultations with ESL teachers and browsing the library resources of the English Language Center. Several short novels were inspected for content and length and then run through the Range software (Heatley, Nation and Coxhead, 2002), which lists the words in a given text according to their frequency and word families. The final selection was a short novel *Goodbye Mr. Hollywood* by John Escott, which is a stage 1 (400 headwords) graded reader made available through the Bookworms Library, Oxford University Press. It is available in print with a word count of 5400 (642 types and 372 word families) and classified under thriller and adventure stories. The text was cut down to 4649 words (595 types and 394 word families) by adjusting encounters of target words and taking out unnecessary details. Accordingly, the lexical density (types/tokens ratio) was not high (12.9%). Range output confirmed that the lexical coverage of the story is at the 5,000 word level. Table 3 outlines the lexical distribution of the text across frequency levels.

Table 3

Lexical profile of "Goodbye Mr. Hollywood"

Word List	Tokens (*percentage*)	Types (*percentage*)	Families
1000	4074 (87.6%)	479 (80.5%)	328
2000	309 (6.65 %)	61 (10.25%)	49
3000	38 (0.82%)	12 (2.02%)	9
4000	13 (0.28%)	6 (1.01%)	4
5000	22 (0.47)	6 (1.01%)	4
Not in the lists	193 (4.15%)	31 (5.21%)	
Total	4649	595	394

The lexical profile for the novel shows that the text is densely populated with high-frequency vocabulary (the first 1000 words) and less populated with words at the 2000 level while very few tokens appear from the rest of the levels. The tokens not in the list constituted only proper nouns and names of people and places. Two versions, which differed only in terms of the target words assigned in each one, were created of the original text. To determine the list of target words, Range lexical analysis was inspected for frequency (number of occurrences) of certain vocabulary items. A sample chapter of the original story is provided in Appendix C.

2.3.4 Target words

The final list of the target words consisted of 40 items with occurrences ranging from 1 to 30. These words were equally split into two lists (20 items each), of which a given participant saw one list as experimental items (i.e., pseudo words) and the other list as familiar English controls. Each vocabulary item in the first list matched another item in the second list in part of speech, and number of letters and syllables. Because the graded reader contained all familiar words that were estimated to be a part of participants' lexical repertoire, the experimental items in each version were replaced by matching pseudo words retrieved from online resources

especially the ARC Nonword Database (http://www.cogsci.mq.edu.au/~nwdb/) and previous

vocabulary research (Godfroid et al., 2013; Webb, 2007, 2008). The pseudo words in one version

of the story appeared in their familiar forms in the other version and vice versa. With this

procedure, the two versions were counterbalanced and every pseudo word in a given context had

a familiar counterpart in the other text version. To minimize item effect, a single pseudo form

was made to substitute two different words: one in each story version.

A large list of pseudo words was passed over to two native speakers who intuitively

edited or excluded some of them. Additionally, the same list of pseudo forms was passed around

in an ESL class with 9 international students who were asked to judge whether the listed items

could be actual English words. These students were not participating in the actual study. Taking

all feedback in consideration, a finalized list of 20 pseudo items was created to substitute for the

40 target items in the experiment. Each pseudo item matched the real word in number of letters

and syllables to minimize visual effects on eye movements. Table 4 outlines the target words in

the two versions, the number of times they appeared in text and their substitute pseudo words.

The total number of pseudo tokens in each version was 121, which accounted for 2.6 %

of the total tokens in the text. This guaranteed that the reading material provided approximately

97.4% of lexical coverage, which falls within the recommended lexical coverage range of (95%-

98%) to ensure reading comprehension and the ability to guess novel words from context

(Nation, 2006). Based on these criteria, pseudo words were inserted and the text was divided into

shorter parts (seven chapters) and shorter paragraphs in preparation for programming.

Table 4

Pseudo forms and their frequency in the reading text

Version A targets	Version B targets	Pseudo words	Number of encounters
hotel	table	fozle	30
café	room	gube	18
face	desk	mave	10
stop	meet	tund	9
tall	busy	leam	7
kill	push	blef	6
party	money	toker	6
pocket	window	bannow	5
bag	gun	mot	5
picture	airport	fonteen	4
quiet	happy	dangy	4
garden	letter	windle	4
shirt	dress	neech	3
accident	hospital	redaster	3
rich	cold	dook	2
sleep	drink	tance	1
cinema	camera	pamery	1
famous	hungry	tantic	1
plane	noise	dorch	1
chair	shoes	smick	1

2.3.5 Comprehension packet

A 50-item comprehension test (5-8 items per chapter) was created to monitor readers'

understanding of the main content of the story. The items included a combination of true/false

statements and multiple choice questions depending on the content of each chapter. The test was

printed out in seven pages (one page per chapter) along with characters' illustrations copied from

the story book to foster reader engagement and visualize the content. A sample page of the

packet is provided in Appendix D.

2.3.6 Reading perception questionnaire

To gauge readers' interest and enjoyment during reading, a short 10-item questionnaire,

adapted from Uden, Schmitt and Schmitt (2014), was used as a post-reading task. The items were

in the form of short statements with a six-point Likert scale where 1 indicates 'strongly disagree'

and 6 indicates 'strongly agree'. The statements mainly revolved around readers' enjoyment,

ease of reading and their overall comfort through the experiment. See Appendix E for a copy of

the questionnaire.

2.3.7 Vocabulary tests

To obtain a multi-faceted picture of incidental lexical development, it was important to

include multiple measures of vocabulary knowledge (Nation, 2001; Schmitt, 2008, 2010). Three

vocabulary tests were prepared to measure form recognition, meaning recognition and meaning

recall of the target pseudo words. In general, only these target words were identical in all the

tests while distracter items differed. Because the target words carried two different meanings

according to the story versions, all the tests were adjusted to accommodate this factor by

including the two meaning options in the meaning recognition test and considering two different

responses in the scoring procedures of the meaning recall test. All tests were scored in a binary fashion where zero means "no response' or incorrect answer and 1 refers to the correct response.

Form recognition test. This test comprised 100 vocabulary items including the 20 target pseudo words, familiar words from the text and other sources and pseudo words out of the text. The instruction for the task is to circle only the words that were seen in the reading material. A copy of the test is found in Appendix F.

Meaning recall test. This test included the 20 target pseudo words in addition to 10 distracter items that represented off text pseudo words, familiar words from the list and other low frequency English words. The task was to recall meanings, synonyms, related words or semantic fields for the given items. A sample of the test is provided in Appendix G.

Meaning recognition test. This is a multiple choice test with 30 items covering the target words along with other additional pseudo words, familiar words and low frequency words. Each item had five meaning options in addition to an 'I don't know' option to minimize guessing. A sample of the test is shown in Appendix H.

2.4 Procedure

2.4.1 Apparatus

Before any participants were invited into the lab, the reading material was programmed into the desk-mounted EyeLink 1000, an eye-tracker manufactured by SR Research (http://www.sr-research.com/). The story was copied into the Experiment Builder and set up in two versions so that a participant can selectively be assigned to one experiment file at the time of participation. The text was typed in Courier New font size 18, on a 19-inch computer monitor set up 55 cm from the participants' eyes. The font color was black on a light grey background.

The full experiment file consisted of 87 screens including introductory pages, instructions and break transitions. The main story content was thus provided in 70 screens, each containing 60-70 words in double spaced text. Minor editing was performed on the displayed text to confirm that target words did not appear in the beginning of slides and/or at the beginning and end of sentences. Each chapter was captured in a range of 7 to 11 screens. Breaks were offered at the end of each chapter in the story. Eye calibration was set to be performed at the beginning of the experiment and after the return from breaks. Participants moved across screens using a button on the right side of a hand-held controller. Drift correction was set up at the beginning of each page. Participants placed their heads on a chin and forehead rest during reading time to minimize head movements.

2.4.2 The reading session

The participants for the study arranged individual meeting with me in the eye tracking lab run by Second Language Studies Program at Michigan State University. After signing the consent form and filling the background questionnaire, they took the vocabulary size test and prepared for the eye tracking session. I started with an introduction about the story and main characters and asked each participant if he/she had read it before and what was his/her expectations about the incidents in the story based on the title and illustrations of main characters. I told participants that they would be tested on the content of the story so that they would pay attention during reading but I did not explicitly forewarn them about any vocabulary testing. Once the participant was warmed up for reading, I started giving directions regarding the use of the eye tracking equipment then I performed the initial calibration. Two participants did not pass the calibration stage so they could not continue with the experiment.

I randomly assigned participants to either version A or B in the experiment builder. Once a participant is done with a chapter, a break prompt appears on the screen. At the beginning of the break, he/she would take the comprehension packet on the side of the desk and respond only to the questions on the chapter he/she has just finished. Whenever he/she was ready, the participant would return to the chin rest and perform calibration for the following chapter. The same procedure continued with the rest of the chapters. I gave a longer break by almost the middle of the story (around chapter 4 or 5) and provided snacks for the reader. When the participant reached the end of the story, he/she would complete the last page in the comprehension packet then move away from the eye tracker to another desk in the lab. The reading session for each participant including calibration, breaks and comprehension check took an average of 45 to 70 minutes.

2.4.3 The testing session

The testing session started with the reading perception questionnaire to gauge their attitudes and feelings about the eye tracking experience and the story. They then took the vocabulary tests in the following order to avoid transfer effects: form recognition, meaning recall and meaning recognition. The testing session for an average participant took a maximum of 10 to15 minutes and it was the final task required from participants.

2.4.4 Modified cloze procedure

To retrieve predictability information of target words, we needed to look at the text from a native speaker perspective. A norming study was designed in which the two original versions of the story, with target words deleted from context, were circulated online to English native speakers in order to intuitively fill in the gaps with appropriate words. This procedure was termed in previous research as modified cloze procedure (Schwanenflugen and LaCount, 1988 ;

Rayner and Well,1996). A high percentage of agreement on a specific item in a given context would be interpreted as strong predictability for the vocabulary item and a lack of agreement would mean low or zero predictability.

To create a user-friendly cloze task, I worked with a doctoral candidate in computer science engineering to build a web-based interactive survey that can be easily adminstered and analyzed. The story versions were provided in the same format L2 readers saw them but with 121 gaps representing the target tokens in each version. Participants were asked to log in one version only. To guarantee maximum responses, every chapter was presented in a single web page with a submit button at the bottom so that once a respondent submits a chapter, the answers get recorded to the server. An incentive of a $50 drawing was announced to encourage more respondents. The survey was open for three weeeks then closed for data analysis. A sample snapshot of the survey is shown in Appendix I.

A total of 136 entries were recorded in the server for the whole survey. All respondents were undergraduate and graduate native speakers of English at Michigan State University. After cleaning procedures and exclusion of blank entries and non-native respondents in both versions of the story, a sample of 108 valid entries were considered (56 in version A and 52 in version B). The output was orgnaized around target words with each column recording the entries for a specific gap in the text. The predictability is calculated as the proportion of correct answer over the total number of responses for an item. In previous literature, an agreement percentage of 78% - 100 % was considered for high predictability, 55 % - 77 % for medium and 0 % - 54 % for low predictability. For the purpose of the current study, the predictability values were entered as a numerical variable to measure for the role of context in online processing and vocabulary learning.

2.5 Analyses

2.5.1 Definition of variables

I distinguish between online and offline effects on vocabulary outcomes. Online variables refer to the information in eye movement records that includes early measures of processing (e.g.; first fixation and first pass time) and late processing measures (e.g.; gaze duration and total time). First fixation duration captures the time of the first look at the target area (for example, a novel vocabulary word) when encountered for the first time during forward reading. Gaze duration combines first fixation duration along with any other fixation made on the target area at the initial visit before the eyes move forward or backward to the next target area. Total reading time is the sum of all fixation durations on the target area (see Winke, Godfroid, & Gass, 2013). I also report skipping rates, regressions-in and regressions-out of the interest areas. Regressions-in refer to instances when readers returned to a target word after first pass. Regressions-out refer to times when readers went back to a previous part of the sentence on first pass. These processing measures are reported for each of the target tokens as well as summed over vocabulary items to test if eye movement behavior predicts learning outcomes in token-based and item-based analyses.

Offline variables refer to the textual factors of total exposure and predictability. Total exposure is an item-based factor that represents the number of times a vocabulary item was seen in the text. Based on exposure, each item contributed different number of tokens. The instance of meeting a single token was labeled as an 'encounter'. In a similar manner, I distinguish between token-based predictability and item-based predictability. Token-based predictability is the specific predictability score of a given encounter with a word. Item-based predictability is its maximum reported predictability among all its tokens. For example, the pseudo word 'gube'

received a range of predictability scores between 0 and 38.3 over its 18 tokens in version A of the story, and between 5 and 84 in version B. For token-based analysis, all predictability scores of 'gube' were used in the model to predict learning success. In item-based analysis, the item 'gube' was assigned its maximum reported predictability (38.3 in version A and 84 in version B). In this way, it was possible to test the effect of predictability at each encounter and also test if readers exhibited different learning patterns vocabulary items based on their predictability levels. To further elucidate the role of context in word learning, I categorized item maximum predictability into two levels: predictable and less predictable. Previous literature has defined a range around 78 % as a cutoff for high predictability (Ehrlich, & Rayner, 1981; Rayner, & Well, 1996; Schwanenflugen, & LaCount, 1988). Based on the distribution of predictability data in the study, I set a cutoff point of 77 %, yielding equal number of items in predictable and less predictable categories.

Because there are two versions of the reading material where target and control words were counterbalanced, each participant contributed reading times to two conditions: experimental and control. The factor of condition was used to describe differences in processing measures and reading behavior between the target words (pseudo words) and control words (familiar English words). Individual factors included learners' vocabulary size, L2 proficiency, reading speed and reading comprehension scores. Vocabulary outcomes, scored as 0 or 1 for each item, represented three categorical dependent variables in the statistical models, one for form recognition, one for meaning recognition and the third for meaning recall. Table 5 briefly outlines and defines the variables and terminology used in presenting the results.

Table 5

Definitions of variables and terminology in the study

Term	Definition
Condition	Whether the word appeared as a pseudo or familiar token
Encounter	Each target or control token in the reading text
First fixation duration (FFD)	The time of the first look at the target word
Gaze duration (GD)	The sum of fixations made on the target word at the initial visit
Interest area(target area)	A word for which eye movement measures were recorded and analyzed
Item predictability	The maximum predictability score for a vocabulary item across all tokens
Online processing measures	Recorded Eye movement times on target and control words
Regression-in	When readers went back to the target word after the first pass
Regression-out	When readers returned to an earlier part of the sentence on first pass
Skipping	The absence of a fixation on a target word at first pass
Summed processing times	The eye movement measures summed for each vocabulary item across all encounters; the cumulative attention measure over specific target words regardless of encounters.
Token predictability	The reported predictability for each token in the text
Total Fixation Duration (TFD)	The sum of all fixation durations on the target word
Total exposure	The number of times each vocabulary item was seen in the text

2.5.2 Data structure

A total of 57 data files were screened and reviewed for errors. Several data files were excluded from the analysis because they showed blank or poor captures of eye movements; that is, irregular and/or incomplete recordings. Therefore, the offline vocabulary tests and other information associated with these recordings were excluded from the data. Forty-two valid

samples were considered for analysis. The data sheet was organized by subjects and item

information. Each subject reported 242 observations, representing the total number of

experimental and control tokens. In this fashion, the layout of the data showed items nested

within subjects, and encounters nested within items. Figure 1 shows an example of this structure

for a given reader in the experiment.
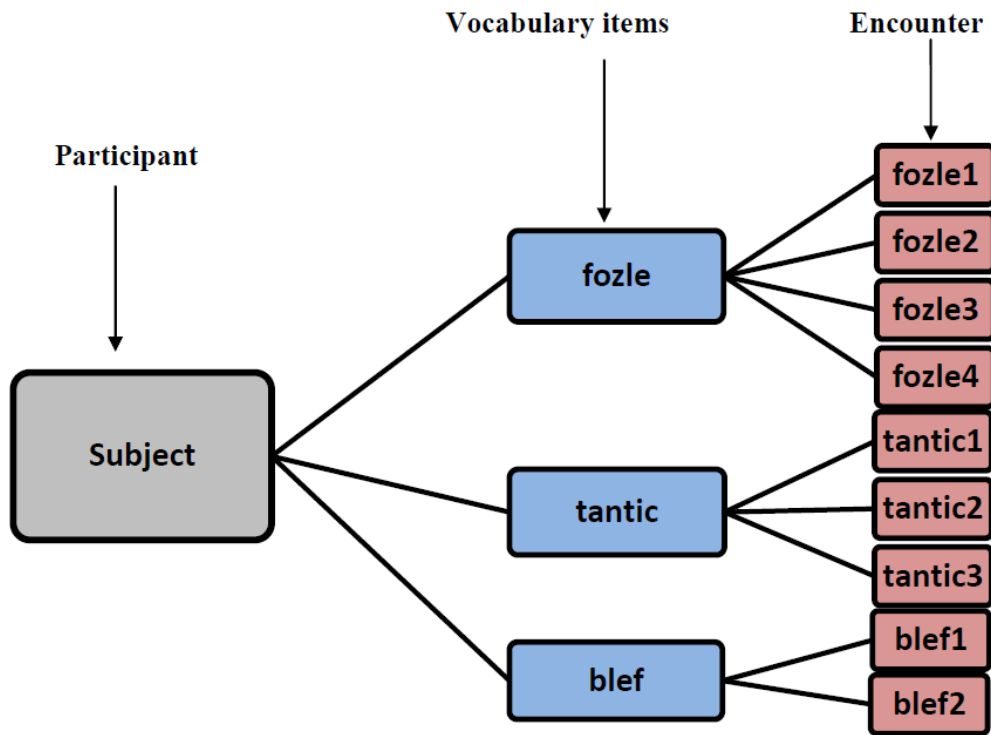
Figure 1. Data structure for participants and target words

Based on this hierarchical structure, I adopted a Generalized Linear Mixed Model

(GLMM) to fit the appropriate regression that can accommodate multiple levels (Heck, Thomas,

& Tabata, 2012). In the light of Figure 1, GLMM is conducted with two levels when we test by

vocabulary item so that the model would only include subject variables and item variable in

repeated measures. The model expands to three levels when we need to test the level of

encounter including information about all the tokens of all items.

2.5.3 Statistical tests

Online reading patterns. Reading patterns were averaged from first encounter to last encounter over all items to investigate how processing times changed from early to repeated exposure to target words. I then examined the role of condition, encounter and predictability in the online reading patterns and reading behavior exhibited by second language readers in the study. Three sets of GLMM were conducted with online processing measures as continuous dependent variables; condition, encounter and predictability as fixed factors; word length as a control variable and subject and items as random factors. Upon inspection, online reading data (first fixations, gaze durations and total times) was not normally distributed and was largely skewed to the positive side, which made the use of a linear regression model inappropriate. One alternative test in this case is Gamma regression, which uses a log link function to fit non-normal positive dependent variables (McCullagh, & Nelder, 1989). Interaction terms were estimated for encounter, predictability and condition under each model. Similar tests were performed to predict the patterns of skipping, regressions and summed reading times on target and control words.

Item-based effects. I presented descriptive statistics for the average vocabulary gains in the three post tests and investigated the role of total exposure and maximum item predictability on learning outcomes through three sets of two-level GLMM to fit a binary logistic regression for each of the dependent variables: form recognition, meaning recognition and meaning recall. Using similar models, the effects of summed processing measures on learning outcomes were also investigated. Total exposure and maximum predictability were entered as covariates in the same model to estimate how offline and online predictors compare in explaining the variance in vocabulary learning.

Token-based effects. Three-level GLMM were used to fit binary logistic regressions to estimate the effect of online reading times for every token in the text on the probability of learning novel words. By combing token-based predictability and online fixation times in a single model, it was possible to examine which of the two was the more important factor and whether they interacted in predicting learning outcomes.

Individual differences. The final part of the results shows descriptive statistics for the reading questionnaire followed by logistic regressions to investigate the role of L2 proficiency, vocabulary size, comprehension and reading speed on incidental learning from reading.

2.5.4 Reporting results

The GLMM output calculates the probability of the incidence of a dependent variable in terms of an odds ratio (OR), quantifying the predicted change in the dependent measure as a function of a one unit increase in a given predictor (Ferguson, 2009). An OR larger than 1 indicates a positive relationship and an OR less than one indicates a negative relationship. The interpretation of OR varies according to the type of the dependent variable. For example, if encounter predicted fixation times (continuous variable) with an OR of 0.25, this would indicate that one additional encounter predicts a percent decrease in fixation times by 75 % ($1 - 0.25 *$ 100%). On the other hand, if repeated exposure predicted form recognition (binary categorical variable) with an OR of 1.75, this would indicate that one extra exposure was associated with an increase in the odds of correct responses in form recognition by 75 % ($1.75 - 1 * 100$ %). In addition to the odds ratio (OR), I report the 95% confidence interval of the effect size of the predictor variable. The predictor was considered significant at the .05 level while the strength of the relationship was interepreted through OR. A strong relationship starts at OR < 0.33 or OR > 3 (Ferguson, 2009; Menard, 2010; Powers, & Xie, 2008).

CHAPTER 3: RESULTS

The results presented in this chapter are organized by research questions. I first compare token-based and item-based online reading measures in both the control and experimental condition and investigate how encounter and predictability in increase or decrease reading times. In the second part, I provide descriptive statistics for vocabulary gains and estimate the effects of token-based online processing and token-based predictability on learning outcomes. In the third part, I explain the role of summed processing measures and compare their effects with those of offline textual factors; i.e., item predictability and total exposure. Finally, I present findings regarding the role of individual differences in incidental learning from L2 reading.

3.1 Encounter and predictability data

To retrieve the predictability data, I calculated respondents' percentage of agreement for each token in the cloze procedure of the norming study (see section 2.4.4). These percentages were entered as numerical values between 0 and 100 to represent token predictability in the model. For item-based analyses, the highest predictability score for each target word was assigned as that word's predictability score. Table 6 shows item-based information about number of exposures and highest predictability for target words in the two versions of the text. A detailed information about token predictability in the two versions of the story are provided in detail in Appendix J.

Table 6

Encounter and predictability data for target vocabulary

Number of encounters	Pseudo word	Meaning in version A	Maximum item predictability	Meaning in version B	Maximum item predictability
30	fozle	hotel	90	table	96
18	gube	café	38.3	room	84
10	mave	face	85	desk	75
9	tund	stop	90	meet	94.5
7	leam	tall	60	busy	56.3
6	blef	kill	77.5	push	88
6	toker	party	27.5	money	82
5	bannow	pocket	77.5	window	72.7
5	mot	bag	85	gun	77.1
4	fonteen	picture	90	airport	77.3
4	dangy	quiet	62.5	happy	72.9
4	windle	garden	5	letter	81.3
3	neech	shirt	60	dress	44
3	redaster	accident	97.5	hospital	54.5
2	dook	rich	72.5	cold	72.7
1	tance	sleep	65	drink	59
1	pamery	cinema	71.7	camera	77
1	tantic	famous	10	hungry	37.5
1	dorch	plane	57.7	noise	45.8
1	smick	chair	72.5	shoes	50

3.2 Online reading patterns

Before considering the relationship between online processing and vocabulary learning, I needed to investigate input and contextual factors that influenced reading patterns and lexical processing. To test these effects, online reading measures were entered as continuous dependent variables in a GLMM to run a Gamma regression analysis with condition, encounter and token predictability as predictors and word length as a control variable.

The Gamma regression is an alternative to linear regression which uses a log link function to fit non-normal positive dependent variables (Heck, Thomas, & Tabata, 2012; McCullagh, & Nelder, 1989). The beta coefficient for gamma regression does not provide meaningful interpretation unless odds ratios are calculated. The quotient of odds ratio in a significant relationship is interpreted as a percent change in the incident rate of the continuous outcome either negatively or positively. For each processing measure, I created a line graph and a scatter plot to identify decreases and major cutoffs, if any, in fixation times. When cutoff points were observed, follow up analyses were made to explain any discrepancies between early encounters and late encounters with target words.

3.2.1 First fixation durations

Figure 2 shows that first fixation durations on target words started at an average of 264 ms (SD = 124) and ended with 215 ms (SD = 88) while first fixations on control words started at 227 ms (SD = 86) and ended at 218 ms (SD = 88). Visually, there were no major changes in fixation times from first to last encounter or major differences between conditions. Table 7 summarizes the regression output for the effect of text-based factors on first fixation durations.

Table 7

Effects of text-based factors on first fixation durations (FFD)

	Odds	OR	95% CI		p
Intercept	5.022		4.82	5.22	< .001 ***
Condition		1.08	1.04	1.12	< .001 ***
Encounter		0.98	0.96	0.99	< .001 ***
Predictability		0.89	0.82	0.91	.001 **
Encounter * Predictability		0.995	0.991	0.999	.008 **
Condition * Encounter		1.01	1.005	1.03	< .001 ***
Condition * Predictability		0.89	0.84	0.95	.001 **

Note: The (*) marks signify the level of significance of the *p* value

Regression output showed that condition significantly predicted first fixation (*OR* = 1.08, 95% CI = [1.04, 1.12], *p* < .001). Comparing the odds ratio against the odds of the intercept, fixation times in the experimental condition was significantly longer than in the control condition and that the probability of fixating longer on a target words increases by about 2 % when the word is unfamiliar. Encounter was slightly associated with a decrease in FFD (*OR* =0.98, 95% CI = [0.967, 0.993], *p* < .001), implying that adding more encounters was associated with a negative change in first fixation durations by a factor of 1 %. There was a significant interaction between encounter and condition (*OR* = 1.01, 95% CI = [1.005, 1.03], *p* < .001), which implied that target and control words started to behave similarly as encounters increased. Token-based predictability was negatively associated with FFD (*OR* = 0.89, 95% CI = [0.82, 0.91], *p* = .001), suggesting that an increase in the token predictability resulted in a negative change in first fixation durations by 11 %. The negative effect for an interaction between encounter and predictability (*OR* =0.995,

95% CI = [0.991, 0.999], p = .008) suggested that the effect of predictability was slightly more

evident by later encounters.
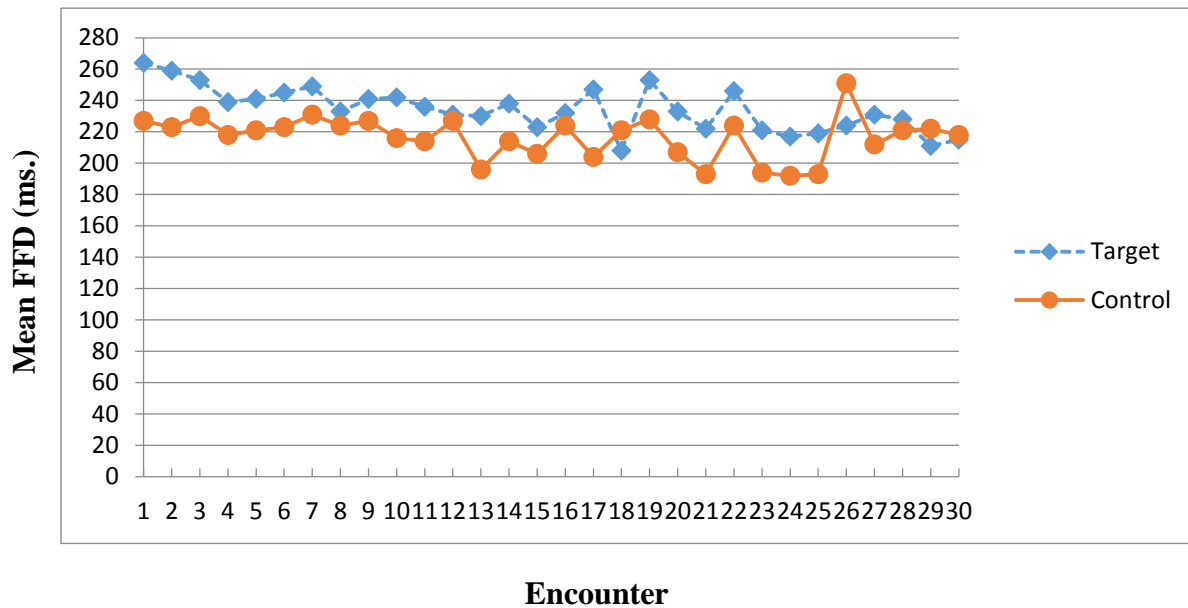
Figure 2. Mean fixation times (in milliseconds) on target and control words by encounter

There was also a small negative interaction between condition and predictability (*OR*

=0.89, 95% CI = [0.84, 0.95], p = .001), which suggested that the effect of predictability may

have been less pronounced for the pseudo words in the text. To visualize the interaction between

predictability and condition, I categorized token predictability into predictable and less

predictable bands based on a cutoff point of 77 %. I then created a graph for mean first fixation

durations in both target and control conditions with separate lines for predictability bands to

investigate how predictability influenced fixation times. Figure 3 shows that less predictable

tokens received more fixation times and that the effect of predictability in reducing the amount

of processing time was slightly more pronounced for familiar words than for pseudo words. The

scatter plot for first fixation durations across encounters did not show major cutoff points.

Accordingly, no follow up analyses were done for specific encounters.
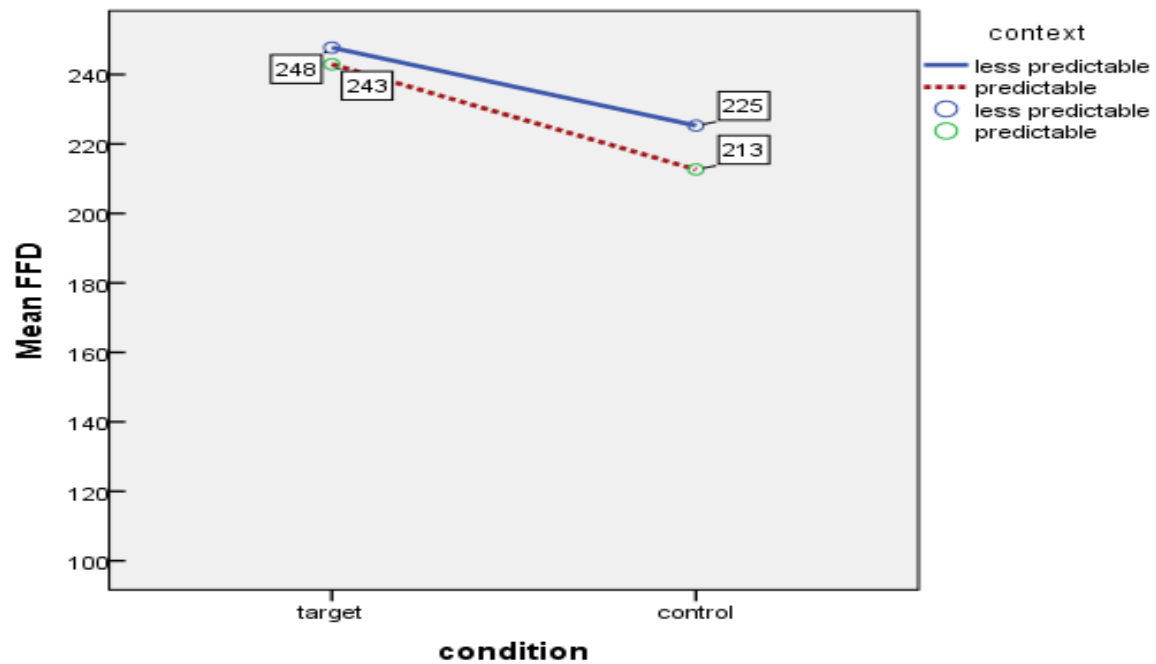
Figure 3. The interaction of condition and predictability in first fixation durations (FFD)
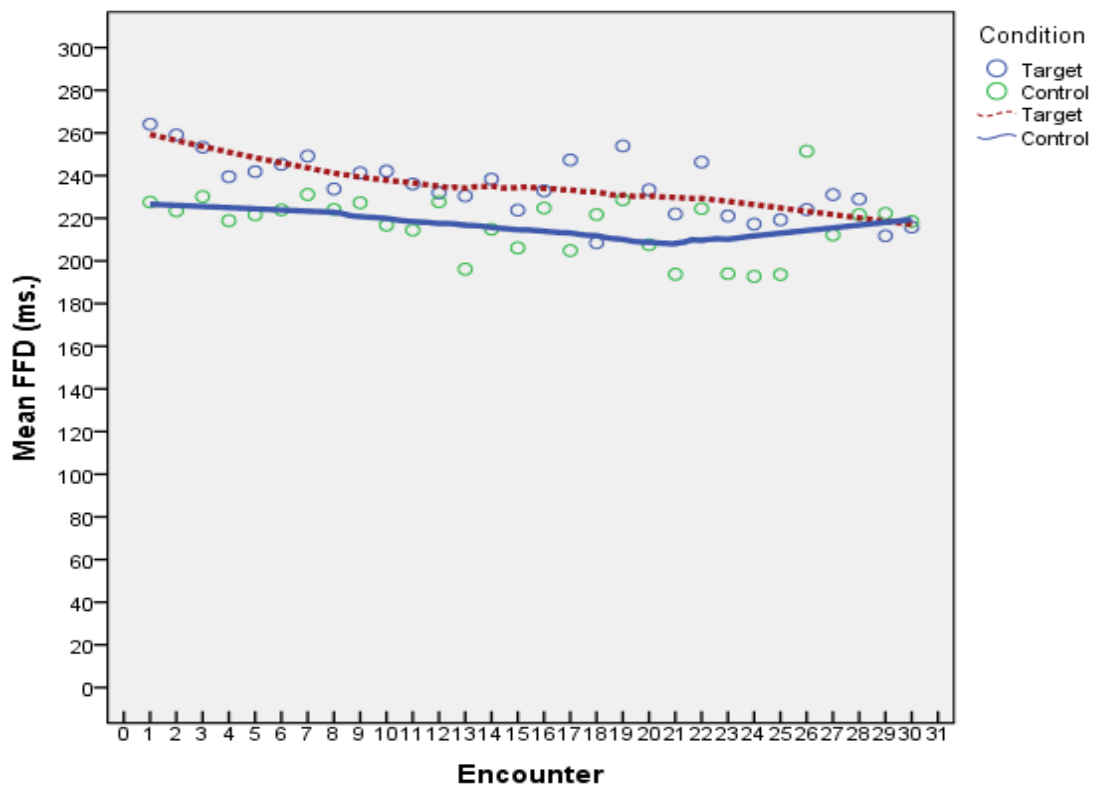
Figure 4. Scatter plot for mean first fixation durations by encounter and condition

3.2.2 Gaze durations

Figure 5 shows that gaze durations (GD) on target words started at 393 ms (*SD* = 282) and ended at 237 (*SD* = 118). A scatter plot was created to identify cutoff points.
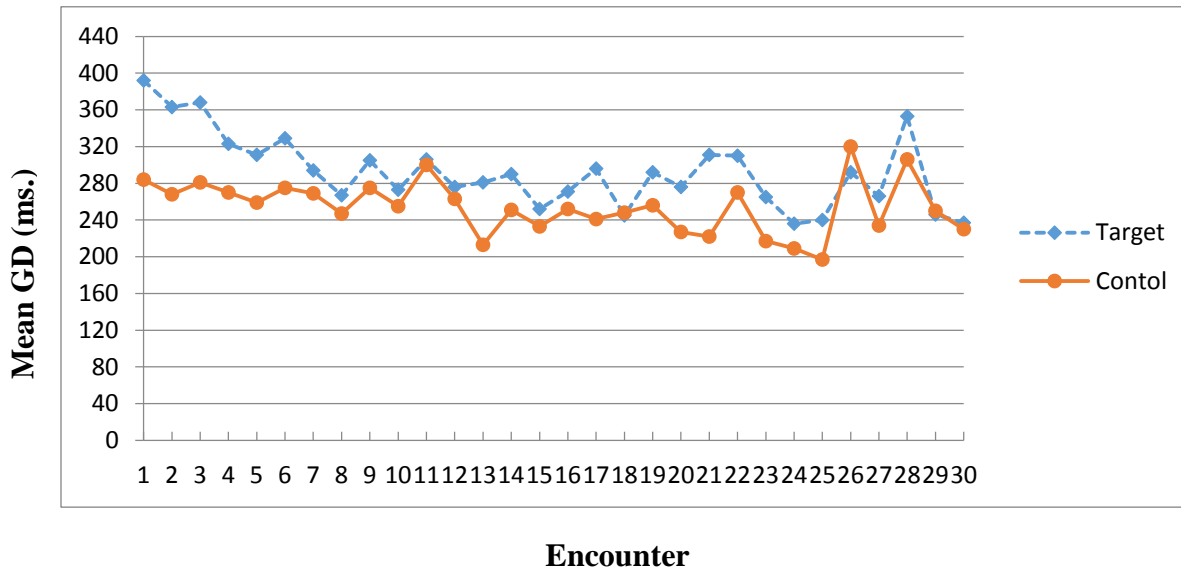
Encounter

Figure 5. Mean gaze durations (in milliseconds) on target and control words by encounter
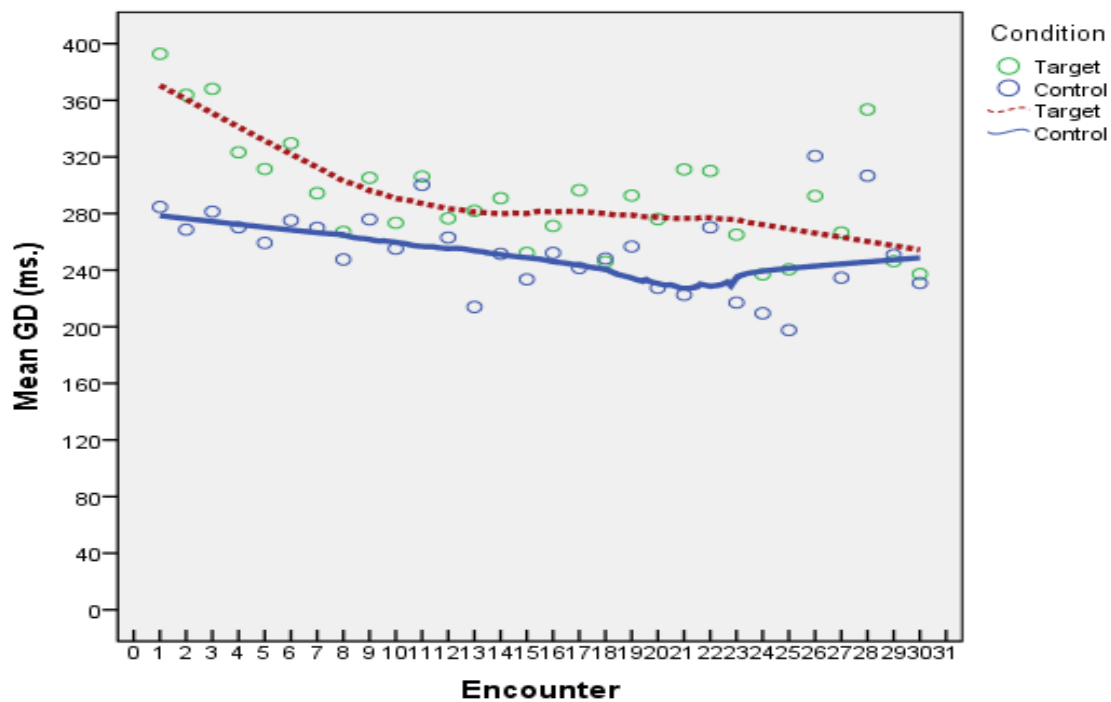
Figure 6. Scatter plot for gaze durations by encounter and condition

49

The scatter plot in Figure 6 shows a larger decrease until encounter 12, after which there was a slow but steady decrease until the last exposure. Table 8 summarizes these textual factors on the patterns of gaze durations.

Table 8

Effects of text-based factors on gaze durations (GD)

	Odds	OR	95% CI		p
Intercept	4.18		3.84	4.53	< .001 ***
Condition		1.24	1.17	1.32	< .001 ***
Encounter		0.95	0.94	0.97	< .001 ***
Predictability		0.91	0.85	0.96	.003 **
Encounter * Predictability		0.99	0.97	1.01	.172
Condition * Encounter		1.04	1.01	1.10	< .001 ***
Condition * Predictability		0.93	0.89	0.95	.045 *

Note: The (*) marks signify the level of significance of the *p* value

In the light of Table 8, condition was a significant predictor of the variance in gaze durations ($OR = 1.24$, 95% CI = [1.17, 1.32], $p < .001$). Comparing the odds of the intercept with odds ratio, the probability of longer gaze durations on pseudo words increased by around 16 % when target words were unfamiliar. Each additional encounter predicted a decrease in gaze duration by about 5 % ($OR = 0.95$, 95% CI = [0.94, 0.97], $p < .001$). There was a small interaction between encounter and condition ($OR = 1.04$, 95% CI = [1.01, 1.10], $p < .001$). The increase in token predictability was associated with a decrease in gaze durations ($OR = 0.91$, 95% CI = [0.85, 0.96], $p = .003$). The interaction between condition and predictability was

significant ($OR = 0.93$, 95% CI = [0.89, 0.95], $p = .045$). Figure 7 illustrates this interaction

showing that predictability had almost equal effects on both target and control words.
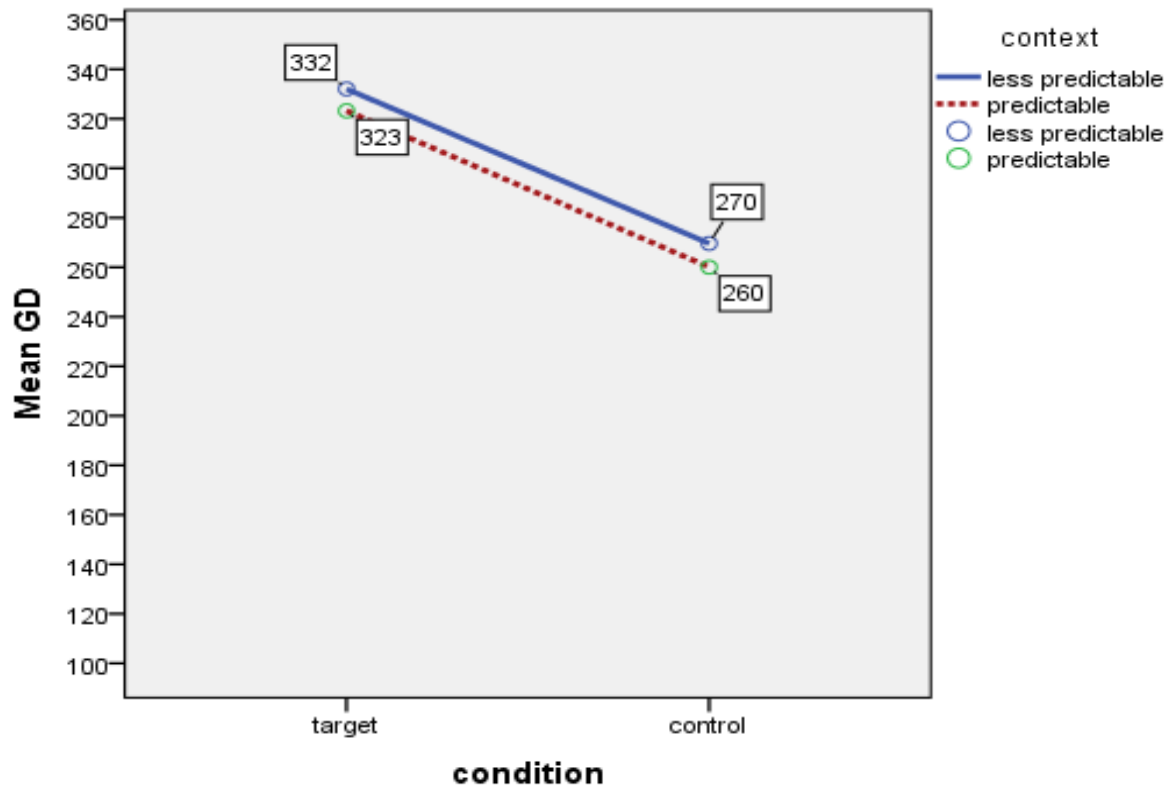
Figure 7. The interaction of condition and predictability in gaze durations (GD)

Based on the scatter plot in Figure 6, follow up analyses were conducted for early

encounters (1-12) and late encounters (13-30). Results indicated that the effect of condition was

larger in early encounters ($OR = 1.41$, 95% CI = [1.32, 1.65], $p < .001$) than in later encounters

($OR = 1.09$, 95% CI = [1.02, 1.12], $p = .035$), suggesting that target and control words started to

be read generally fast after 12 encounters. Encounter was only significant in the first 12

exposures ($OR = 0.84$, 95% CI = [0.81, 0.89], $p < .001$), confirming the visual observation that

the decrease was larger and more significant than in later encounters. On the other hand, token

predictability was significant only in later exposures ($OR = 0.87$, 95% CI = [0.82, 0.94], $p <$

$.001$) and not in early encounters ($OR = 0.97$, 95% CI = [0.96, 1.002], $p = .152$).

3.2.3 Total reading times

Figure 8 shows that total reading times recorded highest on the first encounter of target words ($M = 702$ ms, $SD = 512$) and lowest by the final encounter ($M=265$ ms, $SD=130$).
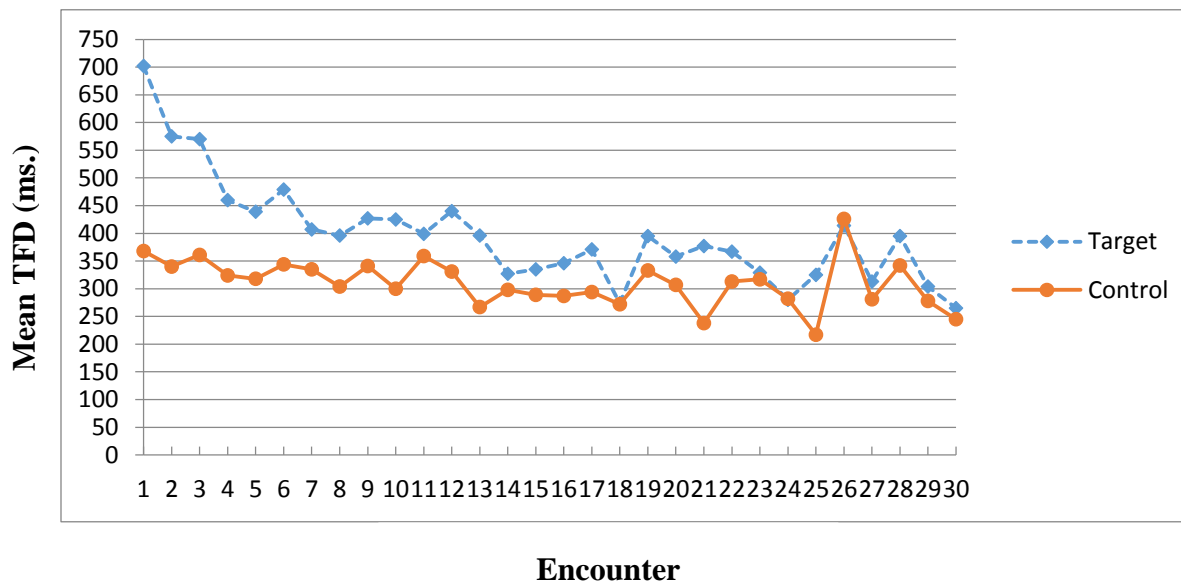
Encounter

Figure 8. Mean total reading times (in milliseconds) for target and control words by encounter
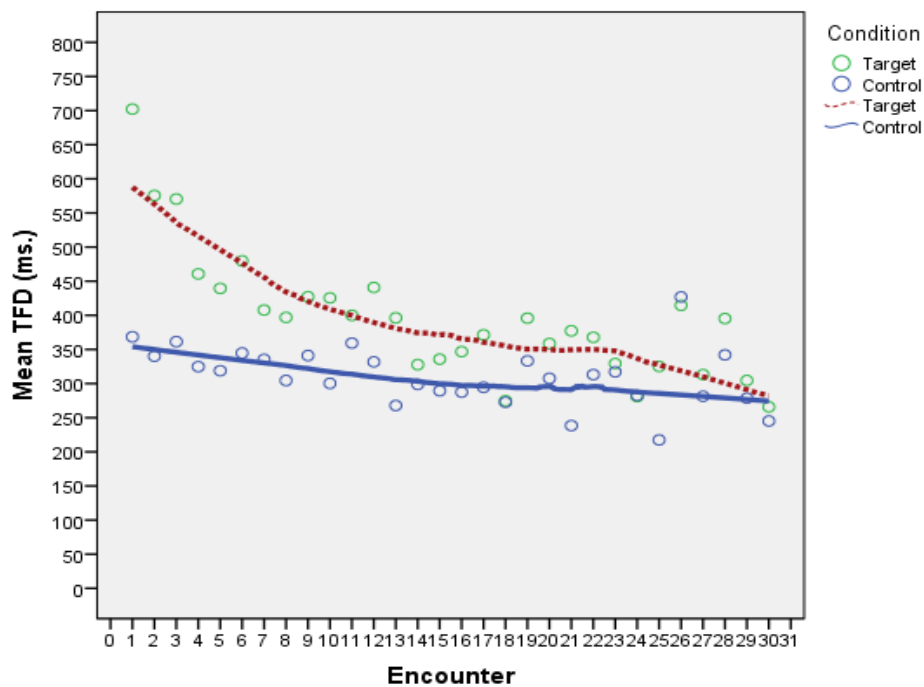
Figure 9. Scatter plot for mean total durations by encounter and condition

The scatter plot in Figure 9 shows that the decreasing pattern was more evident until encounter 11. Later encounters from 12 to 23 showed slower decline in reading times, after which the line dropped steadily until the last exposure. Table 9 summarizes main textual effects on total reading times.

Table 9

Effects of text-based factors on total reading times (TFD)

	Odds	OR	95% CI		*p*
Intercept	3.30		2.88	3.73	< .001 ***
Condition		1.68	1.42	1.77	< .001 ***
Encounter		0.88	0.87	0.91	< .001 ***
Predictability		0.93	0.88	0.95	.001 **
Encounter * Predictability		0.988	0.981	0.995	< .001 ***
Condition * Encounter		1.020	1.016	1.023	< .001 ***
Condition * Predictability		0.85	0.77	0.93	.001 **

Note: The (*) marks signify the level of significance of the *p* value

Table 9 shows that condition was a significant predictor of the variance in total reading times ($OR = 1.68$, 95% CI = [1.42, 1.77], $p < .001$), indicating that readers took longer times on target words than known words. Each additional encounter was associated with a decrease in total times ($OR = 0.88$, 95% CI = [0.87, 0.91], $p < .001$), which was modulated by a small interaction between encounter and condition ($OR = 1.020$, 95% CI = [1.016, 1.023], $p < .001$). There was a significant association between token predictability and the decrease in total times ($OR = 0.93$, 95% CI = [0.88, 0.95], $p = .001$), which was modulated by a small interaction

between encounter and predictability ($OR = 0.988$, 95% CI = [0.981, 0.995], $p < .001$) and between condition and predictability ($OR = 0.85$, 95% CI = [0.77, 0.93], $p = .001$). This interaction is illustrated in Figure 10, which shows that the effect of predictability was slightly more pronounced on familiar words rather than on pseudo words.
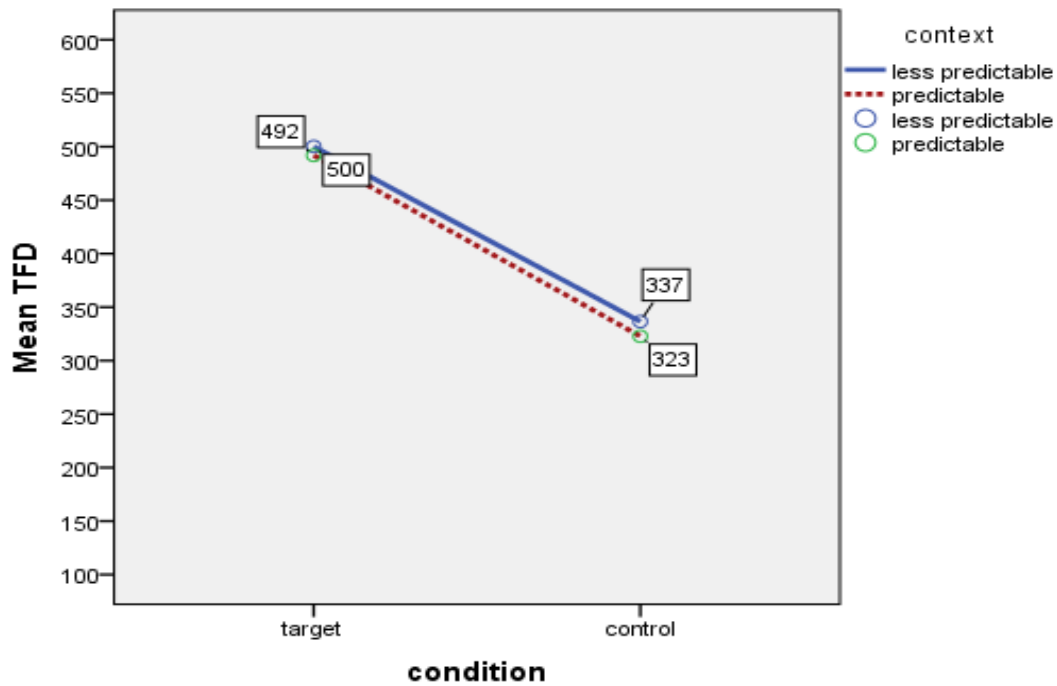
Figure 10. The interaction of condition and predictability in total fixation durations (TFD)

Based on the scatter plot for total times in Figure 9, follow up analyses were conducted for early encounters (1-11) and late encounters (12-30) separately. Results showed that the effect of condition was larger in early encounters ($OR = 1.85$, 95% CI = [1.75, 1.96], $p = .001$) than late encounters ($OR = 1.44$, 95% CI = [1.21, 1.74], $p = .013$). The decrease in total times was also greater in early encounters ($OR = 0.75$, 95% CI = [0.71, 0.80], $p = .001$) than in later ($OR = 0.94$, 95% CI = [0.90, 0.98], $p = .005$). The effect of token predictability was only significant in late encounters ($OR = 0.87$, 95% CI = [0.83, 0.92], $p = .001$) but not on early encounters ($OR = 0.96$, 95% CI = [0.91, 1.01], $p = .189$).

3.2.4 Reading behavior

Other aspects of reading behavior such as skipping and regressions contributed to the patterns of total times. Skipping rates refer to the instances when interest areas were skipped on first pass. Skipping was more frequent in the control condition than in the experimental condition (around 21% of target occurrences and almost 26% of control occurrences).

Regressions-in refer to the instances when readers regressed to the target or control word after the first pass. Readers returned more to target items (almost 25%) than to control items (almost 15%). Regression-out refers to the instance when readers launched a regression from the target or control word after first pass. Regressions-out occurred on almost 27 % of target observations and on 22 % of control observations. Table 10 summarizes these reading behavior patterns on both target and control conditions. Because skips and regressions are binary data, logistic regression analyses were performed as the next step to identify what factors predicted their occurrence.

Table 10

Mean percentages of skipping and regressions (%) on target and control words

	Skipping rate	Regression in	Regression out
Target words	21.3 (40.9)	24.6 (43.1)	26.7 (44.2)
Control words	25.8 (43.7)	14.07 (34.8)	22.4 (41.7)
Total	23.5 (42.4)	19.04 (39.5)	24.6 (43.05)

Skipping. Condition was a significant predictor of skipping ($OR = 0.82$, 95% CI = [0.71, 0.95], $p = .007$), meaning that the odds of skipping decreased by around 18 % when the target words were familiar. The effect of encounter was not significant ($OR = 0.96$, 95% CI = [0.90, 1.03], $p = .273$). Token predictability showed to be a significant predictor of skipping ($OR =$

1.19, 95% CI = [1.03, 1.39], p = .001), suggesting that higher predictability triggered more skipping. No interaction was found between encounter and condition (OR = 0.97, 95% CI = [0.95, 1.02], p = .582) or between condition and predictability (OR = 1.002, 95% CI = [0.98, 1.003], p = .174).

Regressions-in. Condition was a strong predictor of regression-in rates (OR = 2.79, 95% CI = [2.42, 3.22], p < .001), which indicated that the odds of regressing-in significantly increased by 2.79 times when the target was unfamiliar. Each additional encounter decreased the odds of regressing to the target word by about 28 % (OR = 0.72, 95% CI = [0.67, 0.77], p < .001), implying that regressions-in were more frequent in initial encounters. Token predictability did not have a significant effect on regressions-in (OR = 0.88, 95% CI = [0.72, 1.09], p = .251). There was an interaction between encounter and condition (OR = 0.90, 95% CI = [0.86, 0.94], p < .001) suggesting that the rates of regressions-in become similar for target and control words by later encounters. There was no interaction between encounter and predictability (OR = 1.007, 95% CI = [0.98, 1.03], p = .622) or condition and predictability (OR = 0.98, 95% CI = [0.96, 1.09], p = .351).

Regressions-out. Condition was a significant predictor of regression-out rates (OR = 1.21, 95% CI = [1.09, 1.34], p < .001), which shows that the odds of regressing-out increased by about 21% when the target was unfamiliar. Each additional encounter decreased the odds of regressing out of the interest area by 2 % (OR = 0.98, 95% CI = [0.97, 0.99], p = .010). Token predictability was not a significant predictor of regression out (OR = 0.99, 95% CI = [0.84, 1.17], p = 0.971). No further interactions were found significant between predictability, condition and encounter. Table 11 summarizes the roles of textual factors on skipping and regressions.

Table 11

Effects of text-based factors on skipping and regression rates

	Skipping		Regression-in		Regression-out	
	OR	*p*	*OR*	*p*	*OR*	*p*
Condition	0.82	.007 **	2.79	< .001 ***	1.21	<.001***
Encounter	0.96	.273	0.72	< .001 ***	0.98	.010 *
Predictability	1.19	.001**	0.88	.251	0.99	.971
Encounter * Predictability	0.99	.362	1.007	.622	1.001	.521
Condition * Encounter	0.98	.582	0.90	< .001***	0.96	.187
Condition * Predictability	1.002	.174	0.98	.351	0.99	.231

Note: The (*) marks signify the level of significance of the *p* value

3.2.5 Summed reading times

Summed processing measures are the sum of all times spent on a given item over all its encounters. Because repetition generally invites more reading, I arranged mean summed fixation measures by exposure bands (the number of times a word was seen). Items with a higher frequency of occurrence in the text generated higher summed times and that pseudo words received more attention than control words. Table 12 outlines the average summed fixation measures on all pseudo and control words by exposure bands.

Summed first fixations were significantly different by condition ($OR = 1.14$, 95% CI = [1.11, 1.18], $p < .001$) and were significantly predicted by exposure band ($OR = 1.09$, 95% CI = [1.03, 1.12], $p < .001$). Controlling for exposure, maximum item predictability was associated with a decrease in the summed FFD ($OR = 0.88$, 95% CI = [0.81, 0.91], $p < .001$). No significant interactions were found between exposure and condition or exposure and predictability.

57

Table 12

Mean summed fixation measures (in milliseconds), with SD in parentheses, by exposure bands

Exposure band	Summed FFD		Summed GD		Summed TFD	
	target	control	target	control	target	control
1	347 (952)	331(622)	460 (769)	410 (780)	748 (695)	543 (607)
2	419 (623)	357 (126)	491 (198)	457 (292)	891 (480)	762 (705)
3	734 (179)	580 (252)	1344 (846)	844 (518)	2282 (1411)	1092 (752)
4	934 (251)	814 (206)	1477 (651)	1037 (428)	2329 (1120)	1255 (534)
5	1174 (243)	946 (271)	1544 (546)	1172 (354)	2249 (997)	1396 (448)
6	1443 (459)	1191 (281)	1974 (720)	1415 (373)	3179 (993)	1848 (711)
7	1593 (367)	1362 (237)	2060 (556)	1507 (334)	2974 (933)	1985 (776)
9	2063 (500)	1808 (416)	2795 (804)	2212 (671)	4873 (1834)	2932 (1037)
10	2281 (392)	1993 (448)	3002 (717)	2443 (608)	4798 (952)	3123 (854)
18	3793 (866)	3098 (877)	4550 (1310)	3527 (1057)	6121 (843)	4289 (1433)
30	6109 (991)	5388 (987)	7676 (2551)	6291 (1815)	10220 (634)	7766 (2854)
Mean	1375 (1325)	1222 (156)	1852 (3073)	1473 (2470)	2801 (3152)	1864 (3287)

Summed gaze durations were significantly predicted by condition ($OR = 1.27$, 95% CI = [1.19, 1.34], $p < .001$) and exposure band ($OR = 1.07$, 95% CI = [1.01, 1.11], $p < .001$). Item predictability was associated with a decrease in summed gaze durations ($OR = 0.95$, 95% CI = [0.87, 0.99], $p = .019$). No significant interactions were found between exposure and condition or exposure and predictability.

Summed reading times were significantly different by condition ($OR = 1.49$, 95% CI = [1.41, 1.57], $p < .001$) and exposure band, ($OR = 1.12$, 95% CI = [1.09, 1.17], $p < .001$). This

confirms the fact that pseudo words took more processing times than control words regardless of

the number of exposures. Item predictability was not a significant predictor of summed reading

times ($OR = 0.99$, 95% CI = [0.95, 1.02], $p = .581$). A significant interaction was found between

total exposure and condition with a small effect ($OR = 0.99$, 95% CI = [0.98, 0.99], $p = .001$).

Table 13 summarizes the role of text-based factors in summed processing measures.

Table 13

Effects of text-based factors on summed processing times

	Summed FFD		Summed GD		Summed TFD	
	OR	*p*	*OR*	*p*	*OR*	*p*
Condition	1.14	< .001**	1.27	< .001***	1.49	< .001***
Encounter	1.09	< .001**	1.07	< .001***	1.12	< .001 ***
Predictability	0.88	< .001**	0.95	.019 *	0.99	.581
Encounter * Predictability	0.99	.362	0.99	.522	0.99	.001 **
Condition * Encounter	0.98	.182	0.91	.145	0.96	.187
Condition * Predictability	1.01	.274	0.98	.131	0.99	.231

Note: The (*) marks signify the level of significance of the *p* value

3.2.6 Interim summary

Online reading patterns showed a clear distinction between the processing of pseudo words and known words in context. Condition was significant on all reading measures, showing that readers looked longer and spent more time processing unfamiliar words. It was also shown that pseudo words invited less skipping and more regressions than known words. There was evidence of a growing sense of familiarity with target words as additional encounters were associated with shorter reading times and less regressions. Line graphs of gaze duration and total reading times (Figures 6 and 9) demonstrated that readers dwelled more on early encounters until around exposures (11-13), after which the decrease in fixation times became slower. The difference between conditions was larger in early encounters, and the data suggests that target and control words started to behave similarly after encounters 12-13.

Token predictability was generally associated with shorter reading times and more skipping of target items. This effect became more important in late encounter than in early encounters. This may imply that predictability started to play a role later when pseudo words became better integrated in the sentence structure. The interaction between condition and predictability pointed to the fact that context effects, though significant, were less pronounced on pseudo words than on control words, implying that word familiarity may interfere with predictability in real time processing. Repeated exposure generated higher summed processing times with a significant effect of condition in terms of total times spent on vocabulary items. Predictability was associated with reduced first fixation and gaze duration times.

3.3 Vocabulary knowledge gains from reading

3.3.1 Descriptive statistics

In overall vocabulary measures participants reported the highest gains in form recognition, followed by meaning recognition and finally meaning recall. Table 14 indicates that participants were able to retain the forms of an average 42 % of target words while they recognized the meanings of 30 % of the words and recalled the meanings of only 13 % of the same target items.

Table 14

Average word gains, with standard deviations in parentheses, for the vocabulary post tests

Test	*M* (*SD*)	Percentages (%)	Minimum	Maximum
Form recognition	8.36 (3.16)	41.8	1 (5%)	16 (80%)
Meaning Recognition	6.06 (3.27)	30.3	1 (5 %)	13 (65%)
Meaning recall	2.59 (2.32)	12.9	0 (0 %)	8 (40 %)

To investigate the effect of amount of exposure on vocabulary learning, I analyzed participants' responses by exposure bands (refer to Table 6) to estimate how many hits (correct responses) each item received from participants in each test. Figure 11 reveals a wide difference between highest and lowest exposure bands but variable patterns were noted for middle bands particularly in meaning recognition and recall.

To elucidate the role of context in word learning, I categorized maximum item predictability into two levels: predictable and less predictable based on a cutoff point of 77 % Figure 12 shows the average percentages of vocabulary gains by context type, indicating that context richness increased chances of learning words in all vocabulary tests in a similar manner.
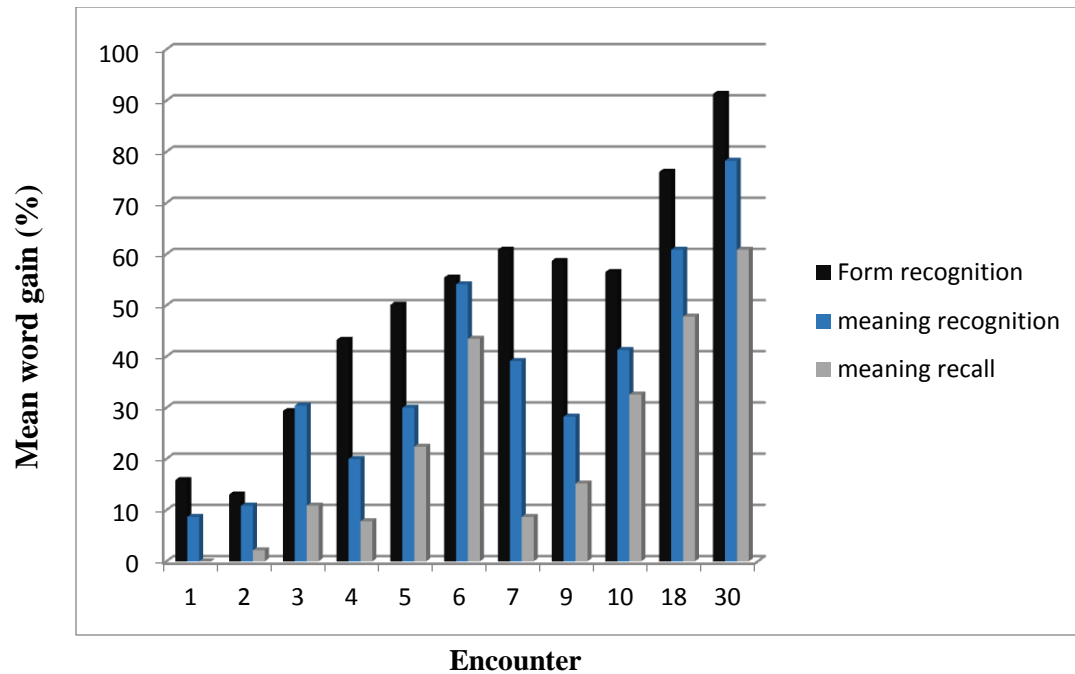
Figure 11. Mean percentages of vocabulary gains in the vocabulary posttests by exposure bands
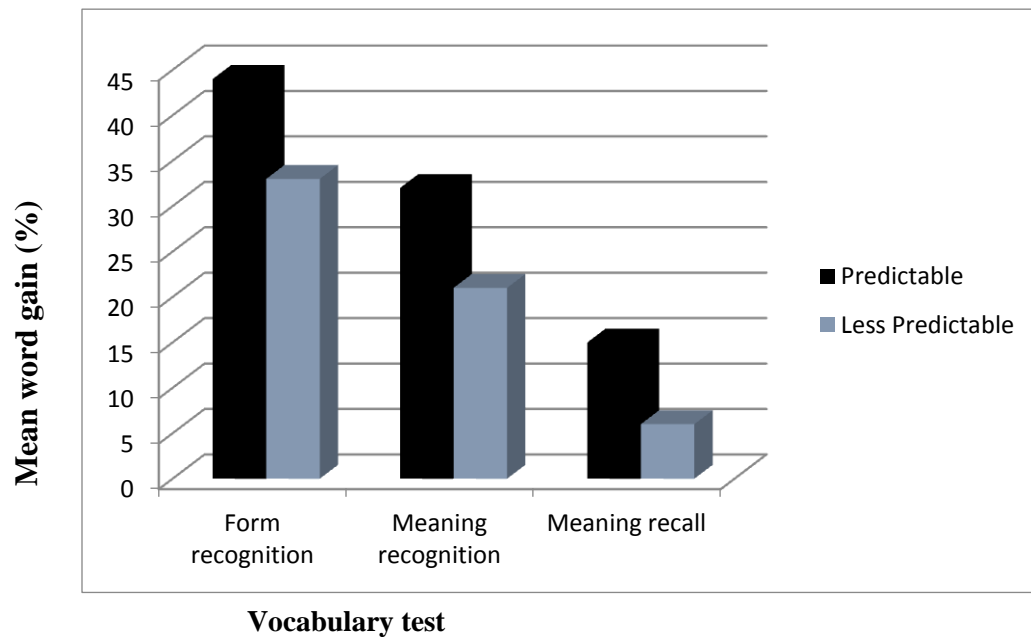
Figure 12. Mean percentages of word gains by context type

3.3.2 Text-based characteristics and vocabulary learning

To explain the variance in learning outcomes based on text-based factors, I looked at how item exposure and maximum item predictability contributed to vocabulary learning. Because the vocabulary variables are binary, I fitted a logistic regression using a two-level GLMM for every vocabulary test. Controlling for item effects and word length, logistic regression output showed that total exposure was a significant predictor for all the vocabulary outcomes but to somewhat different degrees: form recognition (OR = 1.21, 95% CI = [1.05, 1.40], p = .010), meaning recognition (OR = 1.29, 95% CI = [1.15, 1.44], p < .001), and meaning recall (OR = 1.42, 95% CI = [1.27, 1.61], p < .001). By comparing the odds ratios with the odds of the intercept in the three models, we calculate the difference between the probability of learning outcomes and the baseline probability of the intercept [OR * odds/ (odds+1)]. Regression output (Tables 15-17) indicated that each additional exposure increased the probability of form recognition by around 2 %, meaning recognition by around 3 % and meaning recall by 2 %.

Item predictability was most strongly associated with meaning recall (OR = 1.63, 95% CI = [1.36, 1.95], p < .001) followed by meaning recognition (OR = 1.24, 95% CI = [1.08, 1.42], p = .002) yet it did not have a significant relationship with form recognition. Tables 15 through 17 summarize these effects showing positive effects for both exposure and predictability. However, the interaction between exposure and predictability yielded odds ratios < 1, implying a negative impact on meaning recognition and meaning recall although ratios were very close to 1 as shown in the tables. Figures 13 through 15 illustrate the interacting effects of context and repetition on vocabulary gains.

Table 15

Regression output for the effects of exposure and predictability on form recognition

	Odds	OR	95% CI		p
Intercept	0.13		0.054	0.30	$< .001$ ***
Total Exposure		1.21	1.05	1.40	.010 **
Item predictability		1.11	0.99	1.24	.691
Exposure * predictability		0.99	0.98	1.02	.874

Note: The (*) marks signify the level of significance of the *p* value

Table 16

Regression output for the effects of exposure and predictability on meaning recognition

	Odds	OR	95% CI		p
Intercept	0.058		.016	0.20	$< .001$ ***
Total Exposure		1.29	1.15	1.44	$< .001$ ***
Item predictability		1.24	1.08	1.42	.002 **
Exposure * predictability		0.98	0.97	0.99	.012 *

Note: The (*) marks signify the level of significance of the *p* value

Table 17

Regression output for the effects of exposure and predictability on meaning recall

	Odds	OR	95% CI		p
Intercept	.012		.002	.068	$< .001$ ***
Total Exposure		1.43	1.27	1.61	$< .001$ ***
Item predictability		1.63	1.36	1.95	$< .001$ ***
Exposure * predictability		0.97	0.96	0.99	.002 **

Note: The (*) marks signify the level of significance of the *p* value
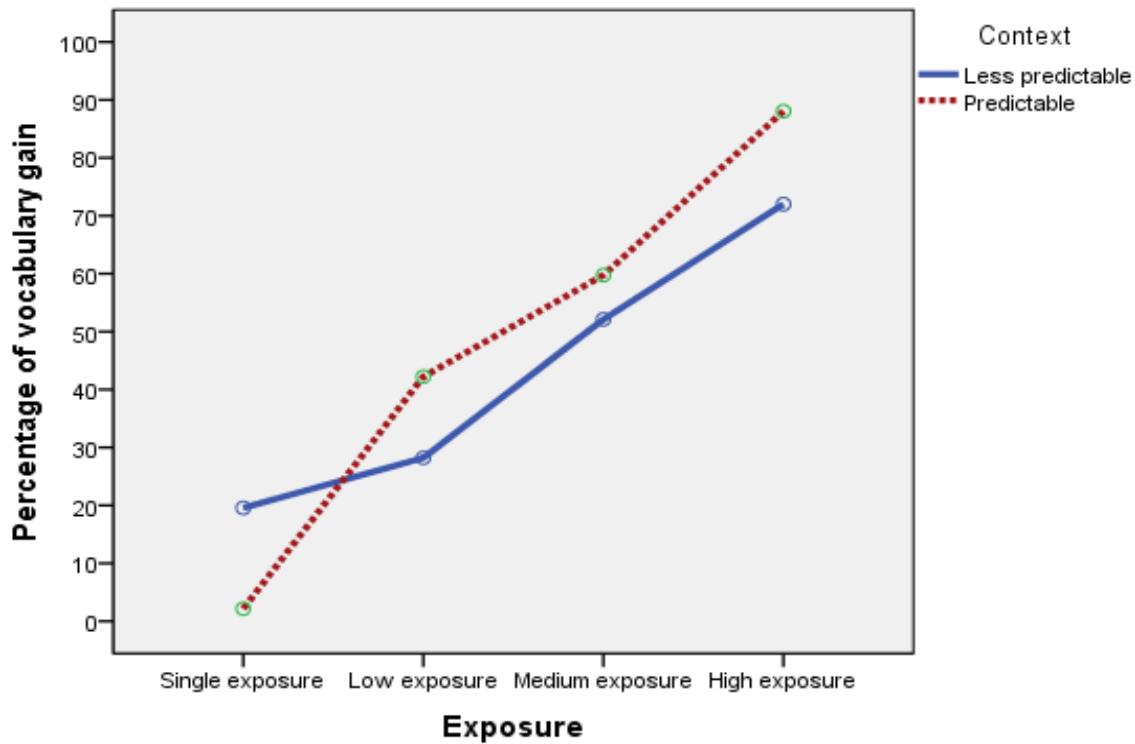
Figure 13. The interaction between exposure and predictability in form recognition

With contextual constraint categorized into predictable and less predictable bands, I divided exposure bands into four categories: single exposure, low exposure (2-5), medium exposure (6-9), and high exposure (10 and more). Figure 13 illustrates that highly predictable items yielded relatively better gains except for single exposure words. On the other hand, Figure 14 on meaning recognition indicates that predictable context makes the largest difference in the medium-exposure band. In meaning recall, this variance becomes clearer as high context words are more likely to be recalled in all exposure bands while less predictable words with single, low and medium exposures were not recalled as much (see Figure 15). Overall, repetition was effective in all vocabulary gains but context predictability enhanced these gains, especially in the medium-exposure band.
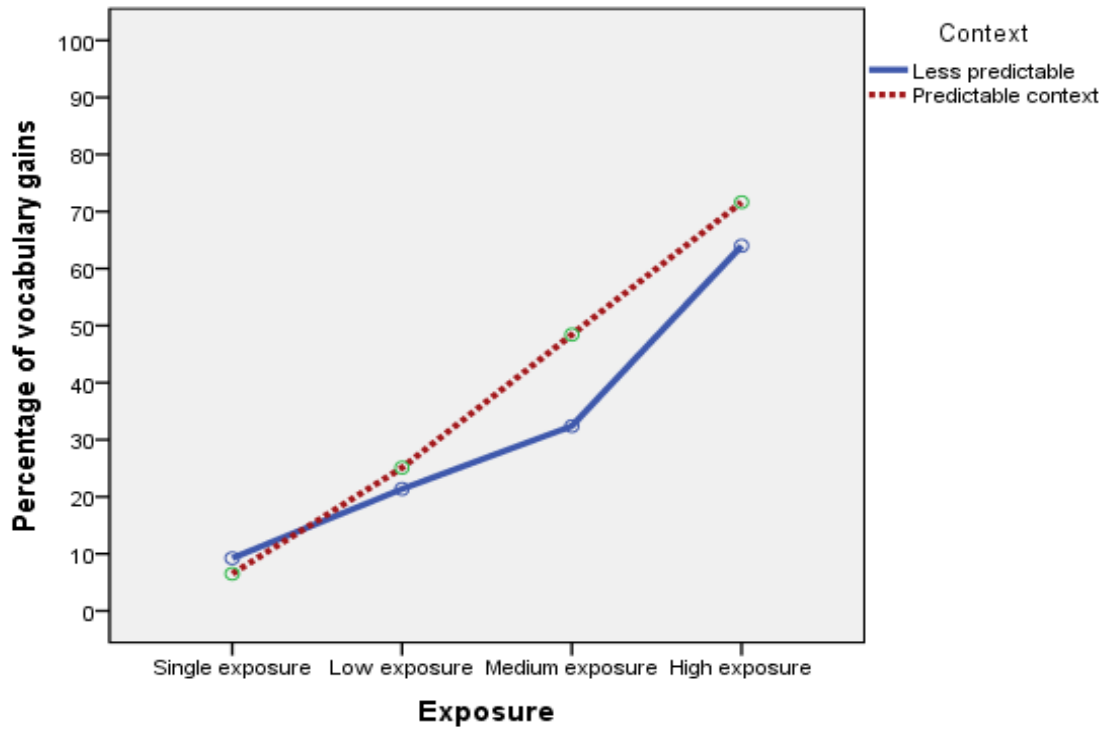
Figure 14.The interaction of exposure and context in meaning recognition
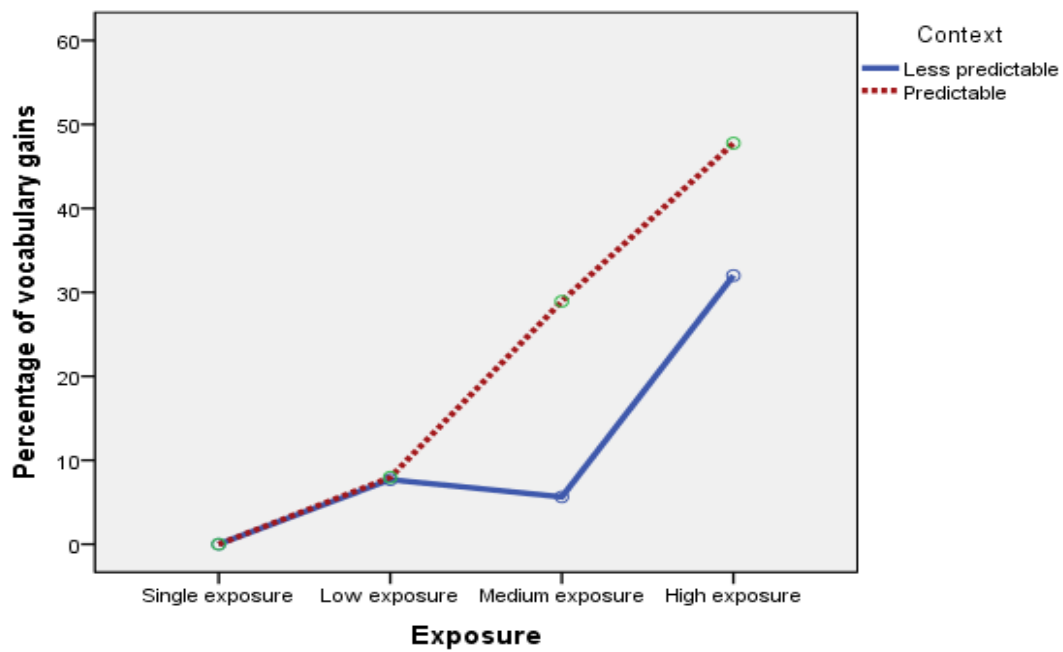
Figure 15. The interaction of exposure and context in meaning recall

3.3.3 Real time processing and vocabulary learning

In this section, I investigate real time processing of the target words in text and how moment-by-moment eye movement measures and token predictability can predict that certain vocabulary items will be acquired from reading. Because the data included items nested within subjects and encounters nested within items, I fitted a binary logistic regression in a three-level GLMM for each vocabulary test. Online reading measures and token predictability were entered as fixed factors with subjects and items as random factors and word length and total exposure as control variables.

Results yielded significant positive relationships of form recognition with first fixation durations ($OR = 1.21$, 95% CI = [1.13, 1.32], $p = .035$) and total reading times ($OR = 1.42$, 95% CI = [1.12, 1.80], $p =.004$), indicating that a one second increase in first fixations and total times spent on a target occurrence increased the probability of form recognition success by 4 % and 7 % respectively. Token predictability was not a significant predictor for form recognition ($OR = 0.91$, 95% CI = [0.96, 2.32], $p = .084$. Table 18 outlines the regression output for form recognition.

Table 18

Token-based predictors of form recognition

	Odds	OR	95% CI		*p*
Intercept	0.26		.075	0.93	.038 *
First fixation duration		1.21	1.13	1.32	.035 *
Gaze duration		1.16	0.73	1.82	.533
Total time		1.42	1.12	1.80	.004 **
Token predictability		1.45	0.91	2.32	.116

Note: The (*) marks signify the level of significance of the *p* value

Meaning recognition results pointed to a positive effect of total reading times on vocabulary outcomes (OR =1.33, 95% CI = [1.03, 1.72], p = .029). A one second increase in reading times of each token increased the probability of meaning recognition by 3 %. In addition, token predictability was highly associated with meaning recognition success (OR = 2.81, 95% CI = [1.81, 4.34], $p < .001$), implying that one unit increase in the predictability of individual encounters increased the chance of meaning recognition by 22 %. Table 19 summarizes token-based predictors of meaning recognition.

Table 19

Token-based predictors of meaning recognition

	Odds	OR	95% CI		p
Intercept	0.10		.034	0.32	$< .001$ ***
First fixation duration		2.65	0.81	8.67	.106
Gaze duration		1.54	0.71	1.88	.560
Total time		1.33	1.03	1.72	.029 *
Token predictability		2.81	1.81	4.34	$< .001$ ***

Note: The (*) marks signify the level of significance of the p value.

Meaning recall was significantly predicted by gaze durations (OR = 2.19, 95% CI = [1.22, 3.77], p = .005) and total reading times (OR = 1.73, 95% CI = [1.14, 2.63], p = .010). One additional second spent on target tokens increased the probability of meaning recall by 3 %. Token predictability showed a strong positive effect on meaning recall (OR = 5.68, 95% CI = [3.19, 10.25], $p < .001$), implying that an increase in predictability increased the probability of meaning recall by almost 17 %. Table 20 outlines the regression output for meaning recall.

Table 20

Token-based predictors of meaning recall

	Odds	OR	95% CI		*p*
Intercept	.051		.008	0.30	.002 **
First fixation duration		0.77	0.26	2.31	.650
Gaze duration		2.19	1.22	3.77	.005 **
Total time		1.73	1.14	2.63	.010 *
Token predictability		5.68	3.19	10.25	< .001 ***

Note: The (*) marks signify the level of significance of the *p* value

3.3.4 The role of cumulative online processing in vocabulary learning

Because summed fixation times reflected the cumulative processing effort devoted to target words, it was interesting to test how these measures would compare with text-based factors (exposure and item predictability) in explaining the variance in vocabulary outcomes. I fitted binary logistic regressions using a two-level GLMM because holistic effects based on items rather than the encounter level are of interest.

Table 21

Regression output of the online vs. text-based predictors of form recognition

	Odds	*OR*	95% CI		*p*
Intercept	0.18		0.094	0.35	< .001 ***
Summed FFD		0.53	0.34	0.83	.006 **
Summed GD		1.17	0.85	1.61	.332
Summed TFD		2.16	1.21	3.38	< .001 ***
Total exposure		1.29	1.18	1.41	< .001 ***
Item predictability		1.01	0.98	1.09	.704

Note: The (*) marks signify the level of significance of the *p* value

Table 21 points to a negative relationship between summed first fixation durations and form recognition in that a one second increase in total first fixation durations decreased the probability of successfully recognizing word form by almost 6 % ($OR = 0.53$, 95% CI = [0.34,0.83], $p = .006$). On the other hand, total reading times spent on target words positively increased the chances of learning form ($OR = 2.16$, 95% CI = [1.21, 3.38], $p < .001$), indicating that looking for one extra second at target words increased the probability of form recognition success by 13 %. At the level of text-based features, total exposure positively influenced form recognition although the effect was somewhat smaller than online processing times ($OR = 1.29$, 95% CI = [1.18, 1.41], $p < .001$).

Meaning recognition was significantly predicted by summed reading times ($OR =1.47$, 95% CI = [1.25, 1.72], $p < .001$) and total exposure ($OR =1.38$, 95% CI = [1.21, 1.58], $p < .001$). Meaning recall followed the same pattern with total reading times ($OR =3.27$, 95% CI = [1.28, 5.33], $p < .001$) and total exposure ($OR =1.27$, 95% CI = [1.13, 1.41], $p < .001$). In both models, item predictability was significant, although with a modest association strength. Tables 22 and 23 summarize the predictors of meaning recognition and recall.

Table 22

Regression output of the online vs. text-based predictors of meaning recognition

	Odds	*OR*	95% CI		*p*
Intercept	0.15		.060	0.38	< .001 ***
Summed FFD		0.81	0.48	1.39	0.45
Summed GD		0.78	0.51	0.21	. 284
Summed TFD		1.47	1.25	1.72	< .001 ***
Total exposure		1.38	1.21	1.58	< .001***
Item predictability		1.10	1.01	1.21	.047 *

Note: The (*) marks signify the level of significance of the *p* value

Table 23

Regression output of the online vs. text-based predictors of meaning recall

	Odds	*OR*	95% CI		*p*
Intercept	0.033		0.007	0.15	< .001 ***
Summed FFD		1.07	0.61	1.88	.825
Summed GD		0.80	0.49	1.28	.352
Summed TFD		3.27	1.28	5.33	< .001 ***
Total exposure		1.27	1.13	1.41	< .001 ***
Item predictability		1.16	1.03	1.30	.016 *

Note: The (*) marks signify the level of significance of the *p* value

A general overview of Tables (21-23) indicates that holding the effects of total exposure and item predictability constant, summed reading times strongly predicted learning success in all vocabulary measures particularly in form and meaning recall. This might suggest that individual attention on the part of the reader can be more important in explaining vocabulary learning above and beyond repeated exposure.

3.4 Individual differences in learning from reading

Eye Link trial reports showed that readers spent an average of 19.2 minutes ($SD = 4.59$) on the actual text with a mean reading speed of 258 words per minute. The range of their TOEFL IBT scores (79-100) and vocabulary sizes (2950- 4200) out of 5000 suggested that most participants were at upper-intermediate to advanced levels in English proficiency and that the reading material, with pseudo tokens, met the lexical coverage threshold necessary to ensure adequate reading comprehension and the possibility of learning from reading. Average comprehension scores ranged from 60 % to 100% ($M = 87$, $SD = 8.6$), indicating a generally good understanding of story content.

Post-reading questionnaire included items on a 6-point scale ranging from 1 (strongly disagree) to 6 (strongly agree). Average responses indicated that readers enjoyed the story (*M*= 4.9, *SD*=.86) while they did not express much discomfort with reading on the eye tracker (*M*=3.6, *SD*=1.6). They also expressed that the text was easy to read and there was no need to use a dictionary. Table 24 summarizes descriptive statistics for the items in the reading questionnaire.

Table 24

Mean responses on the reading perception questionnaire

Question	Mean response (*SD*)
Ease of reading	4.6 (.94)
Reading enjoyment	4.9 (.86)
Need of dictionary	2.9 (1.5)
Reading comfort	4.3 (1.3)
Eye tracking discomfort	3.6 (1.6)

Individual differences in reading speed, reading comprehension scores and vocabulary size did not yield significant effects on any of the vocabulary measures. However, a small effect of proficiency scores was found on form recognition (OR =1.014, 95% CI = [1.001, 1.025], $p =$ .039) and meaning recall (OR =1.023, 95% CI = [1.00, 1.067], $p = $ .049), indicating that more proficient readers may have shown slightly better retention of form and meaning of pseudo words encountered in the text.

3.5 General summary of results

Participants in this study read the graded reader '*Goodbye Mr. Hollywood*', a stage 1 story of 4649 words, which was found to be well within their current English proficiency levels. The percentage of pseudo tokens in the story was below 3 %, and the lexical coverage of the story was satisfactory (Nation, 2001, 2006). Readers met pseudo words and familiar control words in equal number of exposures ranging from 1 to 30, yielding 121 tokens in each condition. The predictability of tokens ranged between 0 and 96 based on English native speaker cloze agreement percentages. Eye movements were recorded during reading to compare online and text-based effects on incidental vocabulary learning.

Online reading patterns pointed to significant differences between attention to target items and familiar items. First fixations, gaze durations and total times decreased as a result of additional encounters to target words, pointing to a gradual increase in familiarity with pseudo words in the text as they were repeated. The decrease in reading times was more significant in early encounters (1-12) than in later encounters. After about 12 encounters, both conditions started to elicit similar processing patterns. Conversely, the role of token predictability in reducing processing load was more important in later encounter than in early encounters. The interaction between condition and predictability suggested that the role of predictability might have been slightly more pronounced in processing familiar control words than with pseudo words. Analyses of regressions and skips confirmed, as was to be expected, the extra attention devoted to pseudo words in early encounters. The overall summed fixation times were significantly influenced by condition and exposure band, suggesting that more repetition normally invited more attention and that pseudo words elicited longer summed processing times than known words over repeated exposures.

Readers displayed learning outcomes in form recognition followed by meaning recognition and finally meaning recall. Total exposure predicted all vocabulary outcomes while maximum item predictability supported meaning recognition and recall. The interaction between total exposure and predictability in text-based effects suggested that a rich context may have mitigated the positive effect of repetition in the process of retaining word meanings from reading. Overall, repetition was effective in all vocabulary gains while context predictability enhanced these gains, especially in the low-exposure and medium-exposure bands.

Token-based online processing measures demonstrated that Total time was a positive indicator of learning success in all vocabulary tests while first fixations only predicted form recognition and gaze durations only predicted meaning recall. Token-based predictability was an indicator of meaning recognition and recall but not of form recognition. When aggregating processing measures on all encounters, it was shown that only summed total time was positively associated with learning outcomes. After accounting for total exposure and item predictability, it was estimated that a one second increase in total times is a significant indicator of vocabulary learning particularly form and meaning recall. This suggested that word-based attention and utilization of context on the part of the reader can represent independent additive effects in the process of incidental learning from L2 reading.

Overall, participants represented a relatively homogenous group in terms of proficiency and vocabulary size. They expressed that the text was an easy and enjoyable piece of reading and reported a good comprehension of the details of the story. A slight effect of proficiency scores was observed in form recognition and meaning recall but no other effects of comprehension scores, reading speed or vocabulary size were found.

CHAPTER 4: DISCUSSION

Research on extensive reading has provided ample evidence on the role of repetition in lexical learning and called for further research on the role of contextual richness in vocabulary acquisition from L2 reading (e.g. Horst, 2005; Waring, & Nation, 2004; Webb, 2007, 2008). On the other hand, eye movement studies on reading behavior documented the cognitive effects of repetition and context quality on lexical processing and associated lexical retention in terms of online processing patterns and the eye-mind link hypothesis (Godforoid et al., 2013; Juhasz, & Pollatsek, 2011; Rayner, & Well, 1996). The present study aimed to bring together methods from both strands to investigate incidental vocabulary acquisition from L2 reading and track the cognitive effects of repetition and context predictability on the development of different aspects of vocabulary knowledge. In this chapter, I discuss the findings of the study in the light of the research questions and draw implications from extensive reading and eye movement research.

4.1 Lexical processing in repeated encounters

The first research question sought to investigate how second language readers processed unknown words in the graded reader '*Goodbye Mr. Hollywood*', and what textual factors influenced their reading patterns in real time. It was shown that readers gave relatively more attention to pseudo words as compared to familiar words, particularly in early encounters. Gaze durations and total times were inflated between encounters 1 and 12, after which target and control words started to exhibit similar processing patterns. Steady decreases were more significant in early encounters than in later encounters. The cutoff point around the $11^{th} - 12^{th}$ encounter was not clear for first fixation durations which showed significant yet small decreases across encounters and few differences by condition. A possible explanation can be provided in the light of the E-Z Reader model that postulates different stages of lexical processing (e.g.

Pollatsek, Reichle, & Rayner, 2006). In this model, first fixations reflect an early stage of familiarity check, and do not capture later events of reanalysis, word recognition or form-meaning mapping. The unfamiliarity of target words triggered subsequent fixations that fed into gaze durations, and the reported frequent regressions to target words ultimately fed into total times. This scenario may have caused the notable rise of attention exhibited in early encounters.

The fact that readers did pay more attention to pseudo words and particularly on early encounters was also confirmed by other evidence from reading behavior. In particular, skipping was less frequent on pseudo words while regressions occurred more frequently particularly in early encounters. Skipping instances occurring at pseudo words does not contradict with the fact that readers attended more to novel items. Parafoveal processing may have occurred for new words making them less likely to be skipped at first pass. Less skipping and more regressions indicated increased processing and reanalysis of target words, which may have supported the form-meaning mapping process. Repeated encounters were associated with shorter fixation times, which is consistent with previous eye movement research (e.g. Joseph et al., 2014), suggesting a gradual increase in familiarity with target forms over time.

The interaction between encounter and condition pointed to a possible exposure threshold after which pseudo words are read as fluently as familiar words. In extensive reading studies, 10 or more repetitions supported word learning (Pellicer-sanchez, & Schmitt, 2010). This may imply that a full knowledge of meaning might have been established after sufficient exposures, triggering more fluent reading. The observed cutoff point in gaze durations and total times after 12 or 13 encounters may point to this stage of meaning acquisition although it does not exclude the possibility that some readers internalized word meanings at earlier encounters or at least accumulated partial knowledge over successive exposures. A further support for this assumption

comes from the finding that regressions became less frequent in later encounters, indicating that readers might have formed plausible hypotheses about target word meanings in later encounters, which increased their fluency and made them proceed with reading with less hesitation.

The role of predictability in the present study was consistent with previous research that associated high context predictability with reduced reading times and higher skipping rates (Kleigel et al., 2004; Rayner, & Well, 1996). One further finding in the light of online processing results was that the role of predictability became more important in later encounters than early encounters with target words. A possible explanation for this observation is that pseudo words were better integrated in the sentence structure by later encounters because form retrieval became more fluent as a function of repetition. Due to the novelty of word forms, readers needed more repeated encounters to recognize them before they could rely on context to guess their meanings. This explanation is also consistent with assumptions from the E-Z Reader model presented by Reichle, Warren, and McConell (2009) who postulated a post-lexical integration stage that begins immediately after word identification. In this stage, readers may require additional time to construct higher-level representations such as linking the word to its syntactic structure, creating a context-based semantic representation or incorporate the word meaning at the discourse level. This explains the additional time shown for pseudo words in the present study, and the regression rates reported in early encounters.

The interaction between condition and predictability (as shown in Figures 3, 7 and 10) confirmed that highly predictable tokens required less processing in target and control condition although it can be noted that this effect was slightly more pronounced with control words. This finding is consistent with the perceived effect of form unfamiliarity, which interfered with the role of predictability in early encounters. It can also highlight the effect of lexical frequency on

processing based on previous research reviews which maintained that low frequency vocabulary attract longer processing times (Clifton, Rayner, & Staub, 2007; Rayner, Raney, & Pollatsek, 1995; Rayner, 2009, 2007; William, & Morris, 2004). From a lexical perspective, the pseudo words integrated in the text can be claimed to share features with low-frequency vocabulary in English. Previous eye movement studies found that the level of frequency and predictability independently affected reading times and interacted with the number of exposures (e.g. Ashby, Rayner, & Clifton; Rayner, Raney, & Rayner, 1995). Further research can shed more light on the hypothesized interaction between word frequency and context predictability.

Taking a broader perspective, the role of context predictability can in fact extend beyond lexical tokens because pseudo words can incrementally acquire higher predictability over later encounters at the discourse level. The readers' engagement with the content of the story at the discourse level can eventually feed into the predictability of individual words, particularly those that were repeated more often. This explains the fact that estimated item predictability reduced the summed first fixations and summed gaze durations on individual target words.

4.2 Text-based effects on vocabulary learning

In line with previous literature on vocabulary acquisition (Nation, 2001; Schmitt, 2008, 2010), knowledge of form seemed to be the first component to develop followed by meaning recognition and finally meaning recall. These differential learning rates can be explained in terms of a progression from the lowest to the highest cognitive demands on the learner's memory. In form recognition, the learner only needs to access the orthographic form of the target word from memory traces while in meaning recall the learner had to have sufficient informative encounters to guess meanings correctly and subsequently decontextualize words and retain them in memory, which is even more demanding task than meaning recognition where learners are given several

options that trigger access to memory of contextual information encountered in the text. The overall picture of learning outcomes shows plausible and predictable patterns in line with the incremental nature of vocabulary knowledge development (Schmitt, 2008, 2010).

Form recognition was mainly influenced by total exposure (see Table 15) while gains in meaning recognition and recall were determined by an interaction between total exposure and item predictability (Tables 16 and 17).

Repeated exposure of items that were categorized as highly predictable yielded increases in the chances of meaning recognition and recall while low context items did not show that linear trend, implying that the ambiguity of certain items attenuated the effects of repeated exposure in the acquisition of word meanings. A further finding is that the effect size of predictability was strongest in meaning recall, implying that the minimum gains reported in the meaning recall test were associated with the most predictable items in the text. These findings are in line with previous vocabulary research (Webb, 2008) that repetition supports knowledge of form while context quality supports knowledge of meaning.

Readers were able to retain traces of word forms due to repetition regardless of context while acquiring meaning required further contextual support which was not available with the same degree in all exposures. When a vocabulary item was highly predictable, high exposure was an ideal setting for accurate guessing and retention of word meanings while a combination of low context and high exposure was more conducive to form recognition and inconsistently associated with meaning gains. Overall, repetition was effective in all vocabulary gains while context predictability enhanced these gains, especially in the low-exposure and medium-exposure bands (Figures 13, 14 and 15).

4.3 Early indicators of vocabulary intake

Token-based analyses were conducted to explore whether lexical processing patterns and the predictability of individual encounters provided early predictions of the probability of retaining new vocabulary items in the three types of the posttest. Controlling for total exposure, it was found that total time spent on individual tokens was associated with successful intake in all the three vocabulary measures. Additionally, first fixations predicted form recognition while gaze durations predicted meaning recall.

The kind of associations found between online processing and different types of vocabulary gain aligns with the claim that different eye movement measures tap into different cognitive processes. Within the framework of the E-Z reader model (Reichle, Rayner, & Pollatsek, 2003; Pollatsek, Reichle, & Rayner, 2006), lexical processing has been posited to proceed in two stages: an early stage called 'familiarity check', and a later stage referred to as the completion of lexical access. The fact that first fixation durations predicted form recognition conforms to this hypothesis in that early lexical processing is largely form-focused (Reichle, Warren, & McConnell, 2009). Gaze duration, as the total duration of early processing, predicted meaning recall, which may indicate that subsequent lexical processing of form-meaning mapping and encoding into memory becomes more important with subsequent fixations on the target word. The same principle would explain why total time, as a late measure, predicted all types of vocabulary learning. Because total time marks the completion of lexical access and sentence integration, it was indicative of the total attention devoted to each individual token in the text. As the total time spent on every encounter of a target word increased, there was more chance that the reader would retain that word in all vocabulary measures.

Token predictability strongly supported meaning recognition and recall. This suggests that, after controlling for reading times, the contextual properties of individual encounters offered crucial support for retaining knowledge of meaning. Similar to the item-level analyses, token predictability was not significantly associated with form recognition. The input characteristics required for form retention seemed to be largely dependent on repetition regardless of context levels. It can be generalized that total reading times and token predictability were the major early indicators of vocabulary intake. As readers paid more attention to individual tokens, they were more likely to acquire knowledge of form and meaning about vocabulary items. The chance of meaning retention was boosted further when readers spent more time on highly predictable tokens.

4.4 Combined measures of attention and exposure

Another perspective in investigating attention to target words was to combine fixation times over individual tokens of each item. The goal of such analysis was to compare overall attention, as reflected in summed fixation times, with text-based effects of total exposure and item predictability. It was clearly shown that summed total reading times positively predicted learning outcomes in all vocabulary measures. This confirmed an association between online processing and lexical retention as documented by previous research (Godfroid et al., 2013).

The fact that summed total times were a strong predictor of learning outcomes after controlling for total exposure (Tables 21, 22 and 23) might indicate that individual attention to target words can explain the variance in vocabulary learning above and beyond mere repeated exposures. This finding aligns with lexical processing data which showed that readers invested more time in initial encounters checking for familiarity and reanalyzing context. From a reader's perspective, exposures were not equal in the amount of context and information they provided

about target words. Thus, when we compare online times with total exposure, we are actually comparing two dimensions of exposure that I may distinguish as dynamic versus static exposure. Dynamic exposure involves the sum of all the information that readers have accrued from all encounters with a given word while static exposure mainly represents an offline scale variable; that is, a number. In the present study, the dynamic exposure captured readers' interaction with target words and all the stages of lexical integration (Reichle et al, 2009) that have contributed to the incremental development of word knowledge as a byproduct of exposure. From this perspective, it was plausible to find that the way readers utilized their repeated encounters with target words strongly predicted learning outcomes beyond encounters per se.

It was interesting to note that the effect of item predictability was somewhat attenuated by summed reading times in meaning recognition and recall. This may suggest that, while context information remains important for meaning acquisition, dynamic exposure to specific target words may moderate this effect to some extent based on individual reading behavior.

4.5 Overview

Some general statements can be presented in the light of the above discussion. First, vocabulary learning from reading is not a byproduct of a single factor but it is rather influenced by multiple variables with variable effect sizes. What makes this type of statistical analysis complicated is that the model controls for the effects of other variables in the equation before assessing the effect of the variable of interest. Therefore, it may moderate or reduce other effects. For example, text-based variables were found to be good predictors of learning. However, their roles were moderated, showing that there were factors beyond the text that had more important roles, particularly reader's processing behavior.

Another general statement concerns the differences between token-based and item-based processing times. In real time processing of individual words (i.e., tokens), it was found that first fixation durations predicted form recognition while gaze durations predicted meaning recall. This was explained in the light of the E-Z Reader model. However, when these measures were summed by item their effects did not transfer for the most part (Tables 21 and 23). A possible explanation for this is that the two analyses provided two different pictures because the summed measures of first fixations or gaze durations only combine partial events of processing and do not reflect all the stages of lexical access. On the other hand, summed total time consistently predicted learning gains in both token-based and item-based analyses. This may well be because total times and summed times reflected a more inclusive inventory of lexical processing events.

Finally, the role of predictability has been consistent for meaning recognition and recall in both text-based and token-based analyses. It may be reasonable to assume that predictability complemented the role of exposure and boosted the effects of reading times in the development of word meanings from context. However, an interesting aspect of predictability that was shown in previous literature as well as in the present study is that high predictability induces shorter reading times and more frequent skipping (e.g. Kliegel et al, 2004; Rayner, & Well, 1996). Does this imply that high predictability tokens or words received less attention? One explanation for this apparent tension is that readers might have looked at high predictable tokens relatively less than other items because not as much processing was necessary for successful form-meaning mapping given that the context already provided part of the solution.

The overall vocabulary gains were good, relative to the amount of reading material and the limited time spent on task (around 41 % in form recognition, 30 % in meaning recognition and 13 % in meaning recall). Although individual differences had minor effects on learning

outcomes, the observed effect of attention as a holistic measure highlights the role of differences in reading behavior regarding incidental vocabulary learning. Because total times predicted vocabulary learning above and beyond total exposure, it can be concluded that it is the reader's use of exposure opportunities and context information that determines the amount and quality of learning from reading in addition to the static exposure and context properties of the written input.

CHAPTER 5: CONCLUSION

This final chapter is divided into three parts. First, I summarize major results and contributions of the study. Next, I present the practical and pedagogical implications informed by the results. Finally, I conclude with a brief discussion of limitations of the study, in addition to suggestions for future research.

5.1 Summary of the findings

The present study investigated incidental vocabulary acquisition from L2 reading using methods from extensive reading and eye movement research to highlight the role of cognitive processing in incidental vocabulary learning. An important contribution of the study was to introduce a natural reading task of reasonable length in an eye movement setting, which can represent a further step in understanding real time processes involved in incidental learning from reading. The experiment provided a considerable ecological validity regarding the reading task, making use of available authentic graded readers in a close approximation of leisure reading input that learners occasionally encounter in their ESL reading resources.

Readers exhibited signs of increased familiarity and reading fluency on target words over encounters whereas they paid more attention to new words in early exposures. Most learning was shown in form recognition, followed by meaning recognition and finally meaning recall. All learning outcomes were significantly predicted by total number of exposures while predictability only aided meaning recognition and recall. Total times on individual encounters provided early indicators of vocabulary learning success in all measures. Additionally, first fixations predicted form recognition whileras gaze durations predicted meaning recall. When aggregating processing measures, it was found that summed reading times predicted learning outcomes above and

beyond text-based characteristics, which highlights the important role of readers' individual attention and their optimal use of input to infer and retain meaning from context.

Results of the study emphasize the significant role of leisure reading in the incremental development of vocabulary knowledge, starting with the word form and gradually building connections with meanings and retaining them for immediate recall. Lexical properties influenced how readers interacted with text and identified new word forms and meanings. Repeated exposure to new vocabulary seems to be a key factor that guarantees sufficient processing opportunities for successful intake. However, learning opportunities were further enriched when highly predictable items were repeatedly encountered in text. The probability of retaining word meanings was most closely associated with the most predictable vocabulary items in the text. On the cognitive level, the consistent relationship between total processing times and learning outcomes indicated that readers were more successful in gaining lexical knowledge from reading when they paid more attention to new vocabulary, taking advantage of exposure opportunities and context features to gather information about lexical items across several encounters.

The present study sheds more light on the cognitive aspects of engagement (Schmitt, 2008) and involvement (Laufer, & Hulstijn, 2001), which were emphasized in vocabulary acquisition research and particularly within the incidental learning framework. Reader engagement with lexical items is reflected in online measures which capture ongoing processing of new vocabulary in different contexts. This adds another dimension to extensive reading as a source of vocabulary development, distinguishing between learning opportunities offered by the text and the expected learning outcomes based on textual features and readers' engagement.

5.2 Practical and pedagogical implications

The results of the study are mostly relevant to second language vocabulary learning and teaching. Maximizing exposure to vocabulary in rich contexts is a recommended strategy to ensure the best conditions for internalizing partially known words or acquiring new vocabulary. Exposure is not only confined to reading, but can also be extended to task-based learning where different input modalities (speaking, listening, reading and writing) can integrate vocabulary learning goals in variable contexts (Brown, Waring, & Donkaewbua, 2008). Task-based learning can extend beyond the classroom to include online courses that can be adapted to enhance the opportunities for incidental exposure to vocabulary in self-study modules. To increase the chances of vocabulary acquisition, it is recommended to recycle new vocabulary repeatedly through different types of teaching tasks over several class sessions to provide different contexts for targeted words.

The present study corroborates previous research on the role of extensive reading mainly in developing reading fluency along with creating possible opportunities for learning new vocabulary. Increasing reading fluency can be an early stage that sets the scene for acquiring new vocabulary. One relevant implication for extensive reading is that it can afford more familiarity with new lexical items but this does not guarantee successful internalization of new word meanings in a limited time frame. This is particularly true in light of the fact that the effects of extensive reading are longitudinal in nature. Reading programs should be evaluated over longer periods of time, considering all factors of input, textual features and individual reading behavior as well as learners' motivation.

Vocabulary reviews have shown that word knowledge is multifaceted and being able to retrieve the meaning of a given L2 word is just one aspect of this knowledge (Nation, 2001;

Schmitt, 2008). Lexical gain results in the present study corroborate this principle and relate it to multiple exposures and levels of predictability in the text. Teachers should consider this fact in their testing material so as to accurately gauge different levels of their students' lexical knowledge and set up plans for their vocabulary building strategies.

5.3 Limitations and further research

The current study provides additional insights in SLA vocabulary research and extends further understanding of the cognitive aspects of incidental vocabulary acquisition. As a newly integrated technology in second language vocabulary research, the eye-tracking technique can answer specific questions about learners' interaction with L2 material with considerable temporal and spatial accuracy. Implementing eye-tracking methodology in SLA is likely to open new avenues of investigation to uncover detailed cognitive processes in language acquisition in general and vocabulary development in particular.

Some methodological issues need to be discussed regarding the nature of tasks and participants in the present study. Using a head mount and a chin rest during the reading task might have interfered with the natural reading behavior of readers to some extent. Further eye-tracking research can make use of more advanced techniques to maximize the ecological validity of task performance without jeopardizing the accuracy of eye movement measures. The second point concerns the use of pseudo words for the study. As learners were expected to know the real words for the target items, they may have concluded that the novel words they encountered in reading were less frequent synonyms of the words they already knew, an impression that may have reduced their motivation or cognitive effort to incorporate the new lexical items. Moreover, the lab-controlled experiment condensed the number of exposures into one experimental session, which may not exactly match the typical incremental route that learners go through in incidental

88

learning, where repeated exposures are spaced over longer periods of time. For practical reasons, delayed post tests were not conducted. Further research should consider the role of repetition and context on vocabulary retention over time.

The roles of repeated exposure and context quality can be investigated in different modalities and different teaching tasks. Findings from task-based vocabulary learning research have associated vocabulary acquisition with the concepts of engagement (Schmitt, 2008) and involvement load (Laufer, & Hulstijn, 2001). Looking at the cognitive perspectives of task performance through eye tracking techniques can shed more light on how learners respond to tasks with different levels of difficulty, how engagement is reflected in their online processing and how their attention resources are divided between task completion and lexical processing of unknown words.

Vocabulary acquisition from L2 reading is usually characterized as incidental when learners are not forewarned of a vocabulary test after receiving input. In the current study, the amount of attention measured through eye movements seemed to be learner-driven because there was no external motivation that manipulated the existence or amount of attention on target vocabulary. Future research can examine how drawing attention of readers to focus on novel words in L2 input can yield different processing patterns and subsequently reflect on the amount of vocabulary gains. However, this kind of methodological manipulation should point to vocabulary gains in terms of a clear distinction between incidental and intentional learning setting.

Individual differences in proficiency and vocabulary size did not yield significant effects on how much they were able to utilize context to learn new words. However, it may be interesting for further inquiry to investigate how diverse L1 backgrounds may have a role in

89

facilitating or hampering lexical acquisition from reading particularly in form recognition. Future research can address effects of script differences on incidental learning and how it interacts with learners' proficiency in vocabulary development.

Finally, the ideal extensive reading study will be longitudinal in nature and it evaluates learning outcomes from several readings over longer periods of times (Horst, 2005). The present study provided a model for further large-scale research that can consider a wider variety of reading material and more authentic texts with different populations of second language learners. Although eye movement research can provide precise quantitative account of lexical processing, it would be an additional asset in future studies to apply stimulated recalls or think-aloud protocols to explore qualitative aspects of attention to target words and reading fluency and their relationship to vocabulary acquisition (Rott, 2005; Rott, & Williams, 2003). Generally speaking, combining quantitative and qualitative methods to explore lexical learning from reading would add to our understanding of attention and engagement in reading comprehension and provide further implications on the process of incidental vocabulary learning from L2 reading.

APPENDICES

Appendix A: Participant Information

Table 25 Participants' proficiency and vocabulary size chart

Participant ID	Age	Gender	L1	Proficiency	Vocabulary size
1	38	M	Arabic	79	3300
2	21	M	Portuguese	80	3800
3	29	M	Twi	86	2870
4	35	F	Spanish	100	4490
5	19	F	Portuguese	88	3190
6	23	M	Chinese	80	4730
7	30	F	Arabic	79	4570
8	19	F	Japanese	86	3140
9	19	M	Chinese	80	4420
10	29	F	Spanish	100	4930
11	20	F	Russian	90	3250
12	21	M	Chinese	86	4260
13	40	F	Portuguese	96	4630
14	18	M	Ndebele	88	3880
15	20	F	Korean	83	4130
16	19	M	Chinese	92	4290
17	24	M	Shona	97	2830
18	19	F	Chinese	92	2980
19	19	F	Japanese	82	3800
20	20	M	Japanese	90	4630
21	20	M	Chinese	79	3900
22	31	M	Chinese	84	3900
23	18	M	Spanish	93	2790
24	30	F	Chinese	87	2730
25	30	F	Amharic	100	4680
26	21	F	Spanish	94	4660
27	19	F	Hindi	97	4230
28	26	M	Chinese	88	4070
29	24	M	Chinese	79	3890
30	27	F	Chinese	80	4360
31	21	F	Yoruba	100	3300
32	20	F	Portuguese	86	3510
33	28	M	Swahili	86	4520
34	23	F	Hindi	100	3550
35	20	F	Chinese	79	2820
36	22	F	Portuguese	90	4470
37	21	M	Polish	100	4430
38	22	F	Arabic	86	3290
39	25	M	Arabic	79	3920
40	27	F	Hindi	90	4570
41	28	F	Japanese	88	3580
42	22	F	Japanese	79	4630

Appendix B: Background questionnaire

PLEASE FILL OUT THE FOLLOWING BACKGROUND INFORMATION:

1.　　First name: _____　　Last name: _____

2.　　Age: _____

3.　　Gender:　☐ Male　　☐ Female

4.　　Year of study:

　　☐ Freshman　　☐ Sophomore　☐ Junior　　☐ Senior　☐ Graduate Student

5.　　Major _____

6.　　In which section of English are you enrolled?　……………………..

7.　　How many years have you been studying English?　_____

8. How old were you when you started learning English　………………………….

9. Your native language:　…………………………

10. Other languages you studied ………………………………………………….

11. Language(s) spoken at home ………………………………………..

12. Your recent TOEFL/IELTS score …………………………….

13. On a scale from 1 to 9, how do you rate your skills in the following areas of English?

Reading	1	2	3	4	5	6	7	8	9
Writing	1	2	3	4	5	6	7	8	9
Vocabulary size	1	2	3	4	5	6	7	8	9
Overall Proficiency	1	2	3	4	5	6	7	8	9

Thank you very much for your time

Appendix C: Sample of reading material 'Goodbye Mr. Hollywood'

Chapter 1: Mystery Girl

It all began on a beautiful spring morning in a village called Whistler, in Canada- a pretty little village in the mountains of British Columbia. There was a café in the village, with tables outside, and at one of these tables sat a young man. He finished his breakfast, drank his coffee, looked up into the blue sky, and felt the warm sun on his face. Nick Lortz was a happy man. The waiter came up to his table.' More coffee? 'He asked. 'Yeah. Great,' said Nick. He gave the waiter his coffee cup. The waiter looked at the camera on the table. 'On vacation?' he said. 'Where are you from?' 'San Francisco,' Nick said. He laughed. But I'm not on vacation - I'm working. I'm a travel writer, and I'm doing a book on mountains in North America. I've got some great pictures of your mountain. The two men looked up at Whistler Mountain behind the village. It looked very beautiful in the morning sun. 'Do you travel a lot, then? Asked the waiter. 'All the time, 'Nick said. I write books, and I write for travel magazines. I write about everything – different countries, towns, villages, rivers, mountains, people, the waiter looked over Nick's head. , There's a girl across the street, 'he said. , Do you know her?' Nick turned his head and looked. "No, I don't.", 'well, she knows you, I think,' the waiter said. , She's watching you very carefully. He gave Nick a smile. .Have a nice day! 'He went away, back into the cafe.

Nick looked at the girl across the street. She was about twenty-five, and she was very pretty. She is watching me, Nick thought. Then the girl turned and looked in one of the shop windows. After a second or two, she looked back at Nick again. Nick watched her. She looks worried,' he thought. 'What's she doing? Is she waiting for somebody? Suddenly, the girl smiled. Then she walked across the street, came up to Nick's table, and sat down. She put her bag down on the table. The bag was half open. 'Hi! I'm Jan,' she said. 'Do you remember me? We met at a party in Toronto. "Hi, Jan,' said Nick. He smiled. 'I'm Nick. But we didn't meet at a party in Toronto. I don't go to parties very often, and never in Toronto.' 'Oh, 'the girl said. But she didn't get up or move away. 'Have some coffee,' said Nick. The story about the party in Toronto wasn't true, but it was a beautiful morning, and she was a pretty girl. , Maybe it was a party in Montreal. Or New York.' The girl laughed. 'OK. Maybe it was. And yes, I'd love some coffee. 'When she had her coffee, Nick asked, what are you doing in Whistler? Or do you live here? 'Oh no,' she said. 'I'm just, err, just travelling through. And what are you doing here?' 'I'm a travel writer,' Nick said, and I'm writing a book about famous mountains.' 'That's interesting,' she said. But her face was worried, not interested, and she looked across the road again. A man with very short, white hair walked across the road. He was about sixty years old, and he was tall and thin. The girl watched him. 'Are you waiting for someone?' asked Nick. 'No,' she said quickly. Then she asked, where are you going next, Nick?' 'To Vancouver, for three or four days, 'he said. 'When are you going?' she asked. 'Later this morning,' he said.

There was a letter in the top of the girl's half-open bag. Nick could see some of the writing, and he read it because he saw the word 'Vancouver' - . . . and we can meet at the Empress Hotel, Victoria,Vancouuer island, on Friday afternoon . ' . 'So she's going to Vancouver too" he thought. Suddenly the girl said, 'Do you like movies?' 'Movies? Yes, I love movies, 'he said' "Why?' 'I know a man, and he - he loves movies, and going to the cinema,' she said slowly. 'People call him "Mr Hollywood".' She smiled at Nick. , Can I call you "Mr Hollywood" too?' Nick laughed.' OK, 'he said, And what can I call you? She smiled again. Call me Mystery Girl,

she said. 'That's a good name for you, said Nick. Just then, the man with white hair came into the cafe. He did not look at Nick or the girl, but he sat at a table near them. He asked the waiter for some breakfast. Then he began to read a magazine.

The girl looked at the man, then quickly looked away again. 'Do you know him? Nick asked her. No,' she said. She finished her coffee quickly and got up. 'I must go now, 'she said. Nick stood up, too. "Nice to –, he began. But the girl suddenly took his face between her hands, and kissed him on the mouth. , Drive carefully, Mr. Hollywood. Goodbye, she said, with a big, beautiful smile. Then she turned and walked quickly away. Nick sat down again and watched her. She walked down the road and into a big hotel. 'Now what,' thought Nick, was that all about?

The man with white hair watched Nick and waited. After four or five minutes, Nick finished his coffee, took his books and his camera, and left the cafe. His car was just outside the girl's hotel, and he walked slowly along the street to it. The man with white hair waited a second, then quickly followed Nick. From a window high up in the hotel, the girl looked down into the road. She saw Nick, and the man with white hair about fifty yards behind him. Nick got into his car, and the man with white hair walked quickly to a red car across the street. Five seconds later Nick drove away in his blue car, and the red car began to follow him. When the girl saw this, she smiled, then went to put some things in her travel bag.

Appendix D: Sample page from the comprehension packet

Chapter TWO: hand in the back

<u>Check True or False</u>

1-Nick asked the girl to see her again in Vancouver.

2-The girl was not telling the truth about her name.

3-The weather was nice in Vancouver when Nick arrived.

4-A car hit Nick in the middle of the street.

<u>Answer briefly</u>

5-What do you think happened to Nick in the street ? Who did that to him?

………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………

6-How did Nick know more information on the mystery girl?

a)She called him and told him everything

b)from a TV show about her family

c)from a magazine he was reading

d)The police man told him about her

7-Nick learned things about the Mystery girl (check all what applies):

a)She changed her name

b)she is from Torronto

c)she is a daughter of a millionaire

d)she knows the man with white hair

Appendix E: Reading perception questionnaire

Read the following statements and say how much you agree or disagree with them by simply circling a number from 1 to 6.

Strongly disagree	disagree	Slightly disagree	Slightly agree	agree	Strongly agree
1	2	3	4	5	6

I enjoyed reading this book	1	2	3	4	5	6
I would like to read it again	1	2	3	4	5	6
This was comfortable to read	1	2	3	4	5	6
I needed a dictionary to read it	1	2	3	4	5	6
It was easy to read	1	2	3	4	5	6
The story was complicated and not clear	1	2	3	4	5	6
I often lost track of the story	1	2	3	4	5	6
It took longer to read than I expected	1	2	3	4	5	6
In general, I enjoy reading such stories	1	2	3	4	5	6
The eye tracking experience was disturbing	1	2	3	4	5	6

Appendix F: Form recognition test

Circle only the words you have encountered in the story

1	ship	sense	joker	tame	table
2	bannop	hospital	bus	rude	pag
3	fozle	mystery	tance	mave	lame
4	fonteen	gell	chortan	stoll	tund
5	shame	window	nase	gun	camera
6	rim	sind	fake	mork	letter
7	blef	red	rimple	cube	pamery
8	kerp	crasty	lead	mot	shoes
9	bannow	havoc	barn	money	pennem
10	hungry	subid	room	speat	smick
11	happy	bandle	neech	doom	prink
12	jurgs	busy	hair	manage	desk
13	levider	tidge	yelt	noise	airport
14	commute	dress	drink	system	bing
15	bannifet	meet	similar	dillet	tantic
16	windle	mand	push	redaster	vack
17	leam	tantic	popkum	nook	toker
18	fungi	dook	dangy	megole	smile
19	cheem	borch	gotty	tickeny	palk
20	dorch	gube	plampy	cold	dern

Appendix G: Meaning recall test

		No clue	Meaning?	Similar to	Related to ?
1	ferry				
2	mot				
3	blef				
4	fonteen				
5	mystery				
6	windle				
7	rude				
8	leam				
9	redaster				
10	dorch				

Appendix H: Meaning recognition test

	Meaning recognition				
	Circle the best meaning for each of the words below as far as you know.				
word	1	2	3	4	5
upset	Tired	unhappy	poor	cheap	I don't know
jump	lie on top of water	move very fast	stop a car	get off the ground suddenly	I don't know
tantic	annoyed	strong	famous	hungry	I don't know
joke	false statement	kind of humor	way of speaking	way of thinking	I don't know
blef	To hide	To push	To kill	to move	I don't know
windle	club	garden	letter	file	I don't know
neech	Shirt	hair	dress	drawer	I don't know
dangy	dirty	happy	quiet	bright	I don't know
mave	eye glasses	working desk	person's face	big couch	I don't know
bannow	Cap	roof	window	door	I don't know
leam	Slow	fast	tall	busy	I don't know
century	length measure	hundred years	record	past time	I don't know

100

Chapter 3 : A Walk in the Park

The next day was Thursday. Nick stayed in his [____] room. He was busy all morning writing about mountains. Then he drove to Stanley Park in the afternoon. He sat and read a book for an hour, and then he went for a walk under the trees. It was [____], and he could walk and think. He thought about Meg Hutson, and about the man with white hair. He remembered Meg Hutson's last words. Drive carefully, Mr Hollywood. Why did she call him Mr Hollywood?

Suddenly, he heard a noise. He [____]. 'That was a gun!' he thought. There's somebody in the trees with a gun! Then something hit the tree over his head. 'Somebody's shooting at me!' Nick thought. He turned and ran. And somebody began to run after him. Nick ran through the trees. There was no sun in here, and it was half dark. And there were no people. Nobody to help him. He ran on. He could still hear the man behind him, so he ran faster. After three or four minutes, he [____] and listened. Nothing. It was all [____]. Nick was afraid. Why is somebody shooting at me? He waited another second or two, then walked quickly back to his car. Then he saw people, women with young children, some boys with a football, and two men with a dog. He began to feel better. 'Nobody can shoot me now,' he thought. Ten minutes later, he was back at his car. There was a letter on the car door. Nick read it. It said: I'm going to [____] you, Mr Hollywood. He became really afraid.

Nick drove to the nearest police station. He waited for half an hour, then a tired young policeman took him into a small room. Nick told his story, and the policeman wrote it all down. 'So what are you going to do?' asked Nick. 'Nothing,' said the policeman. 'Nothing!' said Nick. 'But somebody shot at me. 'Mr Lortz,' the policeman said tiredly. 'How many people are there in this town with guns?' 'I don't know, 'said Nick. 'You didn't see the shooter. Was it a man, a boy, a woman? Maybe it was an old girlfriend. Maybe somebody doesn't like your travel books or it may be just an [____]'. Think about it in a different way. 'But what about the man with white hair in Whistler? said Nick. 'The girl, Meg Hutson, called me Mr. Hollywood in the [____], and this man heard her. Who is this Mr. Hollywood? 'We all want answers to our questions, Mr Lortz, the policeman said, 'but we don't always get them. Nick walked out of the police station and drove to his [____]. He was angry, and afraid. 'How did the man with white hair find me in Vancouver?' he thought. 'Did he follow me from Whistler? Maybe he's staying at my [____] in the next room waiting to shoot me.

[Submit]

Appendix J: Token predictability data

Table 26 Estimated predictability for target tokens

fozle

Predictability (version A)					Predictability (version B)				
56.7	5	8.3	6.7	62.6	70.9	74.5	90.9	80	36.4
67.5	37.5	80	47.5	90	96.4	90.9	92.7	45.4	85.5
75	87.5	80	20	32.5	48	46	14.6	56.3	83.3
90	77	82.5	52.5	52.5	77.1	68.8	2.1	0	79.2
57.5	7.5	45	10	60	77.1	72.9	77.1	72.9	64.6
5	55	35	52.5	37.5	15.9	75	50	60	52.5

gube

Predictability (version A)						Predictability (version B)					
38.3	15	25	31.7	27.5	28.3	28	79.2	43.3	83.3	84	72.9
30	28.3	15	37.5	32.5	37.5	4.2	58.3	83.3	29.2	31.3	43.8
35	12.5	30	32.5	27.5	25	33.3	35.4	39.6	75	81.8	79.5

mave

Predictability (version A)					Predictability (version B)				
85	66.7	12.5	45	12.5	0	50	50	75	72.9
17.5	70	77.5	80	82.5	12.5	14.6	16.7	64.6	66.7

tund

Predictability (version A)					Predictability (version B)				
75	50	17.5	90	10	92.7	94.5	40	54.2	12.5
32.5	32.5	62.5	10		37.5	63.6	20.5	68.2	

leam

Predictability (version A)				Predictability (version B)			
51.7	15	60	47.5	50	30	56.3	43.8
47.5	27.5	42.5		50	37.5	43.2	

blef

Predictability (version A)			Predictability (version B)		
42.5	70	70	76	82	88
67.5	70	77.5	78	79	64.6

toker

Predictability (version A)			Predictability (version B)		
25	25	21.7	82	56.8	22.7
25	25	27.5	81.8	63.6	61.4

bannow

Predictability (version A)			Predictability (version B)		
47.5	77.5	55	5.5	56.4	49.1
57.5	57.5		0	72.7	

mot

Predictability (version A)			Predictability (version B)		
70	48.3	80	22.9	68.8	72.9
85	85		56.3	77.1	

fonteen

Predictability (version A)		Predictability (version B)	
82.5	80	0	72.7
90	85	75	77.3

dangy

Predictability (version A)		Predictability (version B)	
27.5	62.5	25.5	72.9
17.5	12.5	45.8	8.3

windle

Predictability (version A)		Predictability (version B)	
0	5	1.8	81.3
5	0	16.7	6.8

neech

Predictability (version A)			Predictability (version B)		
5	60	60	44	40	42

redaster

Predictability (version A)			Predictability (version B)		
87.5	97.5	65	54.5	50	18.2

dook

Predictability (version A)		Predictability (version B)	
72	70	13.6	72.7

tance

Predictability (version A)	Predictability (version B)
65	59

pamery

Predictability (version A)	Predictability (version B)
71.7	77

tantic

Predictability (version A)	Predictability (version B)
10	37.5

dorch

Predictability (version A)	Predictability (version B)
57.7	45.8

smick

Predictability (version A)	Predictability (version B)
72.5	50

REFERENCES

REFERENCES

Altarriba , J. , Kroll , J.F. , Scholl , A. , Rayner , K. (1996). The influence of lexical and conceptual constraints on reading mixed language sentences: Evidence from eye fixation and naming times . *Memory and Cognition, 24* , 477 – 492 .

Ashby, J., Rayner, K., & Clifton, C. J. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology, 58A*, 1065-1086.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17*, 364–388.

Beck, I. L., McKeown, M. G., & McCaslin, E. S. (1983). All contexts are not created equal. *Elementary School Journal, 83,* 177–181.

Brown, R., Waring, R., and Donkaewbua, S. (2008). Vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*, (2)136–163.

Brusnighan, S. M., & Folk, J. R. (2012). Combining contextual and morphemic cues is beneficial during incidental vocabulary acquisition: Semantic transparency in novel compound word processing. *Reading Research Quarterly, 47*(2), 172-190.

Bruton, A., Garcı´a Lo´pez, M., & Esquiliche Mesa, R. (2011). Incidental L2 vocabulary learning: An impracticable term? *TESOL Quarterly, 45*, 759–768.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, amq031.

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). Amsterdam: Elsevier, North Holland.

Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge: Cambridge University Press.

Coady, J., & Huckin, T. N. (1997). *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge, U.K.; New York: Cambridge University Press.

Day, R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language, 7*, 541–551.

Dobinson, T. (2001). Do learners learn from classroom interaction and does the teacher have a role to play? *Language Teaching Research, 5*(3), 189-211.

Eckerth, J., & Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research, 16*(2), 227-252.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of verbal learning and verbal behavior, 20*(6), 641-655.

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24*(2), 174–187.

Elley, W. B. (1991). Acquiring literacy in a second language: The effect of book-based programs. *Language Learning, 41*, 375– 411.

Ellis, R. (1999). *Learning a second language through interaction*. Amsterdam: John Benjamins.

Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition, 21*(2), 285-301.

Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language Learning, 44*(3), 449-491.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532-538.

Folse, K. (2006). The Effect of Type of Written Exercise on L2 Vocabulary Retention. *TESOL Quarterly, 40*(2), 273-293.

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

Fraser, C. (2007). Reading rate in L1 Mandarin Chinese and L2 English across five reading tasks. *The Modern Language Journal, 91*, 372–394. doi: 10.1111/j.1540-4781.2007.00587.x

Gass, S. (1999). Incidental vocabulary learning. *Studies in Second Language Acquisition, 21*(2), 319-333.

Gass, S. M., Behney, J., & Plonsky, L. (2013). *Second language acquisition: An introductory course*. New York: Routledge.

Grabe, W., & Stoller, F. (1997). Reading and vocabulary development in a second language: A case study. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 98–123). Cambridge, England: Cambridge University Press.

Grabe, W., & Stoller, F. (2011). *Teaching and researching reading* (2nd ed.). Harlow, Essex:Pearson Education.

Godfroid, A., Housen, A., & Boers, F. (2010). A procedure for testing the Noticing Hypothesis in the context of vocabulary acquisition. In M. Pütz & L. Sicola (Eds.), *Inside the learner's mind: Cognitive processing and second language acquisition* (pp. 169-197). Amsterdam/Philadelphia: John Benjamins.

Godfroid, A. (2012). Eye tracking. In P. Robinson (Ed.), *The Routledge encyclopedia of second language acquisition* (pp. 234-236). New York/London: Routledge.

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. doi: 10.1017/S0272263113000119

Haastrup, K. (2008). Lexical inferencing procedures in two languages. In Albrechtsen, D., Haastrup, K., and Henriksen, B., *Vocabulary and Writing in a First and Second Language: Process and Development.* Basingstoke: Palgrave Macmillan. pp. 67–111.

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). Range [Computer software]. Retrieved from http://www.victoria.ac.nz/lals /staff/paul-nation/nation.aspx

Heck, R. H., Thomas, S., & Tabata, L. N. (2012*). Multilevel modeling of categorical outcomes using IBM SPSS*. New York: Routledge.

Hill, M. and Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *International Review of Applied Linguistics in Language Teaching, 41*(2), 87–106.

Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*, 689–696.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review, 61*(3), 355-382. doi: 10.3138/cmlr.61.3.355

Horst, M., Cobb, T., &Meara, P.(1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207-223.

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

Hu, H.C., & Nassaji, H. (2012). Ease of inferencing, learner inferential strategies, and their relationship with the retention of word meanings inferred from context. *Canadian Modern Language Review, 68*(1), 54-77.

Hu, H. C. M. (2013). The Effects of Word Frequency and Contextual Types on Vocabulary Acquisition from Extensive Reading: A Case Study. *Journal of Language Teaching and Research, 4*(3), 487-495.

Huang, S., Willson, V., & Eslami, Z. (2012). The Effects of Task Involvement Load on L2 Incidental Vocabulary Learning: A Meta-Analytic Study. *Modern Language Journal, 96*(4), 544-557.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition, 21*(2), 181-193.

Huckin, T. N., Haynes, M., & Coady, J. (1993). *Second language reading and vocabulary learning*. Norwood, N.J. Ablex Publishing Corporation.

Hulstijn, J. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P Arnaud &: H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 113-125). London: Macmillan Academic and Professional Limited.

Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal, 80*, 327–339.

Hulstijn, J. H., & Trompetter, P. (1998). Incidental learning of second language vocabulary in computer-assisted reading and writing tasks. In Albrechtsen, D., Hendricksen, B., Mees, M., & Poulsen, E. (Eds.) *Perspectives on foreign and second language pedagogy* (pp. 191–200). Odense, Denmark: Odense University Press.

Hulstijn, J., & Laufer, B. (2001). Some Empirical Evidence for the Involvement Load Hypothesis in Vocabulary Acquisition. *Language Learning, 51*(3), 539-558.

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

Hulstijn, J. (2003). Incidental and intentional learning. In Doughty, C., & Long, M. (Ed.) *The handbook of second language acquisition*. Malden, MA ; Oxford : Blackwell Publishing.

Hyönä, J. & Niemi, P. (1990). Eye movements in repeated movements of a text. *Acta Psychologica, 73*, 259-280.

Jing, L., & Jianbin, H. (2009). An empirical study of the involvement load hypothesis in incidental vocabulary acquisition in EFL listening. *Polyglossia, 16*, 1-11

Joe, A. (2010). The Quality and Frequency of Encounters with Vocabulary in an English for Academic Purposes Programme. *Reading in a Foreign Language, 22*(1), 117-138.

Joseph, H. S., 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.

Juhasz, B. J. & Pollatsek, A. P. (2011). Lexical influences on eye movements in reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook on eye movements* (pp.873–893). Oxford: Oxford 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.

Knight, S. (1994). Dictionary Use While Reading: The Effects on Comprehension and Vocabulary Acquisition for Students of Different Verbal Abilities. *The Modern Language Journal, 78*, 3, 285-299.

Kowen , S. & Kim, H. (2008). Beyond raw frequency: Incidental vocabulary acquisition in extensive reading. *Reading in a Foreign Language. 20*(2), 191-215.

Lai, F. K. (1993). The effect of a summer reading course on reading and writing skills. *System, 21*, 87–100.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review, 59*, 567 587.

Laufer, B. (2005). Focus on Form in Second Language Vocabulary Learning. *EUROSLA Yearbook, 5*, 223-250. doi: 10.1075/eurosla.5.11lau

Laufer, B., & Hulstijn, J. (2001). Incidental Vocabulary Acquisition in a Second Language: The Construct of Task-Induced Involvement. *Applied Linguistics, 22*(1), 1-26.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*, 15–30.

Liversedge, Simon P. and Rayner, Keith (2011) Linguistic and cognitive influences on eye movements during reading. In, Gilchrist, Iain and Everling, Stefan (eds.) The Oxford Handbook of Eye Movements. Oxford, GB, Oxford University Press, 751-766.

Liversedge, S., Gilchrist, I., & Everling, S. (Eds.). (2011). *The Oxford handbook of eye movements*. Oxford University Press.

Lupescu, S., & Day, R. (1993). Reading, dictionaries, and vocabulary learning. *Language Learning, 43*(2), 263-287.

Macaro, E. (2003). *Teaching and learning a second language* : A review of recent research. London; New York: Continuum.

Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21*(4), 557-587.

Mackey, A., & Goo, J.. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A.Mackey(Ed.), *Conversational interaction in second language acquisition* (pp.433-464). Oxford University Press.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition, 22*(4), 471-497.

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal, 82*(3), 338-356.

Matsuoka, W., & Hirsh, D. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language, 22*(1), 56-70.

McCullagh, P., Nelder, J. A., & McCullagh, P. (1989). *Generalized linear models* (Vol. 2). London: Chapman and Hall.

Meara, P. (1992). EFL vocabulary test. Swansea, UK: University College, Centre for Applied Language Studies.

Menard, S. (2010). *Logistic regression: From introductory to advanced concepts and applications*. Thousand Oaks, CA: SAGE Publications, Inc.

Mohamed. A (in press). Task-based incidental vocabulary learning in L2 Arabic: The role of proficiency and task performance. Accepted, *JNCOLCTL.*

Mohamed. A (2012). Investigating incidental vocabulary acquisition in conversation classes: A qualitative and quantitative analysis. *MSU Working Papers in SLS, 3*(1), 30-48.

Mondria, J. A. (2003). The effects of inferring, verifying, and memorizing on the retention of L2 word meanings. *Studies in Second Language Acquisition,25*(04), 473-499.

Mondria, J. A., & Wit-de Boer, M. (1991). The Effects of Contextual Richness on the Guessability and the Retention of Words in a Foreign Language1. *Applied Linguistics, 12*(3), 249-267.

Nagy, W. (1997). On the role of context in first-and second-language vocabulary learning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary description, acquisition, and pedagogy* (pp. 64–83). Cambridge: Cambridge University Press.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal, 24*(2), 237–270.

Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly,37*(4),645–670.

Nation, P., & Wang, K. (1999). Graded readers and vocabulary. *Reading in a foreign language, 12*(2), 355-380.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge; New York: Cambridge University Press. doi: 10.1017/CBO9781139524759

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*, 59–82.

Paribakht, T. S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition, 21*(2), 195-224. doi: 10.1017/S027226319900203X

Parry, K. (1991). Building a vocabulary through academic reading. *TESOL Quarterly, 25*, 629–653.

Pellicer-Sanchez, A., & Schmidt. (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., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German vocabulary through reading: The effect of three enhancement techniques compared. *Language Learning,59*, 113–151.

Pigada, M., &Schmitt, N.(2006). Vocabulry acquisition from extensive reading: A case study. *Reading in a Foreign Language, 22*(1), 1-28.

Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the clockwork orange study using second language acquirers. *Reading in a Foreign Language, 5*, 271–275.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*, 1–56.

Powers, D. A., & Xie, Y. (2008). *Statistical methods for categorical data analysis*. Bingley, UK: Emerald.

Pulido, D. (2007). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning, 57*(1), 155-199. doi: 10.1111/j.1467-9922.2007.00415.x

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*, 720–732

Raney, G., & Rayner, K. (1995). Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology, 49*(2), 151-172. doi: 10.1037/1196-1961.49.2.151

Rayner, K., Raney, G. E., & Pollatsek, A. (1995). Eye movements and discourse processing. In R. F. Lorch and E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 9-36). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372-422. doi: 10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506.

Rayner,K.,&Well,A.D.(1996).Effects of contextual constraint on Eye movements in reading:A further examination. *Psychonomic Bulletin & Review, 3*,504-509.

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24*, 146-161. doi: 10.1017/S0267190504000078

Reichle, E.D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye movement control in reading: Comparison to other models. *Brain and Behavioral Sciences*, *26*, 445–476.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using EZ Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic bulletin & review, 16*(1), 1-21.

Richards, J., & Schmidt, R. (2002). *Longman dictionary of language teaching and applied linguistics. Malaysia*: Pearson Education.

Robb, T.N., & Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language, 5*, 239– 251.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition, 21*(3), 589-619.

Rott, S. (2005). Processing glosses: A qualitative exploration of how form-meaning connections are established and strengthened. *Reading in a Foreign Language, 17*(2), 95-124.

Rott, S., & Williams, J. (2003). Making form-meaning connections while reading: A qualitative analysis of word processing. *Reading in a Foreign Language, 15*(1), 45-75.

Rott, S., Williams, J., & Cameron, R.(2002). The effect of multiple-choice L1 glosses and input-output cycles on lexical acquisition and retention. *Language Teaching Research, 6*, 183-222.

Saragi, T., Nation, I. S. P., & Meister, F. (1978). Vocabulary learning and reading. *System, 6*, 72–78.

Schaffin, R., Morris, R.K., & Seely, R.E. (2001). Learning new words in context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 225-235. doi: 10.1037/0278-7393.27.1.225

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363. doi: 10.1177/1362168808089921

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*, 26–43.

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

Schouten-van Parreren, C. (1989). Vocabulary learning through reading: Which conditions should be met when presenting words in texts. *AILA review, 6*(1), 75-85.

Schwanenflugel, P. J., Stahl, S. A., & Mcfalls, E. L. (1997). Partial word knowledge and vocabulary growth during reading comprehension. *Journal of Literacy Research, 29*(4), 531-553.

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(2), 344.

Uden, J., Schmitt, D., & Schmitt, N. (2014). Can learners make the jump from the highest graded readers to ungraded novels?: Four case studies. *Reading in a Foreign Language, 26*, 1, 1-28.

Van Gompel, R. P. G., Fischer, M. H., Murray, W. S., & Hill, R. L. (EDS.) (2007).*Eye movements: A window on mind and brain*. Oxford: Elsevier.

Vidal, K. (2010). A Comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*(1), 219-258.

Waring, R., & Nation, I. S. P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English speaking world, 4*, 97-110.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from a graded reader? *Reading in a Foreign Langugae, 15*, 130-163.

Watanabe, Y. (1997). Input, intake and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition,19*, 287-307. doi: 10.1017/S027226319700301X.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 27*(01), 33-52.

Webb, S. (2007).The effects of repetition on vocabulary knowledge. *Applied Linguistics,28*,46-65. doi: 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. (2010). Using glossaries to increase the lexical coverage of television programs. *Reading in a Foreign Language, 22*, 201-221.

Williams, R.S., & Morris, R.K. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology, 16*, 312-339.

Winke, P. M., Godfroid, A., & Gass, S. M. (2013). Introduction to the special issue. Eye-movement recordings in second language research. *Studies in Second Language Acquisition, 35*(2), 205-212. doi: 10.1017/S027226311200085X

Wochna, K. L., & Juhasz, B. J. (2013). Context length and reading novel words: An eye movement investigation. *British Journal of Psychology, 104*(3), 347-363.

Yaqubi, B., Rayati, A., & Gorgi, N.(2010). The Involvement Load Hypothesis and Vocabulary Learning: The Effect of Task Types and Involvement Index on L2 Vocabulary Acquisition. *The Journal of Teaching Language Skills.2, 1, 59/4.*

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review, 57*(4), 541-572.

Zimmerman, C. B. (1997). Historical trends in second language vocabulary instruction. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 5-19). England: Cambridge U Press.