Research on extensive reading has provided ample evidence on the role of repetition in lexical learning (Chen, & Truscott; Horst, Cobb & Meara, 1998; Pellicer-Sanchez & Schmitt, 2010; Rott, 1999; Waring and Takaki, 2003; Webb, 2005). However, less insights were provided on how the context quality in each encounter with target words would mediate or facilitate acquisition of new words from reading. Fewer results were reported for the effect of context quality on vocabulary learning outcomes (Fraser, 1999; Haastrup, 2008; Hu, 2013; Joe, 2010; Nassaji, 2003; Webb, 2008; Zahar, Cobb, and Spada, 2001). Paper-and-pencil studies implied that rich context supported word meaning acquisition yet variable parameters were used to operationalize context quality and explain outcomes. On the other hand, eye movement studies on reading behavior documented cognitive effects of repetition and context predictability on lexical processing (Wochna & Juhasz, 2013) and associated vocabulary learning with processing patterns in the light of the eye-mind link hypothesis (Rayner, 1998, 2009). The goal of the present study is to use eye movement methods to explain context processing in second language reading from a cognitive perspective as an initial step to interpret the process of meaning acquisition during reading.

BACKGROUND

Extensive reading and L2 vocabulary

Extensive reading has been shown to be an appropriate alternative to authentic novel reading given that these require at least a vocabulary size of 8000 to 9000 word families for adequate comprehension and new vocabulary intake (Hu & Nation, 2000; Nation, 2001, 2006; Nation & Wang, 1999; Waring & Nation, 2004). Previous research has valued extensive reading as a way of increasing reading fluency, comprehension, and speed of access to frequent words as well as the potential of supporting a smooth transition to authentic novel reading (Day & Bamford, 1998; Elley, 1991; Grabe & Stoller, 2011; Horst, 2005; Lai, 1993; Parry, 1991; Uden, Schmitt, & Schmitt). The majority of research in this area investigated lexical gains in relation to repeated exposure to vocabulary items, advocating an average of 8 to 10 repetitions as appropriate for the development of receptive knowledge of vocabulary with relatively low gains in productive knowledge (Schmitt, 2010). For example, Saragi, Nation and Meister (1978) found that native English speakers were able to learn an average of 76 % of 90 Russian slang words used in a long novel. Lower rates were found in Pitts, White and Krashen (1989) as readers learned about 6.4 % to 8.1 % of 30 target Russian words. Similarly in Day, Omura and Hiramatsu (1991), Japanese EFL learners learned an average of 3 words out of 17 target words in a simplified short novel. Horst, Cobb and Meara (1998) reported that learners could pick up an average of 5 words out of the 45 target words in a simplified novel. Waring and Takaki (2003) distinguished aspects of vocabulary knowledge, finding that learners did best in meaning recognition than productive translation. Pigada and Schmitt (2006) found improvements 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. Pellicer-Sanchez & Schmitt (2010) investigated vocabulary learning outcome from an authentic novel and found that meaning recognition reached 84 % after ten exposures while meaning recall was around 55%.

Contextual richness and vocabulary learning

Previous literature indicates that guessing from context is unreliable in learning vocabulary (Laufer, 2005; Nassaji, 2003).Others point to the essential role of noticing and processing context for retaining word knowledge Fraser, 1999; Paribakht & Wesche, 1999). 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 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. 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) maintained 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.

Cognitive perspectives on lexical learning

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). 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.

Eye tracking. The assumption of an 'eye-mind' link, proposes a connection between overt and covert attention (Rayner, 1998, 2009). The online recording of learners' eye movement behavior, including fixation times (how longer readers look at interest areas) and saccades (the movement of the eyes from one point to the next) have been used to signal different cognitive processes in response to different textual and contextual properties (Godfroid, 2012). Reviews of eye tracking research show that eye movements provide an accurate representation of the cognitive processes in the reader's mind. A large amount of research used recordings of eye movements to explore the psychological processes that control the reading behavior of adult readers (see Rayner, 1998, 2009 for a review). Several computational models were proposed to explain the 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).

Simulations of eye movements in reading studies pointed to a popular model named the E-Z reader as a serial attention hypothesis and was found 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). It 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). In the light of this model, the durations of 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; Juhasz & Pollatsek, 2011). 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.

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) found that readers' fixation times decreased consistently from first to third encounter with target sentences. Raney and Rayner (1995) 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) reported 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.

Regarding lexical processing of context, eye movement studies have consistently shown that high context predictability was associated with shorter fixations and consistent skipping (Ehrlich and Rayner, 1981; Rayner and Well, 1996). 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.

Few studies have investigated a potential association between online processing patterns and learning new words. Chaffin, Morris, and Seely (2001) determined that the familiarity of target words and context quality (informative or neutral) correlated with the amount of time readers spent on the target words. Learners fixated the most on novel words encountered in neutral contexts. Williams and Morris (2004) found that readers spent more processing time on novel words than familiar words, and that there was a positive association between online processing patterns (i.e. reading times), and retention of new word meanings. Brusnighan and Folk (2012) maintained 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. 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.

Recent eye movement studies specifically targeted vocabulary in second language reading. Godfroid et al. (2013). They operationalized attention to novel pseudo words as a quantitative variable reflected in the participants' eye fixation times during reading. They found significant association between the total fixation time on pseudo words and subsequent recognition of these words in a surprise posttest. Pellicer-Sanchez (2015) stressed the same association between looking at pseudo words and retention of meaning traces. She also concluded a process of increased fluency in reading over repeated encounters.

Goals of the study

Neither paper-and-pencil studies nor eye movement research has given sufficient attention to the role of context in processing and learning new words. There has also been a lack of objective operationalization of context types and its interaction with repetition. 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 effect of context predictability on processing and vocabulary intake as it interacts with repeated exposure.

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.

CURRENT STUDY

Research questions

The current study is guided by the following research questions:

RQ1. How do learners of English in the study process novel lexical items in silent reading relative to known control items? And how do context predictability influence lexical processing of repeated pseudo words compared to control words in the text?

RQ2. Does context predictability interact with repeated exposure to influence the acquisition of receptive and productive knowledge of form and meaning of target words in vocabulary posttests?

RQ3. To what extent do eye fixation times and total reading times on target words predict the learning gains of L2 readers in the vocabulary knowledge posttests?

Participants

Forty two advanced second language learners of English (22 females and 20 males) ranging in age from 19 to 35 (M = 22, SD = 4.2) participated in the study. They 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). Their vocabulary sizes, measured at the 5k level using Meara's (1992) vocabulary size test, yielded an average of 3908 (SD = 659).

Material

Vocabulary size test. Adapted from Meara (1992), this 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. Examples of this test can be found online at (http://www.lextutor.ca/tests/).

Reading material. 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 lexical density (types/tokens ratio) was not high (12.9%). Table 1 outlines the lexical distribution of the text across frequency levels.

Table 1

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

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. The experimental items in each version were replaced by matching pseudo words retrieved from online 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. The total number of pseudo tokens in each version was 121, which accounted for 2.6 % of the total tokens in the text. Based on these criteria, pseudo words were inserted and the text was divided into shorter parts (seven chapters) for programming. Table 2 presents these pseudo forms and their control counterparts with the distribution of occurrence in the text.

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.

Table 2

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

Vocabulary tests. 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. The form recognition 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. The meaning recall test required participants to recall meanings, synonyms, related words or semantic fields for the given items. Finally, the meaning recognition test was 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.

Procedure

Apparatus. 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 story content was provided in 70 screens, each containing 60-70 words in double spaced text. 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.

The reading and testing session. Participants were randomly assigned to either version A or B in the experiment builder. After each chapter, participants left the eye tracker for the comprehension check then returned back to resume reading after the calibration for the new chapter. This procedure continued for each chapter until the end of the story. The reading session for each participant including calibration, breaks and comprehension check took an average of 45 to 70 minutes. For testing, participants then took the vocabulary tests in the following order to avoid transfer effects: form recognition, meaning recall and meaning recognition.

Modified cloze procedure. 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. The task was electronically programmed and publicized to native speakers on campus.

A total of 108 entries were received (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.

Analyses

Definition of variables. The analysis in the present study distinguishes between online and offline effects on vocabulary outcomes. Online variables refer to the information in eye movement records including total reading times, skipping rates, regressions-in and regressions-out of the interest areas. 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. For item-based predictability, I used the maximum scored predictability to account for it. Item maximum was further categorized into predictable and less predictable guided by previous research (Ehrlich, & Rayner, 1981; Rayner, & Well, 1996; Schwanenflugen, & LaCount, 1988) with a cutoff point of 77%.

Data structure. Forty-two valid samples were considered for analysis. 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. 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. The model expands to three levels when we need to test the level of encounter including information about all the tokens of all items.

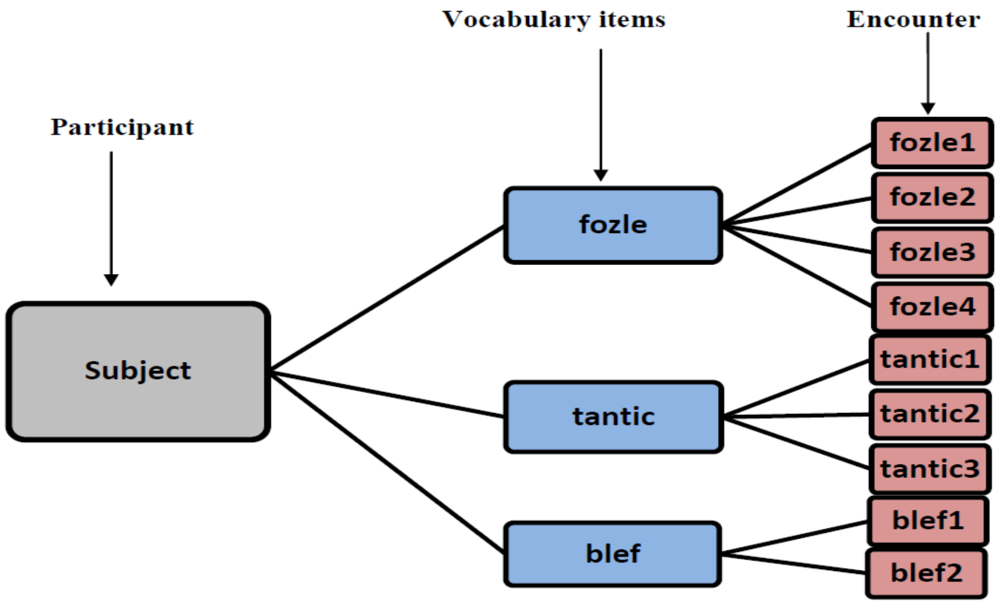

Figure 1. Data structure for participants and target words

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). 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).

RESULTS

Online reading times

The total times spent on every target token in the text were averaged over encounters to investigate the effect of textual features on reading patterns. The scatter plot in Figure 2 shows that total reading times recorded highest at the first encounter of target words (M = 702 ms, SD = 512) and lowest by the final encounter (M=265 ms, SD=130). 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 3 summarizes main textual effects on total reading times.

Table 3

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

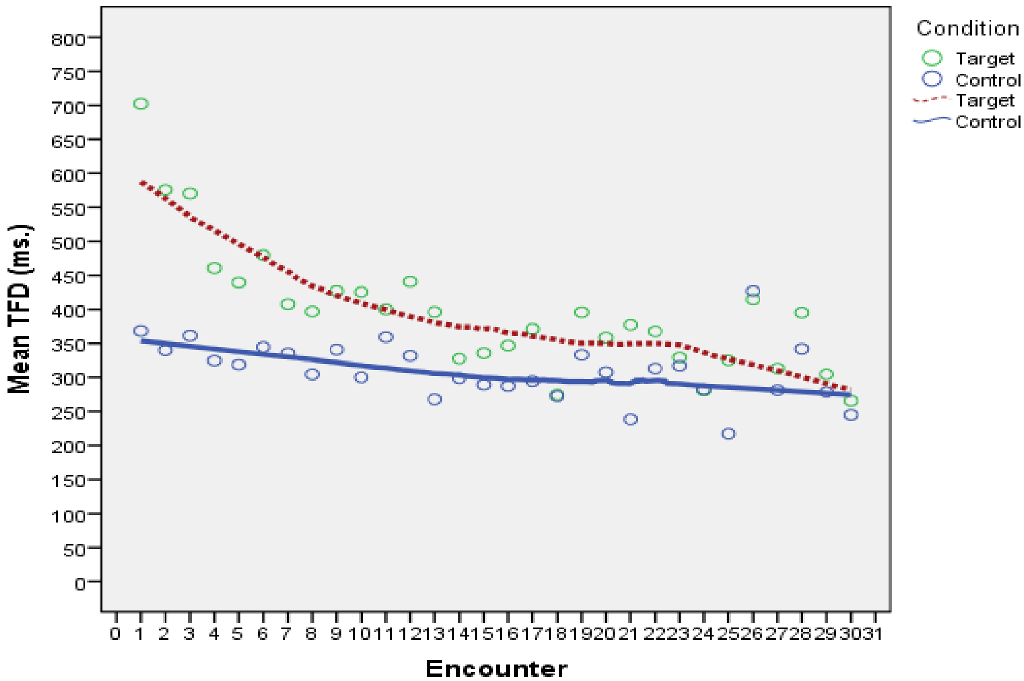

Figure 2. Scatter plot for mean total durations by encounter and condition

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 3, which shows that the effect of predictability was slightly more pronounced on familiar words rather than on pseudo words.

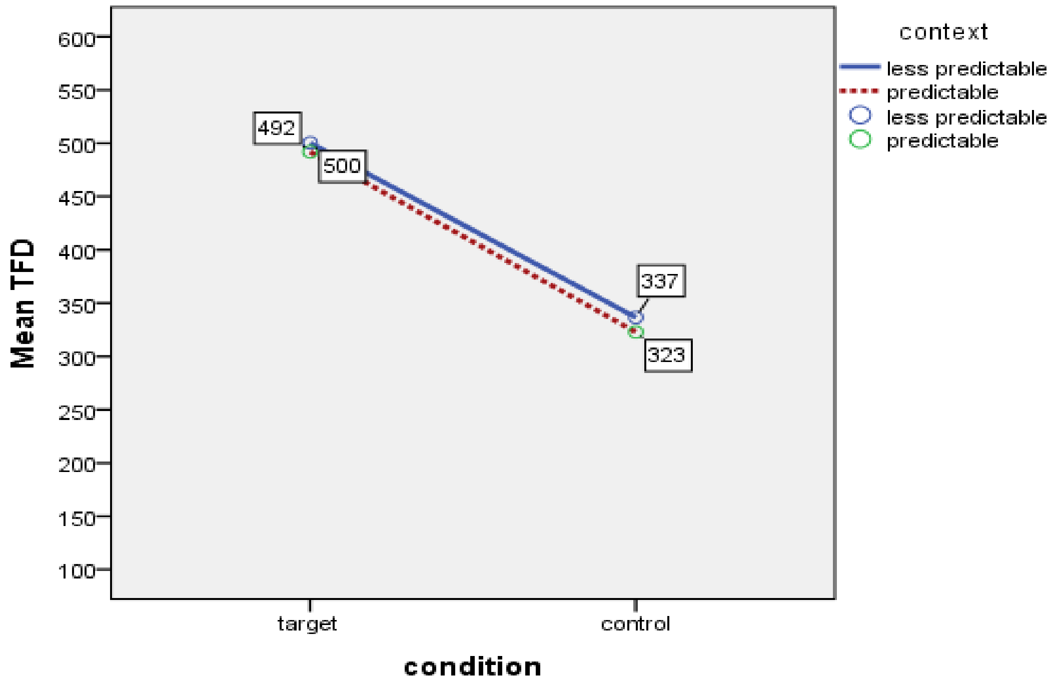

Figure 3. The interaction of condition and predictability in total fixation durations (TFD)

Based on the scatter plot for total times in Figure 2, 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], $p = .189$).

Skipping. Skipping was more frequent in the control condition than in the experimental condition (around 21% of target occurrences and almost 26% of control occurrences). 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], $p = .273$). As predictability increased, there was more probability of skipping (*OR* = 1.19, 95% CI = [1.05, 1.25], $p = .001$).

Regressions-in. Readers returned more to target items (almost 25%) than to control items (almost 15%). 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.

Regressions-out. Regressions-out occurred on almost 27 % of target observations and on 22 % of control observations. 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], $p = .010$). Table 4 summarizes text-based effects on reading behavior patterns.

Table 4

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

Vocabulary knowledge gains from reading

In overall vocabulary measures participants reported the highest gains in form recognition, followed by meaning recognition and finally meaning recall. Table 5 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. 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 6-8) indicated that each additional exposure increased the probability of form recognition by around 2 %, meaning recognition by around 3 % and meaning recall by 2 %.

Table 5

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 %)

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 6 through 8 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 4 through 6 illustrate the interacting effects of context and repetition on vocabulary gains.

Table 6

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 7

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 8

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

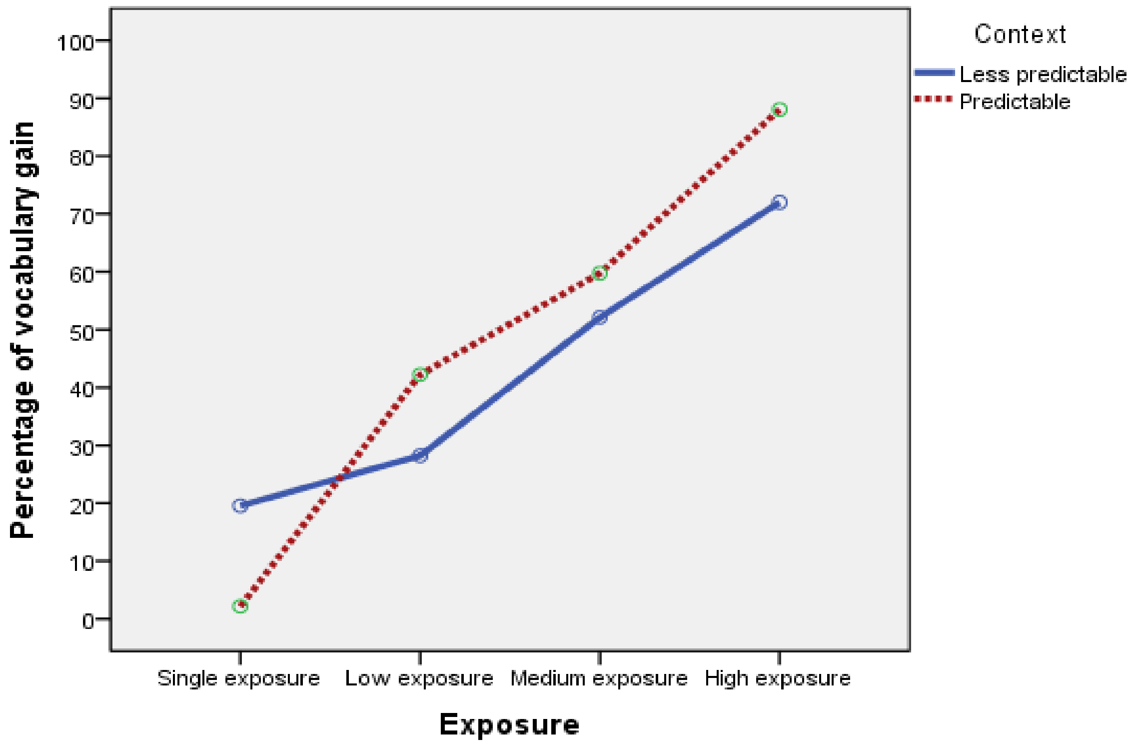

Figure 4. The interaction between exposure and predictability in form recognition

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 4 illustrates that highly predictable items yielded relatively better gains except for single exposure words. On the other hand, Figure 5 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 6). Overall, repetition was effective in all vocabulary gains but context predictability enhanced these gains, especially in the medium-exposure band.

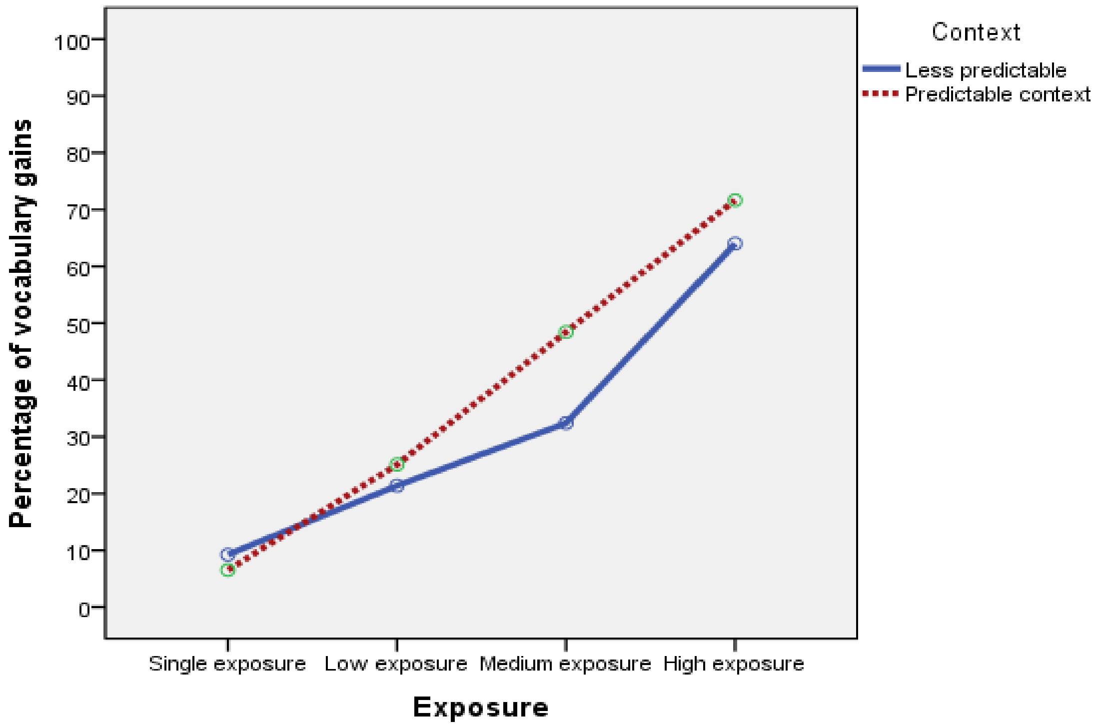

Figure 5.The interaction of exposure and context in meaning recognition

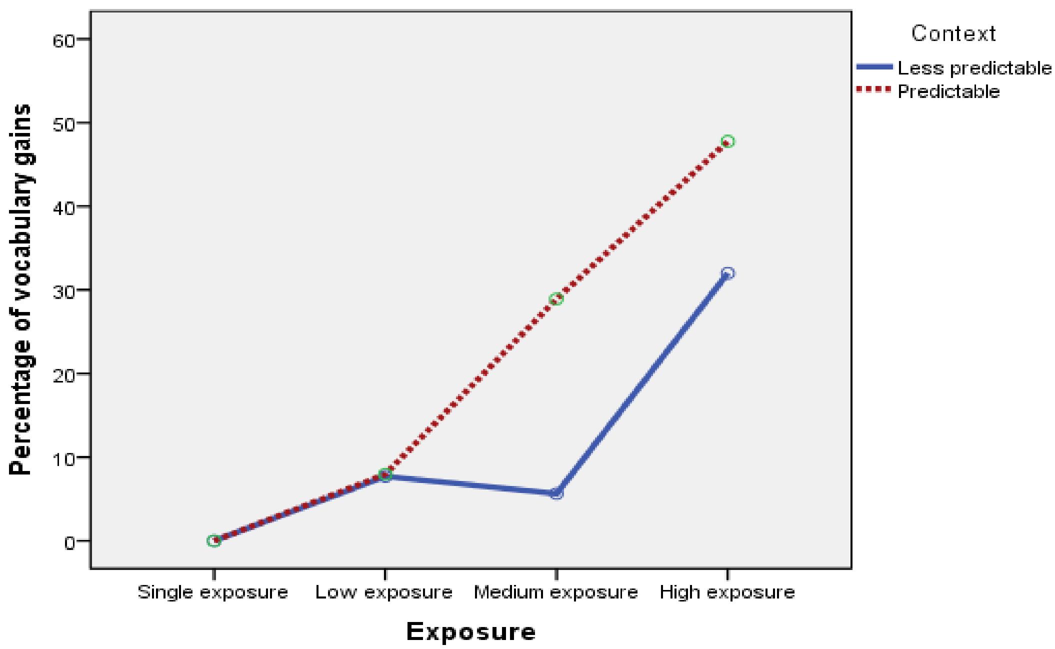

Figure 6. The interaction of exposure and context in meaning recall

Token-based processing and vocabulary learning

Moment-by-moment measures were analyzed in a regression model to test their prediction of successful learning. 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 total times 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 a significant positive relationship of form recognition with 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], $p = .084$. Similarly, there was a positive association between total reading times and meaning recognition (*OR* =1.33, 95% CI = [1.03, 1.72], $p = .029$). 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 %. Meaning recall was significantly predicted by total reading times (*OR* = 1.73, 95% CI = [1.14, 2.63], $p = .010$). 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 9 outlines the regression output for all vocabulary outcomes as a function of online reading times and token predictability.

Table 9

Token-based predictors of vocabulary gains

	Odds	*OR*	95% CI		*p*
<u>Form recognition</u>					
Intercept	*0.26*		0.075	0.93	.038 *
Total time		1.42	1.12	1.80	.004 **
Predictability		1.45	0.91	2.32	.116
<u>Meaning recognition</u>					
Intercept	*0.10*		.034	0.32	<.001 ***
Total time		1.33	1.03	1.72	.029 *
Predictability		2.81	1.81	4.34	< .001 ***
<u>Meaning recall</u>					
Intercept	*0.051*		0.08	0.30	.002 **
Total time		1.73	1.14	2.63	.010 *
Predictability		5.68	3.19	10.25	<.001***

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

Item-based processing and 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 10

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 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 10 shows that 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 11 and 12 summarize the predictors of meaning recognition and recall.

Table 11

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 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

Table12

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 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 (10-12) 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 and context effects.

General summary of results

Online reading patterns pointed to significant differences between attention to target items and familiar items. 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. 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.

Readers displayed the highest 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 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. 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.

DISCUSSION

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.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. Less skipping and more regressions indicated increased processing and reanalysis of target words, which may have supported the form-meaning mapping process.

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 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.

Predictors of 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. 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 8, 9 and 10).

The fact that summed total times were a strong predictor of learning outcomes after controlling for total exposure and predictability 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 and context cues. 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.

CONCLUSION

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.

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.

Limitations and further research

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 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.

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.

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.

REFERENCES

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.

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.

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.

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

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.

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

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

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.

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.

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

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.

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. (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.

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.

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.

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.

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.

Pellicer-Sánchez, A. (2015). Incidental L2 Vocabulary acquisition from and while reading. *Studies in Second Language Acquisition*, 1-34.

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.

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., 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.

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.

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.

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.

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.

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.

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.