Studies in Second Language Acquisition, 2015, **0**, 1–34.
doi:10.1017/S0272263115000224

INCIDENTAL L2 VOCABULARY ACQUISITION *FROM* AND *WHILE* READING

An Eye-Tracking Study

QA Ana Pellicer-Sánchez
University of Nottingham

Previous studies have shown that reading is an important source of incidental second language (L2) vocabulary acquisition. However, we still do not have a clear picture of what happens when readers encounter unknown words. Combining offline (vocabulary tests) and online (eye-tracking) measures, the incidental acquisition of vocabulary knowledge from reading and the online reading of unknown lexical items were examined. L2 English learners read a story containing unknown items while their eye movements were recorded. After eight exposures, L2 readers recognized the form and the meaning of 86% and 75% of the target nonwords, respectively, whereas they recalled the meaning of 55% of the nonwords. After three to four encounters, nonwords were read significantly faster, and by eight encounters they were read in a similar manner to previously known real words. Results also showed a positive relationship between new vocabulary learning outcomes and online reading, with longer reading times associated with higher vocabulary recall test scores. The study was also conducted with first language (L1) readers to provide baseline data for comparison. Results confirmed the L2 findings while also indicating an interesting L1-L2 distinction in the rate rather than in the ultimate outcome of the acquisition process.

Correspondence concerning this article should be addressed to Ana Pellicer-Sánchez, School of English, Trent building A82, University Park, University of Nottingham, Nottingham, NG7 2RD. E-mail: ana.pellicer-sanchez@nottingham.ac.uk

Vocabulary knowledge is essential for fluent second language (L2) use. Learners need a large amount of vocabulary to communicate successfully in a L2, with studies showing that, to reach successful comprehension of a wide range of written and spoken discourse, learners need to know around 6,000 to 7,000 and 8,000 to 9,000 word families, respectively (Nation, 2006).

A major concern of vocabulary research has therefore been to find the most effective way of expanding learners' lexical knowledge. Vocabulary learning studies have followed two main research foci: (a) those exploring the effect of *intentional learning* (i.e., learning that occurs when there is a particular intention to do so; Barcroft, 2009) and (b) those examining the effect of *incidental learning* (i.e., learning that occurs as a result of using language with no particular intention to learn a particular linguistic element; Schmitt, 2010).When these two approaches have been compared, the advantage of intentional learning through explicit teaching has been clear (e.g., Laufer, 2003; Lin & Hirsh, 2012). However, explicit teaching cannot account for the huge amount of words that learners need to know. This is when incidental acquisition from written and spoken input comes into play.

BACKGROUND

Incidental Second Language Vocabulary Acquisition from Reading

The majority of studies investigating incidental vocabulary learning have focused on the acquisition of new words from reading. Despite some evidence suggesting the small effect of reading for vocabulary learning (e.g., Laufer, 2005), there seems to be general agreement that reading is an effective tool for increasing learners' vocabulary knowledge (e.g., Krashen, 1989; Rott, 2007). Studies on incidental learning from reading have traditionally focused on examining the acquisition of new words and their meaning, usually by means of multiple-choice tests. One of the earliest studies in the L2 context was Pitts, White, and Krashen (1989). Using a multiple-choice meaning recognition test, these authors investigated participants' acquisition of the *nadsat* (Russian slang) vocabulary appearing in the first two chapters of *A Clockwork Orange*. Results showed that, though modest (6.4–8.1% gains), there was some vocabulary acquisition through reading. Similarly, Day, Omura, and Hiramatsu (1991) found that both Japanese high school and university students learnt a considerable number of new words from reading a short passage, as measured by a meaning-focused, multiple-choice test. Using the Vocabulary Levels Test format (Nation, 1983), Zahar, Cobb, and Spada (2001) found that their English as a L2 (ESL) participants learned about 8% of the target vocabulary from

reading a story. Higher percentages of vocabulary gains from reading were found by Horst (2005), with her participants acquiring more than half of the unfamiliar words that occurred in the reading materials. However, vocabulary knowledge in this study was measured by a yes-no test in which participants provided self-judgments of their lexical knowledge. These meaning-only and single-test studies showed some incidental acquisition of new words from reading, but they all suggest that the gains are relatively modest.

However, although this receptive form-meaning level of lexical mastery is perhaps the minimum amount of knowledge required to start using the newly acquired vocabulary (at least receptively), there are many other components of vocabulary knowledge that need to be mastered. Nation (2001) provided the most comprehensive list to date of all the different components of vocabulary knowledge, including aspects of *form* (e.g., written form, spoken form, word parts), *meaning* (e.g., form-meaning link, concepts and referents, associations), and *use* (e.g., grammatical functions, collocations, frequency). He also made a distinction between receptive and productive levels of mastery of each of these components. Nation and Webb (2011) highlight the need to address these aspects in vocabulary learning studies using a variety of tests to: (a) measure different types of knowledge learned and (b) measure the strength of that knowledge.

A small but growing number of studies have implemented this multicomponential approach and multitest approach in their designs. Horst, Cobb, and Meara (1998) assessed the amount of vocabulary learned from reading by L2 learners by means of a multiple-choice meaning recognition test and a word association test. Results showed that participants learned 22% of the words that could be learned, and they could also build associations between them. Waring and Takaki (2003) examined incidental acquisition of the meaning of new words at both the recognition and recall levels using a multiple-choice and a translation task, respectively. Their Japanese participants recognized the meaning of 42% of the target words on the immediate multiple-choice test. However, they were only able to provide a Japanese translation for 18% of the items. After 3 months, the meaning recognition score dropped to 24%, but the translation score dropped much more sharply to 4%, showing that recognition knowledge was not only easier to acquire but it was also retained better over time. Similar gains were found by Brown, Waring, and Donkaewbua (2008). After reading a set of graded readers, participants were able to recognize the meaning of 45% of the 28 target words, whereas they could only recall the meaning of 15% of those words. Recognition knowledge remained about the same after 3 months, whereas recall knowledge dropped significantly. Assessing these two same components, Rott (1999) examined the number of words that L2 learners of German acquired from reading short texts. Results of the immediate test also showed better recognition knowledge than recall, with participants

being able to recall the meaning of 45% of the target items and recognize the meaning of 61% of the words after six exposures. These percentages remained the same after 1 week but dropped to 34% and 48%, respectively, after 4 weeks.

The acquisition of spelling, meaning, and grammatical characteristics was examined in Pigada and Schmitt's (2006) study. After 1 month of extensive reading, their L2 learner of French showed some acquisition of 65% of the words tested, with spelling being the most strongly enhanced aspect and meaning and grammatical knowledge being enhanced to a lesser degree. Four components were examined by Pellicer-Sánchez and Schmitt (2010): form recognition, recall of grammatical class, meaning recognition, and meaning recall. After more than 10 exposures in an authentic novel, learners could recognize the meaning and form for 84% and 76% of the words, respectively, whereas they could recall the meaning and word class for 55% and 63% of the words, respectively.

The most striking example of using a multitest approach to vocabulary research is the study conducted by Webb (2005). Vocabulary gains from reading short sentence contexts were assessed by means of 10 different tests measuring five components (orthography, syntax, grammatical functions, association, and meaning-form) at the recognition and recall levels. Results of the first experiment showed that, after reading the 10 target words three times each in short sentence contexts, the gain percentages shown in all postreading vocabulary tests were higher than 73%. Each target item in this study was glossed, underlined, and written in bold, which it can be assumed positively influenced the retention rate. These same five components were examined in a follow-up study (Webb, 2007b). After reading short glossed sentence contexts, his Japanese English as a foreign language (EFL) participants gained a considerable amount of knowledge of all the components measured, with orthography and grammar recognition being the most enhanced aspects (74%) and production of syntagmatic association the least improved (38%). Using a similar methodology, these components were also examined in Webb's (2007a) study. After only one exposure in short contexts, participants showed acquisition of all components investigated, with all receptive scores between 40 and 70% and much lower productive scores (9–50%). These percentages increased with increasing frequency. Overall, the results of these studies showed substantial gains from reading and better acquisition rates at the recognition than at the production level. Webb's studies are a clear example of the advantages of using a multitest and multiaspect approach to vocabulary research. AQ1

Another major concern in vocabulary acquisition research has been the effect of frequency of exposure. Previous studies have shown the positive effect of frequency of exposure on the acquisition of new words from reading, with higher frequencies being associated with stronger gains (e.g., Waring & Takaki, 2003; Zahar et al., 2001). In the first language (L1) context,

Saragi, Nation, and Meister (1978) found that the minimum number of encounters for substantial vocabulary learning to occur was around 10. In the L2 context, Rott (1999) found that, even though after two exposures learners' vocabulary growth was significantly affected, gains were stronger after six exposures. Horst and colleagues (1998) found that target words needed to appear at least eight times for substantial gains to occur. Nation and Wang (1999) claimed that after 10 encounters words were more likely to be learned, although this did not guarantee acquisition. Pigada and Schmitt (2006) also found that by about 10+ exposures, there was a considerable increase in vocabulary learning. This 10+ figure was also confirmed by Pellicer-Sánchez and Schmitt (2010) and Webb (2007a). Results of these studies seem to suggest that, for reliable learning of several lexical aspects, words need to be met around eight to 10 times.

These and other studies (e.g., Grabe & Stoller, 1997; West, Stanovich, & Mitchell, 1993) have shown that L2 learners can acquire new words incidentally from reading, provided that they encounter the new vocabulary enough times within a limited time span. However, these previous studies assessed lexical knowledge using offline postreading tasks. Their findings, although informative, do not tell us much about what happens when learners find unfamiliar words while reading. This question has been addressed by other studies using think-aloud protocols (e.g., Bengeleil & Paribakht, 2004; Fraser, 1999; Huckin & Bloch, 1993; Paribakht & Wesche, 1999) or retrospective interviews (e.g., Godfroid & Schmidtke, 2013). Most of these studies have shown that, when encountering unknown words in reading, L2 learners try to figure out their meaning by means of lexical inferencing (i.e., "guessing the meaning of an unfamiliar word using available linguistic and other cues" [Bengeleil & Paribakht, 2004, p. 225]) and that context-based strategies are most typically used when generating their guesses. However, other studies have shown low rates of success from lexical inferencing (e.g., Nassaji, 2003).

These studies provide further insight into the process of vocabulary learning from reading. However, think-aloud protocols on their own capture only traces of the cognitive processes taking place (Huckin & Bloch, 1993). We still do not know how those unknown words are read in context, how that reading behavior changes as an effect of frequency, or how that reading behavior relates to learning rates. These questions can now be addressed by combining online eye-tracking methodology and offline reading tests.

Eye Movements and Word Recognition in Reading

The investigation of eye movements using eye-tracking methodology is a common research tool in psychology and psycholinguistics, and there

has recently been a growing interest in the use of this technique in applied linguistics research. Eye-tracking methodology has received special attention in reading research. Eye movements are "an inherent behavioural manifestation of the reading process in action" (Radach & Kennedy, 2004, p. 4). A vast number of research studies have investigated eye movements while reading and processing information.[1]

While reading, we move our eyes in a sequence of *saccades* (i.e., very rapid movements), and these saccades are interrupted by *fixations* (i.e., periods of relative stability). It is during eye fixations that visual information can be extracted (Radach & Kennedy, 2004). When reading, readers go back to read parts of the text that have already been read 10% to 15% of the time (i.e., they make a *regression*; Rayner, 1997).

In the examination of eye movements a distinction has been made between temporal measures (such as fixation durations), which are believed to be indicators of processing load, and spatial measures (such as fixation position and saccade amplitudes), which are indicators of the direction and the sequence of processing (Radach & Kennedy, 2004). Some of the most frequently used measures of processing time include

- *First fixation duration*, or the duration of the first fixation on a word or region of interest;
- *First pass reading time* or *gaze duration*, which is the sum of all fixations made on a word or region of interest before exiting the area or word either to the left or to the right;
- *Fixation count*, or the number of all fixations made on a word or region of interest; and
- *Total reading time*, or the sum of all fixation durations made on a word or region of interest.

Figure 1 shows these four measures and hypothetical patterns of eye movements in a sample sentence from the reading text used in this study.

Deciding which one is the best measure of processing time (when the unit of analysis is a single word) is a controversial issue (Rayner, 1998). Rayner (1998) points out that using a single measure of processing time per word is inappropriate, and he recommends the use of several measures as a way of drawing reasonable inferences about the reading processes. Earlier measures of first fixation duration and gaze duration are believed to tap into initial lexical access, whereas later measures of the number of fixations and total reading time are believed to reflect higher order processes like semantic integration (Libben & Titone, 2009).

When we read, about 30% of the words in a text are not fixated (Rayner & Juhasz, 2004). Short words are skipped more frequently than longer words (Rayner & McConkie, 1976), and high-frequency words are

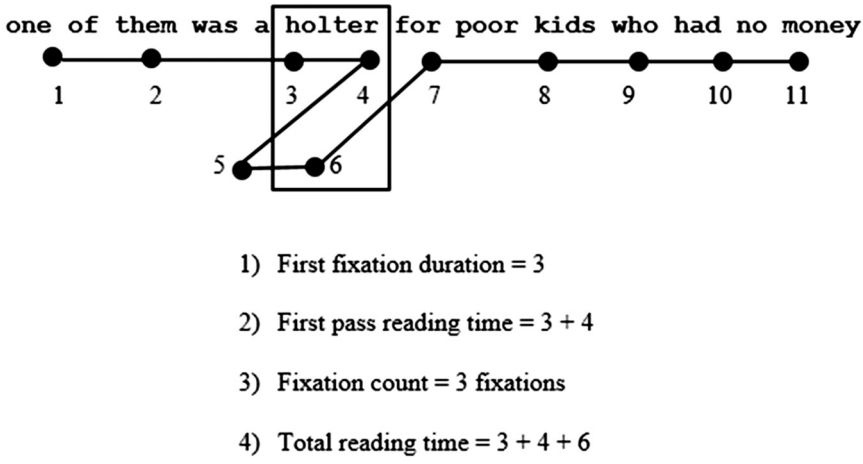

Figure 1. Sample sentence with eye-tracking measures (format of the figure based on Siyanova-Chanturia et al., 2011).

skipped more frequently than low-frequency words (Rayner, Sereno, & Raney, 1996). The fact that words are not fixated does not mean that they are not processed (Rayner, 1998). Due to the speed that characterizes saccades, there is no useful information acquired during saccades. Instead, readers acquire information during fixations (Wolverton & Zola, 1983). Fixation durations are related to the ease or difficulty with which words in a text are comprehended (Rayner, 1997, 1998; Rayner & Pollatsek, 1989). When readers encounter a novel word, they spend more initial processing time on those novel words than on familiar words (Chaffin, Morris, & Seely, 2001; Godfroid, Boers, & Housen, 2013; Williams & Morris, 2004).

These reading patterns are influenced by a number of lexical, semantic, and contextual features.[2] Some of these lexical features include orthographic regularity (e.g., Radach, Inhoff, & Heller, 2004); orthographic familiarity (e.g., White & Liversedge, 2004); length, frequency, and predictability (e.g., Kliegl, Grabner, Rolfs, & Engbert, 2004); word frequency (e.g., Rayner, 1998; Rayner et al., 1996); word familiarity (e.g., Williams & Morris, 2004); age of acquisition (e.g., Juhasz & Rayner, 2006); and lexical ambiguity (e.g., Rayner & Frazier, 1989). In general, increased regularity and familiarity, higher predictability and frequency, as well as shorter length, have been found to lead to shorter fixation durations and fewer fixations. Another factor having an important effect on the reading of fixated words is repetition. Hyönä and Niemi (1990) investigated the process of repeated reading and found that repeated exposures to the same text led to a decrease in fixation durations, number of fixations, and number of regressions. Results showed that, with increased repetitions, reading times on fixated words decreased. This decrease in

reading times in repeated reading is explained by the authors as an effect of the increased familiarity of both the surface features and the content of the text. When readers are exposed to an already-read text, both the visuographic features and the content of the text become more familiar, which leads to the decrease in reading times. Hyönä and Niemi (1990) interpret this finding as a suggestion that both low-level processes (i.e., visuographic features of the text) and higher level comprehension processes are responsible for eye guidance during reading, supporting the view that readers' eyes are guided by both visual and cognitive information.

Eye Movements in Vocabulary Acquisition Research

Eye-tracking methodology has been used to investigate several phenomena in L1 and L2 reading, such as sentence processing (e.g., Altarriba, Kroll, Scholl, & Rayner, 1996), the processing of formulaic language (e.g., Siyanova-Chanturia, Conklin, & Schmitt, 2011), and noticing of corrective feedback (e.g., Smith, 2012). However, only a few studies have used eye-tracking methodology to investigate the process of vocabulary learning from reading.

Chaffin and colleagues (2001) examined L1 readers' eye movements when learning meanings of novel words from sentence contexts. They found longer initial and total reading times for novel words as compared to high-familiarity words. However, they did not use any postreading measure to assess whether participants had indeed learned the meaning of those novel items.

Williams and Morris (2004) measured both participants' eye movements while reading words of different degrees of familiarity in sentence contexts and participants' vocabulary learning from reading by means of postreading vocabulary tests. Overall, results showed a systematic relationship between online processing patterns and memory for new words. However, they found a different and conflicting effect of early and late measures on postreading vocabulary scores. Shorter first pass reading times were associated with higher accuracy in the vocabulary tests, whereas longer second pass reading times were shown for words that were correctly answered in the vocabulary test. This negative relationship between first pass reading time and vocabulary learning scores, which could not be accounted for by the authors, has been explained as a possible effect of other confounding factors such as word length (Godfroid et al., 2013). One main limitation of this study is that it investigated reading in short sentence contexts and, because vocabulary learning was only a secondary focus of the study, it was measured by means of a two-choice synonym recognition test, using only an immediate measure of vocabulary retention.

In the L2 context, Godfroid and colleagues (2013) also used eye-tracking to examine the process of learning unknown vocabulary from reading. Twenty-eight EFL students read 20 short paragraphs containing pseudowords (e.g., *paniplines*) while their eye movements were recorded. After the reading task, they completed a multiple-choice gap-filling exercise. Participants were presented with the experimental paragraphs they had seen in the reading task with the difference that the target pseudoword had been removed. Their task was to fill the gap with one of the 18 options provided. They found that total reading time was a significant predictor of the probability of posttest recognition, with longer reading times being associated with better recognition scores.

In addition to the scarcity of studies using eye-tracking to investigate vocabulary acquisition from reading, most of the studies available used vocabulary tests that did not provide a good representation of the depth of vocabulary knowledge that accrues from reading. Furthermore, none of these studies explored the effect of repetition in the online reading of the unknown items.

Following a multicomponential and multitest approach and combining online measures of eye-tracking and offline vocabulary tests, the present study examined both the incidental acquisition of knowledge of unknown items (word form and meaning) from reading and the online reading of those items in context. Participants read a story containing unknown vocabulary while their eye movements were recorded. To explore the role of frequency of exposure in participants' reading of unknown vocabulary, all target items were repeated eight times. After the reading, participants completed several vocabulary tests to assess their knowledge of the unknown vocabulary. Another component of vocabulary knowledge that was examined in this study was the degree of certainty of participants' responses. It seems obvious to claim that the better or more consolidated one's knowledge of a word, the higher the degree of certainty of that knowledge. Wesche and Paribakht (2000), in their investigation of different text-based vocabulary exercises using think-aloud protocols, found that in some cases learners reported greater certainty in their knowledge of some target words. Confidence ratings have indeed been previously used in the implicit-explicit learning literature as a way of dissociating conscious and unconscious knowledge of grammar (Rebuschat, 2013). However, this subjective measure appears not to have been examined in vocabulary studies in a systematic way.

So far, research on vocabulary acquisition from reading has been able to shed light on the quantity and quality of words that are acquired incidentally *from* reading. However, by combining both online and offline measures, we can also examine how unknown items are recognized *while* reading. The combination of both methods of assessment should provide a fuller account of L2 learners' reading behavior and of their incidental vocabulary acquisition from reading.

The following questions were addressed:

1. Do L2 learners acquire vocabulary knowledge incidentally from reading, as measured by the offline vocabulary tests?
2. How does the reading of unknown items in context change across several encounters, as measured by the online measures?
3. Is there a relationship between the online reading of lexical items and vocabulary knowledge?

To provide a fuller examination of the process of L2 vocabulary learning from reading, the study was also conducted with L1 readers. The comparison with L1 baseline data allows us to examine whether the reading patterns observed in the L2 acquisition process are also found with L1 readers. Ultimately, this comparison allows us to find out whether L2 readers benefit from the reading activity in a similar way to L1 readers.

METHOD

Participants

Thirty-seven L2 speakers of English from various language backgrounds and 36 L1 speakers of English initially participated in this study. Due to cases of drift (i.e., imprecise eye movements indicating a deterioration of the calibration over time) in the areas under examination, data from 14 L2 participants and 11 L1 participants were discarded. Data from 23 L2 participants (10 males; 13 females) and 25 L1 participants (1 male; 24 females) were included in the analysis.

L1 participants were undergraduate students at a U.K. institution, and their ages ranged from 19 to 21 years (M = 19.4). L2 participants were postgraduate students and postdoctoral researchers at a U.K. university. They were from 11 different language backgrounds (nine participants from alphabetic languages; six from logographic languages; and five from syllabic languages or abugidas). Their ages ranged from 22 to 42 years (M = 27). They all had spent a minimum of 12 months and a maximum of 6 years living in an English-speaking environment (M = 2.4 year; SD = 1.7 years). They were advanced learners who had met the university entry requirement of English proficiency (6.0 or above on the International English Language Testing System [IELTS] or equivalent examination). At the beginning of the experiment L2 participants completed a self-rating test of proficiency (on a scale from 1 to 10) in English.[3] Table 1 shows that the mean values for all skills were above 7, and all participants rated their reading skills at 7 or above. L2 participants received

Table 1. Self-rating proficiency scores

Skills	Max	Min	*M*	*SD*
Reading	10	7	8.74	1.01
Writing	10	5	7.78	1.24
Listening	10	6	8.26	1.10
Speaking	10	6	7.87	1.22

a small compensation for their participation, and L1 participants received course credit.

Reading Materials

A short story (2,300 words) was written for the study. Because text comprehension influences vocabulary learning from reading (e.g., Pulido, 2004), vocabulary in the story was controlled to ensure that potential acquisition of the unknown items would not be hindered by lack of knowledge of the remaining items in the text. Of the words in the story, 96.82% belonged to the 3,000 most frequent words of the British National Corpus (BNC; determined by Compleat Lexical Tutor; Cobb, n.d.). Only four words (.17%) were from the 5,000 to 9,000 frequency bands. These were considered adequate percentages to ensure participants' comprehension.

Six nonwords and six control words (real known words) were inserted in the story, each repeated eight times. Having 48 unknown items (2.09%) in the text still maintained the 98% coverage that has been claimed to enable adequate comprehension (Hu & Nation, 2000). This percentage of nonwords ensures comprehension even for participants who may not learn any of the target items throughout the eight encounters. Nonwords were spread to ensure a balanced distribution of unknown items throughout the story. The story was presented on a computer screen (Courier New font, size 18) and divided over 25 screens. The text presented on each screen had a similar length (eight lines; 82–103 words). All screens contained a maximum of two nonwords (23 screens contained two nonwords, and the first and last screens contained only one nonword). The position of target and control items in the text was carefully controlled so that none of the items would be in initial or final position in a line or sentence, because previous research has shown that the first fixation on a line tends to be longer (e.g., Rayner, 1977).

A true-false comprehension test containing 12 statements assessed participants' comprehension.[4] None of the target items appeared in any of the comprehension questions. A shorter (five screens, 423 words)

but similar story (also containing unknown items) and comprehension test were created for a practice session.

Target Items

To ensure lack of previous knowledge of the target items, nonwords (i.e., invented letter strings that look like real words in English) were used. Nonwords were selected from the list developed by Meara and his colleagues and available from Compleat Lexical Tutor (Cobb, n.d.) and modified to suit the required length (two syllables, six letters). They all replaced high-frequency (1,000–3,000 from the BNC), concrete nouns in the text.

After an initial pilot with 10 native and 10 nonnative speakers of English, the candidate nonwords for the study were: *holter* "house," *berrow* "bowl," *bancel* "prisoner," *cambul* "ring," *twoser* "noise," and *soters* "clothes." To make sure that all nonwords were equally guessable from the contexts in which they appeared, they were again piloted with 87 native speakers of English divided in eight groups. Group 1 read the first context in which each of the nonwords appeared (including the nonword sentence, the previous sentence, and the following sentence), Group 2 read contexts 1 and 2, Group 3 read contexts 1 through 3, and so forth. Participants were asked to read the paragraphs and guess the meaning of the nonwords. This pilot study made it possible to check those cases in which several meanings should be considered correct for the same item. For example, *bancel* was guessed as *criminal* and *prisoner*, and both meanings fit all contexts. These two options were considered accurate guesses. A similar case was found for *holter.* It was initially inserted as *house*, but other possible meanings were *workhouse*, *orphanage*, or *shelter*. Results also showed that the vast majority of participants provided the same accurate guesses (*holter*: 93% of the participants, *bancel*: 98%, *twoser*: 97%, *soters*: 98%, *cambul*: 100%, *berrow*: 98%). These percentages were considered a good indication of the similar degree of informativeness of the contexts in which target words were embedded.

To compare the reading behavior of these unknown words with that of already known words and to make sure that any effect observed was not a practice effect that would have also been observed with known words, six control items were included in the story and also repeated eight times. These were real words with the same characteristics as the targets (nouns, six letters, and two syllables). They were all high-frequency words (1,000–3,000 from the BNC): *garden*, *master*, *mother*, *dinner*, *worker*, and *secret*.

Measurement Instruments

For the online measures, participants' eye movements were monitored using a head-mounted SMI Eyelink I eye-tracker manufactured by SR Research. The following four measures were examined: first fixation duration, gaze duration, number of fixations, and total reading time.

For the offline measures, three vocabulary tests were used: form recognition, meaning recall, and meaning recognition. The first vocabulary test assessed participants' ability to recognize the correct form of the target items. A multiple-choice test presented four different options, and participants were asked to select the correct spelling of the target items. The second test measured participants' ability to recall the meaning of the target items. Participants were shown the target items one-by-one and were asked to say everything they knew about the meaning of the item. A third and easier measure of the form-meaning link (i.e., meaning recognition) was included to capture knowledge below the level of meaning recall (Schmitt, 2010). A multiple-choice item was designed for each word, consisting of the target item and five possible options: three distractors, the correct meaning, and an "I don't know" option. Careful attention was given to the design of distracters, which were all semantically related to the content of the story (otherwise their discrimination would have been too easy) and were all of the same word class. In the three tests, participants also had to indicate on a scale from 1 to 4 how certain they were of their responses (1 = *very uncertain*, 4 = *very certain*; see Appendix).

Procedure

Experiments took place individually in a psycholinguistics laboratory. At the beginning of the session L2 participants completed a language background questionnaire. After setting up the eye-tracker, participants were asked to read the story as naturally as possible for comprehension. They were aware that there would be postreading comprehension questions, but they were not aware of the presence of nonwords in the story. After the explanation of the procedure, a nine-point grid calibration was completed. Before completing the experimental reading task, participants read a practice story and answered five comprehension questions. Another calibration was completed between the practice and experimental trials, and another halfway through the experiment. Before each new screen, a fixation point appeared at the left of the screen. After participants had fixated the point and a calibration check was made, the story appeared on the screen. To proceed from one screen to the next, participants had to press the "Enter" button on the keyboard.

During the reading participants could not go back to previously read screens. After reading the story, the true-false comprehension questions appeared one-by-one on the screen. Participants had to respond by pressing the "yes" and "no" buttons on the keyboard. After the comprehension questions, participants completed the vocabulary tests. The first (form recognition) and last (meaning recognition) tests were completed individually in paper format. The second test (meaning recall) was conducted by means of a personal interview. The researcher showed participants A3-size index cards with each of the target items and asked them to say everything they knew about their meaning. The whole procedure lasted around 45 min. The three vocabulary tests were scored using the same system (1 = correct response, 0 = incorrect response). Partially correct responses were not scored. Given the immediate nature of the posttest and the small number of items in the study, a stricter approach to the scoring of responses was considered a more reliable indication of true vocabulary gains in the study.

L2 participants were invited to attend a second session. A delayed posttest session took place in the same location 2 weeks after the first session. Participants were not aware of the content of the second session. Only the offline tests were administered, following the procedure outlined previously. The whole session lasted around 15 min.

The exact same procedure was followed with L1 and L2 participants, with the only difference being that, for practical limitations, L1 participants could not complete the delayed testing session. Only results of the immediate test will be reported for L1 readers.

RESULTS AND DISCUSSION

Offline Measures

Results of the comprehension test showed proper comprehension of the text by both L2 and L1 participants (L2 participants: M = 11.22, SD = 1.09, Min = 8; L1 participants: M = 11.32, SD = .85, Min = 9). Results of independent-samples t tests showed that there was no significant difference between L1 and L2 scores, $t(46) = .37$, $p = .719$. Both L1 and L2 readers showed very similar levels of comprehension.

A one-way within-groups ANOVA was conducted to compare participants' performance on the three vocabulary tests (see means in Table 2). Results indicated that there was an overall significant difference in the scores of the vocabulary tests, $F(2, 44) = 12.85$, $p < .001$. Post hoc tests using the Bonferroni correction showed that, after eight encounters, L2 participants were able to recognize the form of the target items significantly better than to recall their meaning ($p < .001$) and that they were able to recognize their meaning significantly better than to recall

Table 2. Postreading vocabulary test scores

Vocabulary test	L2 participants				L1 participants	
	Immediate posttest		Delayed posttest		Immediate posttest	
	M[a] *N* items	*M* %	*M* *N* items	*M* %	*M* *N* items	*M* %
Form recognition	5.13 (1.01)	85.50	5.14 (0.66)	85.71	5.48 (0.87)	91.30
Meaning recall	3.65 (1.58)	60.87	3.29 (1.94)	54.76	3.92 (1.44)	65.30
Meaning recognition	4.70 (1.19)	78.26	4.50 (1.29)	75.00	5.20 (0.87)	86.60

[a] Max = 6.

it ($p < .001$). The difference between form recognition and meaning recognition was not significant ($p = .459$). Interestingly, out of the total number of unknown responses in the meaning recall test (54 unknown responses), 26% of the responses (14 cases) were ones in which the meanings were recalled correctly but linked to the wrong word. This could be an indication that participants had some initial productive knowledge of the meaning of the target items in place but failed to make the appropriate form-meaning link.

Results of a one-way within-groups ANOVA with the L1 baseline data also showed that there was an overall significant difference in the vocabulary test scores, $F(2, 48) = 30.28$, $p < .001$, with meaning recall being the most difficult aspect to be acquired and the two recognition aspects being the easiest. Post hoc comparisons using the Bonferroni correction showed that, after eight encounters, L1 participants were able to recognize the form and the meaning of the target items significantly better than to recall their meaning ($p < .001$). The difference between form recognition and meaning recognition was not significant ($p = .550$).

Independent-samples *t* test analyses showed that there was no significant difference between L1 and L2 participants in the acquisition of these three components of lexical mastery (form recognition: $t[46] = 1.28$, $p = .206$; meaning recognition: $t[46] = .61$, $p = .543$; meaning recall: $t[46] = 1.69$, $p = .097$).

A nonparametric Friedman test (data not normally distributed) was used to compare participants' degree of certainty on the three different tests (see mean values in Table 3). Results showed that there was a significant difference in certainty scores across the three tests, $\chi^2(2) = 10.92$, $p = .004$. A post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied (resulting significance level at $p < .017$), and results showed that the difference between L2 participants' certainty for meaning recognition was significantly higher than for meaning recall, $Z = -3.45$, $p < .001$, and for form recognition, $Z = -3.48$; $p < .001$. These results suggest that meaning recognition is not

Table 3. Degree of certainty in vocabulary test responses

Certainty of vocabulary knowledge	L2 participants				L1 participants			
	M[a]	*SD*	Min	Max	*M*[a]	*SD*	Min	Max
Certainty of form recognition	3.14	0.95	1	4	3.24	0.91	1	4
Certainty of meaning recognition	3.59	0.60	2	4	3.50	0.72	1	4
Certainty of meaning recall	3.04	1.19	0	4	3.06	0.95	1	4

[a] 1 = *very uncertain*; 4 = *very certain*.

only one of the easiest aspects to acquire (when compared to meaning recall), but it is also a type of knowledge for which participants seem to be more certain. This is not surprising, given the multiple-choice nature of the meaning recognition test. In contrast, meaning recall is not only a difficult aspect to acquire, but, even when acquired, certainty for that knowledge may be more difficult to achieve.

Similarly, results of the Friedman test with L1 baseline data showed that there was a significant difference in participants' certainty levels (see mean values in Table 3), $\chi^2(2) = 17.61$, $p < .001$. A post hoc analysis with Wilcoxon signed-rank tests (using the Bonferroni adjusted alpha value) showed that L1 readers were also significantly more certain about their ability to recognize the meaning of the words than about their ability to recall the meaning, $Z = -4.57$, $p < .001$, and their ability to recognize the correct form, $Z = -2.42$, $p = .016$. Meaning recall was also the component with the lowest degree of certainty for L1 readers.

These gains come from an immediate test, and thus it is not surprising that participants were able to show knowledge of the target items. Results of the delayed test with L2 participants should give a better indication of durable lexical learning. Only 14 L2 participants out of the initial 23 completed the delayed test. The generalizability of results is therefore less robust. Results of the one-way within-subjects ANOVA showed a significant difference in the scores of the delayed tests (see mean values in Table 2), $F(2, 26) = 10.88$, $p < .001$. Results of the post hoc comparisons with Bonferroni correction showed a very similar pattern to that of the immediate test, with scores for meaning recall being significantly lower than meaning recognition and form recognition ($p < .001$). Results of paired-samples *t* tests also showed that the differences between the immediate and delayed posttests were not significant (form recognition: $t[13] = -.37$, $p = .720$; meaning recall: $t[13] = .29$, $p = .775$; Meaning recognition: $t[13] = .82$, $p = .426$).

The possibility of test effect needs to be considered. The vocabulary knowledge shown in the delayed test may have been a consequence not only of the vocabulary learned incidentally from reading but also of the additional focused exposure to the target items in the immediate test. However, because participants did not know about the content of the

delayed test and they did not encounter the target items in the 2 weeks between the two testing sessions, results are still a good indication of durable learning 2 weeks after the initial exposure.

In response to the first research question, results of the present study have shown that, after having encountered the unknown items eight times in a text, L2 and L1 participants learned a considerable number of those items and that, for L2 readers, this percentage of learning seemed to persist after 2 weeks. Results of the offline measures have shown no significant differences between the behavior of L1 and L2 readers, suggesting that the advanced L2 readers in this study benefitted from the reading activity in the same way as L1 readers. L1 and L2 participants acquired a similar amount and type of vocabulary from reading and were similarly certain about the acquired knowledge.

Online Measures

Participants' reading behavior for target and control items was analyzed. Four measures were examined: first fixation duration, gaze duration, number of fixations, and total reading time. Single fixation durations shorter than 100 ms and longer than 800 ms were discarded. Fixation counts greater than or equal to 10 were also discarded (Morrison, 1984). This resulted in the loss of 5% of the L2 data (218 fixations out of the total 3,824 fixations) and 6.5% of the L1 data (227 fixations out of the total 3,262 fixations).[5]

The Kolmogorov-Smirnov test of normality showed that the data was not normally distributed. Results of nonparametric Kruskal-Wallis tests demonstrated that, as shown by previous studies (e.g., Chaffin et al., 2001; Godfroid et al., 2013; Williams & Morris, 2004), in the case of L2 readers, mean reading times for targets were significantly longer than for controls in all four measures examined (first fixation duration: $\chi^2[1] = 21.61$, $p < .001$; gaze duration: $\chi^2[1] = 35.59$, $p < .001$; number of fixations: $\chi^2[1] = 65.24$, $p < .001$; total reading time: $\chi^2[1] = 80.93$, $p < .001$; see mean values in Table 4).

Kruskal-Wallis tests were also conducted to explore the effect of repetition on the four eye-tracking measures. When examining reading times across the eight encounters, it was observed that both controls and targets experienced a decrease. For targets, the effect of repetition was significant in all measures (gaze duration: $\chi^2[7] = 34.38$, $p < .001$; number of fixations: $\chi^2[7] = 67.13$, $p < .001$; total reading time: $\chi^2[7] = 76.46$, $p < .001$), except for first fixation duration, $\chi^2(7) = 13.38$, $p = .063$, whereas for controls it was only significant in two measures, number of fixations, $\chi^2(7) = 24.87$, $p < .001$, and total reading time, $\chi^2(1) = 17.43$, $p = .015$. There was no significant effect of repetition among the eight frequency

Table 4. Mean reading times for targets and controls across encounters (L2 participants)

	First fixation duration (ms)		Gaze duration (ms)		Number of fixations		Total reading time (ms)	
Repetition	Targets	Controls	Targets	Controls	Targets	Controls	Targets	Controls
1	283.30 (70.19)[a]	239.64 (51.17)	421.83 (111.68)	293.82 (71.52)	2.61 (0.85)	1.52 (0.48)	703.16 (228.31)	369.87 (110.34)
2	259.19 (45.26)	208.08 (35.11)	372.17 (121.36)	259.00 (56.72)	2.28 (0.70)	1.46 (0.39)	568.07 (166.78)	324.32 (88.11)
3	256.73 (42.98)	229.71 (54.11)	326.57 (93.42)	265.82 (65.85)	2.30 (0.78)	1.47 (0.37)	596.98 (224.69)	359.03 (101.19)
4	252.25 (52.13)	226.12 (42.01)	309.20 (94.37)	262.72 (50.46)	1.96 (0.54)	1.51 (0.39)	490.96 (162.62)	357.54 (112.73)
5	252.12 (50.96)	230.66 (36.97)	296.08 (78.13)	253.02 (51.17)	1.54 (0.42)	1.22 (0.24)	394.59 (114.16)	303.47 (68.30)
6	246.48 (51.58)	225.35 (36.38)	309.17 (76.44)	271.30 (59.55)	1.60 (0.42)	1.29 (0.36)	395.54 (119.98)	319.39 (92.06)
7	230.19 (31.93)	232.19 (31.87)	296.85 (92.41)	255.99 (46.09)	1.60 (0.42)	1.14 (0.31)	390.19 (95.00)	282.59 (66.18)
8	235.25 (47.69)	235.46 (51.51)	275.45 (60.96)	270.91 (55.27)	1.34 (0.44)	1.30 (0.48)	329.96 (110.67)	324.22 (85.45)
M	251.93	228.40	325.92	266.57	1.90	1.36	483.68	330.06

*Note. SD*s provided in parentheses.

groups for first fixation duration, $\chi^2(7) = 9.30$, $p = .232$, or gaze duration, $\chi^2(7) = 5.38$, $p = .614$ (Table 4).

Both targets and controls experienced a decrease in reading times in some of the measures examined. This was expected due to a general repetition effect, as found by previous studies (e.g., Hyönä & Niemi, 1990; Rayner, Raney, & Pollatsek, 1995). However, the decrease for target items was greater and significant in a higher number of measures.

Results of Kruskal-Wallis tests with the L1 baseline data showed that, as was the case with L2 learners, mean reading times for targets were significantly longer than for controls in all four measures examined (first fixation duration: $\chi^2[1] = 25.33$, $p < .001$; gaze duration: $\chi^2[1] = 33.31$, $p < .001$; number of fixations: $\chi^2[1] = 67.18$, $p < .001$; total reading time: $\chi^2[1] = 76.43$, $p < .001$; see mean values in Table 5). Kruskal-Wallis tests were also run to examine the effect of repetition on the four eye-tracking measures. Both controls and targets experienced a decrease across the eight encounters. For targets, the effect of repetition was significant in all measures (first fixation duration: $\chi^2[7] = 22.75$, $p = .002$; gaze duration: $\chi^2[7] = 56.32$, $p < .001$; number of fixations: $\chi^2[7] = 73.42$, $p < .001$; total reading time: $\chi^2[7] = 72.66$, $p < .001$), whereas for controls it was significant in three out of the four measures (gaze duration: $\chi^2[7] = 15.11$, $p = .035$; number of fixations: $\chi^2[7] = 36.81$, $p < .001$; total reading time: $\chi^2[7] = 37.63$, $p < .001$; first fixation duration: $\chi^2[7] = 6.12$, $p = .526$; Table 5).

Post hoc comparisons were run with Mann-Whitney tests applying the Bonferroni correction (.05/7 = .007). The fixation times and number of fixations at the first encounter were compared to fixation times and number of fixations at each subsequent encounter. Results showed that, in the case of L2 readers, the decrease in gaze duration for targets started to be significant from the third encounter, $Z = -3.18$, $p = .002$, $r = .47$. In the case of number of fixations and total reading time, it was not until the fourth encounter that the effect of frequency of exposure started to be significant (number of fixations: $Z = -2.85$, $p = .004$, $r = .42$; total reading time: $Z = -3.59$, $p < .001$, $r = .53$). Overall, after three to four encounters, we found a significant decrease in three of the four measures examined. The decrease in the number of fixations for control words was not significant until the seventh encounter, $Z = -3.64$, $p < .001$, $r = .54$, and total reading time for controls was not significant until the fifth encounter, $Z = -2.97$, $p = .003$, $r = .44$.

Post hoc comparisons with Mann-Whitney tests with the L1 baseline data (adjusted alpha value = .007) showed that, for targets, the decrease in first fixation duration started to be significant after the fourth encounter, $Z = -2.81$, $p = .005$, $r = .40$. However, in the other three measures examined, the effect of frequency started to be significant earlier. The effect in gaze duration started to be significant from the third encounter, $Z = -2.70$, $p = .007$, $r = .38$, and, in the case of number of fixations and

Table 5. Mean reading times for targets and controls across encounters (L1 participants)

	First fixation duration (ms)		Gaze duration (ms)		Number of fixations		Total reading time (ms)	
Repetition	Targets	Controls	Targets	Controls	Targets	Controls	Targets	Controls
1	240.59 (55.04)[a]	191.37 (43.58)	386.16 (130.67)	230.82 (62.95)	2.64 (0.89)	1.31 (0.35)	626.43 (218.59)	248.35 (67.72)
2	216.13 (44.34)	198.87 (28.32)	288.91 (75.65)	219.77 (37.05)	1.79 (0.57)	1.11 (0.31)	395.12 (155.87)	217.28 (68.92)
3	214.14 (43.14)	200.96 (65.74)	288.53 (85.54)	220.31 (77.34)	1.94 (0.54)	1.07 (0.45)	438.16 (162.99)	208.16 (94.90)
4	222.29 (39.29)	194.40 (30.76)	249.05 (58.34)	235.98 (57.50)	1.47 (0.61)	1.28 (0.40)	325.73 (157.15)	252.13 (76.47)
5	198.85 (32.72)	192.32 (35.20)	227.91 (48.35)	202.54 (43.84)	1.56 (0.56)	0.94 (0.25)	317.84 (135.53)	183.63 (54.28)
6	219.90 (33.84)	197.06 (32.66)	250.68 (59.69)	220.21 (48.36)	1.29 (0.54)	1.15 (0.28)	283.47 (154.44)	225.28 (82.25)
7	190.08 (26.04)	180.77 (32.63)	211.49 (42.57)	187.92 (33.40)	1.30 (0.49)	0.85 (0.30)	266.29 (138.41)	154.48 (66.04)
8	202.40 (39.98)	195.67 (34.06)	214.93 (55.51)	221.34 (64.34)	1.07 (0.44)	0.98 (0.32)	217.55 (105.92)	192.08 (81.96)
M	213.05	193.93	264.71	217.36	1.63	1.09	358.82	210.17

*Note. SD*s provided in parentheses.

total reading time, a significant effect was observed after the first encounter (number of fixations: $Z = -3.25$, $p = .001$, $r = .50$; total reading time: $Z = -3.74$, $p < .001$, $r = .53$). It only took L1 readers one or two encounters for their reading behavior to show a significant effect of repetition in three of the four measures examined. The decrease in the number of fixations and total reading time for controls was not significant until the fifth encounter (number of fixations: $Z = -4.14$, $p < .001$, $r = .59$; total reading time: $Z = -3.60$, $p < .001$, $r = .51$).

The effect of repetition on reading behavior was also explored by means of nonparametric correlation analyses. Results for L2 participants in Table 6 showed that the same pattern of results was found for controls, with significant, negative, and small correlations in only two measures (i.e., number of fixations and total reading times). For targets, results showed a significant negative correlation between frequency of exposure and all four measures examined, with a higher number of exposures being associated with shorter reading times and a lower number of fixations. Spearman correlation coefficients clearly showed that the negative relationship between frequency of exposure and reading times was stronger for the target words. Results of correlation analyses suggest a clearer and stronger effect of frequency of exposure for targets, both in terms of the strength of the correlation and in terms of the number of measures for which a significant correlation was found.

In line with L2 findings, results showed that, in the case of L1 readers, both targets and controls experienced a decrease in reading times in some of the measures examined, with the decrease for target items

Table 6. Nonparametric correlations between frequency of exposure and reading measures

	Targets		Controls	
Rep-Reading measure	Correlation coefficient r_s	Sig. (2-tailed)	Correlation coefficient r_s	Sig. (2-tailed)
L2 participants				
Rep-First fixation duration	–.261	.000*	.116	.115
Rep-Gaze duration	–.389	.000*	–.059	.428
Rep-Number of fixations	–.583	.000*	–.277	.000*
Rep-Total reading time	–.603	.000*	–.208	.005*
L1 participants				
Rep-First fixation duration	–.240	.001*	–.022	.762
Rep-Gaze duration	–.496	.000*	–.140	.048
Rep-Number of fixations	–.570	.000*	–.270	.000*
Rep-Total reading time	–.574	.000*	–.291	.000*

Note. * = significant at the $p < .013$ level (Bonferroni correction applied). Sig. = significance; Rep = repetition.

being greater and significant in a higher number of measures. Nonparametric correlation analyses confirmed these patterns by suggesting a clearer and stronger effect of repetition for targets, both in terms of the strength of the correlation and in terms of the number of measures for which a significant correlation was found (Table 6).

Nonparametric Mann-Whitney tests were also conducted between mean reading times for L1 and L2 readers (see Tables 4 and 5 for mean values), and results showed that, in all measures examined, L1 readers were significantly faster when reading both targets (first fixation duration: $Z = -7.79$, $p < .001$; gaze duration: $Z = -6.87$, $p < .001$; number of fixations: $Z = -4.24$, $p < .001$; total reading time: $Z = -6.80$, $p < .001$) and controls (first fixation duration: $Z = -8.53$, $p < .001$; gaze duration: $Z = -8.54$, $p < .001$; number of fixations: $Z = -7.15$, $p < .001$; total reading time: $Z = -12.12$, $p < .001$).

To further explore the effect of repetition on reading times, a gains analysis was conducted. The general improvement (i.e., overall decrease in reading times and number of fixations after the eight encounters) of targets and controls was compared. A gain score was calculated for each participant for targets and controls, and scores were compared. Results of Wilcoxon signed-rank tests in Table 7 showed that, in the case of L2 participants, the gain scores were significantly higher for targets than for controls in three of the four measures examined.

In the case of L1 participants, results of Wilcoxon signed-rank tests in Table 7 showed that, in line with L2 findings, the gain scores were significantly higher for targets than for controls in the four measures examined.

Nonparametric Mann-Whitney tests were also conducted between the gains experienced by L1 and L2 readers to explore potential differences between them. Results showed that there were no significant differences

Table 7. Gain values for targets and controls

Gains per reading measure	Targets ms (*SD*)	Controls ms (*SD*)	Sig.	*Z*
L2 participants				
First fixation duration	47.33 (76.65)	2.85 (74.34)	.053	–1.93[a]
Gaze duration	136.67 (133.25)	26.36 (74.19)	.003*	–2.98[a]
Number of fixations	1.31 (0.869)	0.28 (0.50)	.000*	–3.92[a]
Total reading time	403.02 (225.16)	73.97 (98.94)	.000*	–4.20[a]
L1 participants				
First fixation duration	38.19 (51.72)	–4.31 (53.50)	.007*	–2.71[b]
Gaze duration	171.23 (106.96)	9.48 (85.65)	.000*	–4.29[b]
Number of fixations	1.57 (0.81)	0.33 (0.43)	.000*	–3.94[b]
Total reading time	408.88 (188.19)	56.27 (89.17)	.000*	–4.37[b]

Note. * = significant at the $p < .0125$ level (Bonferroni correction applied). Sig. = significance.
[a] Based on negative ranks.
[b] Based on positive ranks.

between their gains, both for targets (gains first fixation duration: $Z = -.59$, $p = .556$; gains gaze duration: $Z = -.96$, $p = .337$; gains number of fixations: $Z = -1.08$, $p = .282$; gains total reading time: $Z = -3.30$, $p = .741$) and for controls (gains first fixation duration: $Z = -.73$, $p = .464$; gains gaze duration: $Z = -.33$, $p = .741$; gains number of fixations: $Z = -.26$, $p = .794$; gains total reading time: $Z = -.59$, $p = .556$). This suggests that, although L1 readers were generally faster in reading both types of items, both L1 and L2 readers experienced similar gains in terms of the decrease in reading times across encounters.

The next interesting question was to find out the point at which the difference between the reading behavior of targets and controls disappeared. Results have shown that targets showed overall significant longer reading times when compared to controls and that there is a general decrease in reading times. Wilcoxon signed-rank tests between each of the reading measures for targets and controls for each encounter showed that, in the case of L2 participants, although for first fixation duration and gaze duration the difference between targets and controls started to disappear at around the third and fourth encounter, it was not until the eighth encounter that this difference disappeared in all the measures (see Table 8). This suggests that, after eight encounters, words that readers had never encountered started to be read as words they knew very well.

Wilcoxon signed-rank tests with the L1 baseline data showed that, although some differences started to disappear after the second and

Table 8. Wilcoxon signed-rank tests between target and control reading measures for each encounter[a] (L2 participants)

	First fixation duration Target-Controls		Gaze duration Targets-Controls		Number of fixations Targets-Controls		Total reading time Targets-Controls	
Rep	Sig. (2-tailed)	*Z*	Sig. (2-tailed)	*Z*	Sig. (2-tailed)	*Z*	Sig. (2-tailed)	*Z*
Rep 1	.016	–2.40[b]	.000*	–3.62[b]	.000*	–4.20[b]	.000*	–4.20[b]
Rep 2	.000*	–3.89[b]	.000*	–3.74[b]	.000*	–4.11[b]	.000*	–4.20[b]
Rep 3	.031	–2.16[b]	.003*	–3.00[b]	.000*	–3.71[b]	.000*	–3.74[b]
Rep 4	.067	–1.83[b]	.091	–1.69[b]	.002*	–3.08[b]	.003*	–3.01[b]
Rep 5	.212	–1.25[b]	.035	–2.11[b]	.002*	–3.03[b]	.000*	–3.59[b]
Rep 6	.033	–2.13[b]	.009	–2.63[b]	.005*	–2.81[b]	.000*	–3.65[b]
Rep 7	.783	–0.28[c]	.064	–1.86[b]	.000*	–3.89[b]	.000*	–4.05[b]
Rep 8	.323	–0.99[b]	.784	–0.27[b]	.697	–0.39[b]	.338	–0.96[b]

Note. * = significant at the $p < .006$ level (Bonferroni correction applied). Sig. = significance; Rep = repetition.

[a] For means see Table 3.

[b] Based on negative ranks.

[c] Based on positive ranks.

third encounters, it was not until the eighth encounter that this difference consistently disappeared in all the measures (see Table 9), suggesting that, as in the case of L2 learners, after eight encounters, previously unknown words were read as known words.

In response to the second research question, analyses have shown that reading times for novel target items were initially significantly longer than for controls. Results also showed that there was an overall pattern of decrease in reading times throughout the eight exposures. The difference between the decrease experienced by control and target items (both in terms of the number of measures for which a significant effect was found and the gain analyses) suggests that this was not simply a practice effect and that, on top of the expected repetition effect, unknown items experienced an additional decrease in reading measures. Results have also shown that this effect started to be significant around the third to fourth encounter and that after eight encounters previously unknown words started to be read in a similar manner as known words. Results of these analyses have shown a very similar pattern for L1 and L2 readers. Interestingly, the only difference between L1 and L2 readers was not in the magnitude of the observed effect but in its rate. This significant effect of repetition seemed to start a bit earlier for the L1 readers (i.e., after the first encounter, but no differences were observed in terms of the overall gains in reading times).

Table 9. Wilcoxon signed-rank tests between target and control reading measures for each encounter[a] (L1 participants)

	First fixation duration Target-Controls		Gaze duration Targets-Controls		Number of Fixations Targets-Controls		Total reading time Targets-Controls	
Rep	Sig. (2-tailed)	*Z*	Sig. (2-tailed)	*Z*	Sig. (2-tailed)	*Z*	Sig. (2-tailed)	*Z*
Rep 1	.000*	–3.57[c]	.000*	–4.35$_c$	.000*	–4.31[c]	.000*	–4.37
Rep 2	.045	–2.01[c]	.000*	–3.97[c]	.000*	–3.87[c]	.000*	–4.10
Rep 3	.219	–1.23[c]	.002*	–3.11[c]	.000*	–4.07[c]	.000*	–4.10
Rep 4	.007	–2.71[c]	.904	–0.12[c]	.255	–1.14[c]	.115	–1.57
Rep 5	.367	–0.90[c]	.065	–1.84[c]	.000*	–4.19[c]	.000*	–4.21
Rep 6	.002*	–3.11[c]	.021	–2.30[c]	.343	–0.95[c]	.029	–2.18
Rep 7	.143	–1.47[c]	.008	–2.65[c]	.000*	–3.78[c]	.000*	–3.79
Rep 8	.375	–0.88[c]	.397	–0.85[c]	.378	–0.88[c]	.382	–0.88

Note. * = significant at the $p < .006$ level (Bonferroni correction applied). Sig. = significance; Rep = repetition.
[a] For means see Table 4.
[b] Based on negative ranks.
[c] Based on positive ranks.

Relationship between Offline and Online Measures

Participants' reading behavior on nonwords that had been learned versus those that were not learned (i.e., nonwords for which participants had provided the correct response versus those for which they had failed to provide an accurate response in each of the vocabulary tests) were compared. A mean score for the sum of total reading times on learned and nonlearned items was calculated per participant. Mean values for learned versus nonlearned words for all participants were compared by means of nonparametric Wilcoxon signed-rank tests. Results in Table 8 showed that words for which L2 participants were able to recall their meaning showed significantly longer total reading times than words for which participants did not show recall knowledge. There was no significant relationship between total reading times and the rest of the lexical aspects. In line with the results for L2 learners, results of Wilcoxon signed-rank tests (Table 10) showed a significant relationship between total reading times and L1 participants' ability to recall their meaning. The relationship between the other eye-tracking measures and vocabulary test scores was also examined, but analyses failed to show any significant relationship either for L2 or for L1 readers.

Table 10. Relationship between total reading times and measures of vocabulary learning

	Total reading times			
Vocabulary test	Learned words (ms)	Nonlearned words (ms)	Sig.	Z^a
L2 participants				
Form recognition	3959.76 (*SD* = 971.41)	3701.00 (*SD* = 1299.35)	.424	–0.80
Meaning recall	3844.79 (*SD* = 1040.52)	3352.98 (*SD* = 926.25)	.009*	–2.60
Meaning recognition	3998.13 (*SD* = 1102.02)	3625.53 (*SD* = 1028.53)	.307	–1.02
L1 participants				
Form recognition	2872.22 (*SD* = 995.28)	2699.50 (*SD* = 912.29)	.674	–0.42
Meaning recall	2968.10 (*SD* = 1122.16)	2702.47 (*SD* = 916.33)	.040*	–2.05
Meaning recognition	2875.27 (*SD* = 957.93)	2859.14 (*SD* = 1079.20)	.397	–0.85

Note. Sig. = significance.
[a] Based on positive ranks.
* $p < .05$.

In response to the third research question, results of this study indicate that there was a significant relationship between total time spent reading the target items and participants' ability to recall their meaning, suggesting that longer reading times on unknown items led to better learning of the form-meaning link at the recall level. This supports findings of previous studies that have also shown a relationship between reading times and vocabulary knowledge (e.g., Godfroid et al., 2013; Williams & Morris, 2004).

Overall, these results showed that reading was an effective means for acquiring new L2 vocabulary for all participants, not only in terms of the amount of words learned but also in terms of the speed and fluency of reading those new words. Results of the comparison of L2 readers' behavior with the L1 baseline data have also shown that L2 readers seem to benefit from the reading activity in a similar way to L1 readers.

GENERAL DISCUSSION AND CONCLUSION

It has been claimed that studies addressing different components of vocabulary knowledge have shown more diverse vocabulary learning than previously thought (e.g., Webb, 2005, 2007a, 2007b). Results of this study provide further evidence for the advantage of following a multicomponential and multitest approach to vocabulary research to examine the acquisition of different aspects of lexical knowledge. Results of this study have shown the effectiveness of reading for the incidental acquisition of several components of vocabulary knowledge (i.e., form recognition, meaning recognition, and meaning recall), in line with previous studies (e.g., Pellicer-Sánchez & Schmitt, 2010; Waring & Takaki, 2003; Webb, 2005). The L2 and L1 participants in this study learned a considerable number of the six unknown items after being exposed to them eight times, with meaning recall being the most difficult aspect to acquire, followed by the recognition of form and meaning, providing further evidence that productive aspects of lexical mastery are more difficult to acquire than recognition aspects (e.g., Brown et al., 2008; Pellicer-Sánchez & Schmitt, 2010; Rott, 1999; Webb, 2005, 2007a, 2007b). The present results have also shown that receptive aspects of vocabulary are not only easier to acquire but may also lead to higher degrees of certainty. The comparison between the L2 and L1 data also showed that there are no differences in the lexical gains experienced by L1 and L2 readers. The advanced L2 learners in this study appeared to benefit from the reading activity in the same way as L1 readers. Results of this study are indicative of incidental learning in the sense that these lexical gains occur as a by-product of the activity of reading and without being explicitly asked to learn that vocabulary. However, as shown by the online reading behavior, and as argued by Godfroid and colleagues (2013),

this does not mean that participants did not attend to those unknown items and attempt to infer their meaning.

This study has also examined the online reading of unknown items in context. L1 and L2 participants initially spent longer time reading the unknown items. These longer reading times in the initial encounters may reflect readers' attempts at inferring the meaning (e.g., Bengeleil & Paribakht, 2004; Fraser, 1999; Godfroid et al., 2013; Paribakht & Wesche, 1999). Present results have shown that, as the number of encounters increased, reading times and number of fixations for both L1 and L2 readers decreased. For L2 learners, this decrease started to be significant after three to four encounters and, for L1 readers, the decrease started to be significant after the first encounter. After eight repetitions, unknown items started to be read in a similar way as known words by both L1 and L2 participants. These patterns of reading behavior suggest that by three to four encounters L2 readers might have already integrated lexical and semantic information and attached a meaning to the unknown items. As Paribakht and Wesche (1999) claimed, the intake and subsequent integration of new vocabulary knowledge normally requires repeated input processing during multiple encounters. The results of the study reported in this article suggest that this integration of lexical and semantic information may happen earlier for L1 readers than for L2 learners. Wesche, Paribakht, and Haastrup (2010) claimed that during repeated exposures to new vocabulary, learners develop a more detailed mental representation of those words as well as increasingly fluent access to it. In the present study, this increased fluent access and a consolidation of that lexical knowledge might have been achieved by the eighth encounter and reflected in the similarity between the reading of targets and controls. This would suggest that the optimal figure for achieving a more fluent reading behavior is around eight encounters, in line with what previous studies examining other components of vocabulary knowledge have shown (e.g. Horst et al., 1998).

The patterns of reading found in the study reported in this article could also be interpreted by the five-step model of vocabulary learning proposed by Brown and Payne (1994; in Hatch & Brown, 1995). This model suggests that after encountering a new word (first stage), learners get a clear visual image of the word (second stage) and then connect that form with the meaning (third stage) and consolidate this form-meaning link (fourth stage) until they can use those learned words (fifth stage). L2 participants in this study may have reached the third stage by the third to fourth encounter, whereas L1 readers may have already reached it by the second encounter. The fourth stage of consolidation may have been reached by both L1 and L2 readers by the eighth encounter.

However, it is important to note that the preceding interpretation only accounts for the reading behavior of words for which participants created some sort of form-meaning link, and it is therefore only a prediction

of what the observed reading behavior may reflect. Alternative interpretations need to account for the decrease in reading times for words for which a form-meaning link was not successfully created. In line with Hyönä and Niemi's (1990) interpretation, the decreased reading times in this study can also be explained as an effect of increased familiarity of the visuographic features of nonwords. Participants' increased familiarity with the nonwords after the first encounter is reflected in a decrease in reading times, either with or without successful creation of a form-meaning link. This increased familiarity with the visuographic features of words could also explain the decrease experienced by control items.

The present study has also shown a relationship between longer reading times and participants' ability to recall the meaning of words, supporting results of previous studies (e.g., Godfroid et al., 2013; Godfroid & Schmidtke, 2013). The important role that attention plays in language learning has been stressed by many researchers (e.g., Schmidt, 1995). Alanen (1995) found that increased attention to target items seemed to be related to the acquisition of some aspects of those items regardless of the treatment received. Longer reading times may suggest increased attention, potentially explaining the connection between longer reading times and better recall scores. However, what the results of the study presented in this article strictly show is a link between reading times and vocabulary scores.

Regarding the comparison of L1 and L2 results, this study has shown that advanced L2 readers seem to benefit from the reading activity in a very similar way to L1 readers, both in terms of the lexical knowledge shown in the postreading tests and in terms of the online reading patterns. L1 and L2 readers seem to achieve the same ultimate improvement in reading times. The only difference seems to be in the rate of that improvement, with the effect of repetition happening earlier for L1 readers than for L2 learners. The similarity of these patterns is probably due to the high level of proficiency of the L2 learners in this study. Future studies should examine the behavior of learners of lower proficiency to investigate whether the patterns observed here also extend to other groups of L2 learners.

The results of this study have important pedagogical implications. Results have shown the important effect of frequency of exposure, reinforcing the need to provide learners with enough exposures to the target vocabulary. In addition, if longer reading times lead to better learning, teachers may need to use different techniques to increase the saliency of target vocabulary in reading texts and to drive learners' attention to the target vocabulary, increasing the time spent in reading target vocabulary. This is in line with studies suggesting that drawing learners' attention to words, for instance by highlighting words in the text (e.g., with color, bold, italics), could improve vocabulary gains

(e.g., Laufer & Hill, 2000). Winke (2013) investigated the effect of textual enhancement on grammar learning using eye-tracking and found that enhancement led to increased reading times but that it did not have any effect on learning. Further research should examine whether similar patterns are found for lexical learning.

One limitation of the present study is the use of invented items, which replaced high-frequency words as opposed to low-frequency real words. Although this ensured no previous knowledge of the target vocabulary, it could be argued that this may not represent a natural context of L2 reading and guessing from context. The reading materials of this study were not representative of the reading situations in which L2 learners acquire a new concept or a L2 word for a L1 concept for which they do not have a L2 word yet. However, they are representative of many other L2 reading situations in which learners learn a new label for a concept for which they already have another L2 word and create the form-meaning connection (e.g., when encountering archaic words, specialized vocabulary, or dialectal words in a text). In addition, all target items in this study were concrete nouns, which have been found to be easier to learn than other parts of speech (Macaro, 2003). Future studies that use eye-tracking to explore learning from reading should examine the acquisition of other parts of speech. It is also important to note that the situation in this study is an ideal reading situation: All words in the text are known except for six, relatively short words, which are repeated eight times each in a short time span and which have the same meaning in each of the encounters. Future research should examine other more complex reading situations.

Another important limitation of the present study is the diverse L1 background of the participants. The possible influences of different language backgrounds, L1 writing systems, and L1 reading skills on the processes investigated in this study should be examined in future studies.

Overall, this study has shown the advantages of using a variety of online and offline measures for researching vocabulary acquisition from reading. Results of this study have expanded our knowledge and understanding of the quantity and quality of vocabulary knowledge that can be learned from reading, providing a fuller account of incidental vocabulary learning not only *from* reading but also *while* reading.

Received 10 July 2014
Accepted 17 October 2014
Final Version Received 15 December 2014

NOTES

1. See Rayner (1997, 1998, 2009) for a comprehensive review of findings in eye movements and reading research.

2. See Clifton, Staub, and Rayner (2007) for a comprehensive review of lexical, semantic, and syntactical factors affecting eye movements in reading.

3. The limited reliability of the self-assessment measure needs to be acknowledged. Time constraints did not allow the use of a more direct measure of proficiency. In this study, this limitation is overcome by the other indicators of advanced level of proficiency (i.e., level of studies, the similar context of use of English, and pre-university requirements).

4. The assessment of comprehension was not the focus of the study and was only used to ensure participants' careful reading and to examine reading in the context of general comprehension. The true-false measure was thus considered a sufficient measure of general comprehension.

5. The number of outliers found on control words and on target words was very similar for both L2 and L1 participants (L2 participants: 45% of outliers on control words and 55% on target words; L1 participants: 48% of outliers on control words and 51% on target words).

REFERENCES

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning and teaching* (pp. 259–302). Honolulu, HI: University of Hawai'i Press.

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-fixations and reading times. *Memory & Cognition*, *24*, 477–492.

Barcroft, J. (2009). Effects of synonym generation on incidental and intentional L2 vocabulary learning during reading. *TESOL Quarterly*, *43*, 79–103.

Bengeleil, N. F., & Paribakht, S. (2004). L2 reading proficiency and lexical inferencing by university EFL learners. *Canadian Modern Language Review*, *61*, 225–249.

Brown, C., & Payne, M. E. (1994, March). *Five essential steps of processes in vocabulary learning*. Paper presented at the TESOL Convention, Baltimore, MD.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening. *Reading in a Foreign Language*, *20*, 136–163.

Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning new word meanings from context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 225–235.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. Van Gompel, M. Fisher, W. Murray, & R. L. Hill (Eds.), *Eye movement research: A window on mind and brain* (pp. 341–372). Oxford, UK: Elsevier Ltd.

Cobb, T. (n.d.). Compleat Lexical Tutor (Version 8) [Online resource]. Retrieved from http://www.lextutor.ca/

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

Eyelink, I [Apparatus and software]. Mississauga, ON: SR Research.

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

Godfroid, A., & Schmidtke, J. (2013).What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports and vocabulary learning scores. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 183–205). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye tracking. *Studies in Second Language Acquisition*, *35*, 483–517.

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, UK: Cambridge University Press.

Hatch, E., & Brown, C. (1995). *Vocabulary, semantics, and language education*. Cambridge, UK, and New York, NY: Cambridge University Press.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review*, *61*, 355–382.

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

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

Huckin, T., & Bloch, J. (1993). Strategies for inferring word-meanings in context: A cognitive model. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 153–178). Norwood, NJ: Ablex Publishing.

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

Juhasz, B. J., & Rayner, K. (2006). The role of age-of-acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, *13*, 846–863.

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*, 262–284.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal*, *73*, 440–464.

Laufer, B. (2003). Vocabulary acquisition in a second language. Do learners really acquire most vocabulary by reading? *Canadian Modern Language Review*, *59*, 565–585.

Laufer, B. (2005). Instructed second language vocabulary learning: The fault in the "default hypothesis." In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 286–303). Amsterdam, The Netherlands: Mouton de Gruyter.

Laufer, B., & Hill, M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning & Technology*, *3*, 58–76.

Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 381–390.

Lin, C., & Hirsh, D. (2012). Manipulating instructional method: The effect on productive vocabulary use. In D. Hirsh (Ed.), *Current perspectives in second language vocabulary research* (pp. 117–148). Bern, Switzerland: Peter Lang.

Macaro, E. (2003). *Teaching and learning a second language*. London, UK: Continuum.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 667–682.

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

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

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.

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

Nation, I. S. P., & Wang, K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language*, *12*, 355–380.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary.* Boston, MA: Heinle.

Paribakht, S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition. *Studies in Second Language Acquisition*, *21*, 195–224.

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

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, *18*, 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.

Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, *54*, 469–523.

Radach, R., Inhoff, A. W., & Heller, D. (2004). Orthographic regularity gradually modulates saccade amplitudes in reading. *European Journal of Cognitive Psychology*, *16*, 27–51.
Radach, R., & Kennedy, A. (2004). Theoretical perspectives on eye-movements in reading: Past controversies, current issues, and an agenda for future research. In R. Radach, A. Kennedy, & K. Rayner (Eds.), *Eye movements and information processing during reading* (pp. 3–26). New York, NY: Psychology Press.
Rayner, K. (1977). Visual attention in reading: Eye movements reflect cognitive processes. *Memory & Cognition*, *4*, 443–448.
Rayner, K. (1997). Understanding eye movements in reading. *Scientific Studies of Reading*, *1*, 317–339.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.
Rayner, K. (2009). Eye movements in reading: Models and data. *Journal of Eye Movement Research*, *2*, 1–10.
Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 779–790.
Rayner, K., & Juhasz, B. J. (2004). Eye movements in reading: Old questions and new directions. *European Journal of Cognitive Psychology*, *16*, 340–352.
Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research*, *16*, 829–837.
Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Hillsdale, NJ: Erlbaum.
Rayner, K., Raney, G. E., & Pollatsek, A. (1995). Eye movements and discourse processing. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 9–36). Hillsdale, NJ: Erlbaum.
Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1188–1200.
Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*, 595–626.
Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition*, *21*, 589–619.
Rott, S. (2007). The effect of frequency of input-enhancements on word learning and text comprehension. *Language Learning*, *57*, 165–199.
Saragi, T., Nation, P., & Meister, F. (1978). Vocabulary learning and reading. *System*, *6*, 72–78.
Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.
Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual.* London, UK: Palgrave Macmillan.
Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, *27*, 251–272.
Smith, B. (2012). Eye tracking as a measure of noticing: A study of explicit recasts in SCMC. *Language Learning & Technology*, *16*, 53–81.
Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*, 130–163.
Webb, S. (2005). Receptive and productive vocabulary learning: The effect of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*, 33–52.
Webb, S. (2007a). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*, 46–65.
Webb, S. (2007b). Learning word pairs and glossed sentences: The effects of a single context on vocabulary learning. *Language Teaching Research*, *11*, 63–81.
Wesche, M., & Paribakht, S. (2000). Reading-based exercises in second language vocabulary learning: An introspective study. *Modern Language Journal*, *84*, 196–213.
Wesche, M., Paribakht, S., & Haastrup, K. (2010). Research on the lexical inferencing process and its outcomes. In M. Wesche & S. Paribakht (Eds.), *Lexical inferencing in a*

first and second language: Cross-linguistic dimensions (pp. 3–31). Bristol, UK: Multilingual Matters.
West, R., Stanovich, K., & Mitchell, H. (1993). Reading in the real world and its correlates. *Reading Research Quarterly*, *28*, 35–48.
White, S. J., & Liversedge, S. P. (2004). Orthographic familiarity influences initial eye fixation positions in reading. *European Journal of Cognitive Psychology*, *16*, 52–78.
Williams, R. S., & Morris, R. K. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, *16*, 312–339.
Winke, P. A. (2013). The effects of input enhancement on grammar learning and comprehension. *Studies in Second Language Acquisition*, *35*, 323–352.
Wolverton, G. S., & Zola, D. A. (1983). The temporal characteristics of visual information extraction during reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 41–51). New York, NY: Academic Press.
Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effect of frequency and contextual richness. *Canadian Modern Language Review*, *57*, 541–572.

APPENDIX

EXAMPLES OF OFFLINE VOCABULARY TESTS

Test 1: Form Recognition

Choose the right spelling for the following six words that have appeared in the story (only one is correct) and indicate in the scale on the right how certain you are of your response. Example:

a) ackol	(b) acklon	c) hacklon	d) hackol	1 Very uncertain	(2) Uncertain	3 Certain	4 Very certain
1. a) hotler	b) holter	c) houter	d) houler	**1**	**2**	**3**	**4**
2. a) twoser	b) twonse	c) twiser	d) twines	**1**	**2**	**3**	**4**

AQ2

Test 3: Meaning Recognition

Select one of the five options. Only one is the correct definition. If you don't know the meaning of the word, please select option "e."

1) holter

1	2	3	4

a) basement

b) workhouse

c) prison

d) food hall

e) I don't know

2) twoser

1	2	3	4

a) story

b) punishment

c) noise

d) game

e) I don't know

Author Queries

QA	The distinction between surnames can be ambiguous, therefore to ensure accurate tagging for indexing purposes online (eg for PubMed entries), please check that the highlighted surnames have been correctly identified, that all names are in the correct order and spelt correctly.
AQ1	Does “glossed” modify “sentence” or “sentence contexts”? “short glossed-sentence contexts”?
AQ2	Should this be Test 2? There is no Test 2.