

Research Article

CONTEXTUAL WORD LEARNING DURING READING IN A SECOND LANGUAGE

AN EYE-MOVEMENT STUDY

Irina Elgort*

Victoria University of Wellington, New Zealand

Marc Brysbaert

Ghent University, Belgium

Michaël Stevens

Ghent University, Belgium

Eva Van Assche

Ghent University, Belgium

Abstract

Reading affords opportunities for L2 vocabulary acquisition. Empirical research into the pace and trajectory of this acquisition has both theoretical and applied value. Charting the development of different aspects of word knowledge can verify and inform theoretical frameworks of word learning and reading comprehension. It can also inform practical decisions about using L2 readings in academic study. Monitoring readers' eye movements provides real-time data on word learning, under the conditions that closely approximate adult L2 vocabulary acquisition from reading. In this study, Dutch-speaking university students read an English expository text, while their eye movements were recorded. Of interest were patterns of change in the eye movements on the target low-frequency words that occurred multiple times in the text, and whether differences in the processing of target and

This study was partially supported by a GOA grant from the Research Council of Ghent University (LEMMA Project). The Open Access publication was supported by Grant No. 215684 to Irina Elgort from Victoria University of Wellington.

*Correspondence concerning this article should be addressed to Dr. Irina Elgort, Senior Lecturer, Centre for Academic Development/School of Linguistics and Applied Language Studies, Victoria University of Wellington, P.O. Box 600, Wellington, 6140, New Zealand. Phone: +64 4 4635970; fax: +64 4 463 5284; E-mail: irina.elgort@vuw.ac.nz

Copyright © Cambridge University Press 2017. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

control (known) words decreased overtime. Target word reading outside of the familiar text was examined in a posttest using semantically neutral sentences. The findings show that orthographic processing develops relatively quickly and reliably. However, online retrieval of meaning remains insufficient for fluent word-to-text integration even after multiple contextual encounters.

INTRODUCTION

When reading in a foreign language, readers come across unfamiliar words that become increasingly more familiar with each contextual exposure, and eventually may be incorporated into the reader's lexicon. This process is referred to as incidental vocabulary learning from reading or contextual word learning (CWL).¹ Laboratory and in-situ experimental research into first (L1) and second/foreign (L2) language reading expanded our knowledge about various conditions that make CWL more or less effective.

Experimental studies into CWL tend to use short (usually single-sentence) contexts and manipulate a limited number of selected variables that affect CWL, while holding other variables constant. This approach allows researchers to better understand the effect of specific variables on CWL, but it does not convey the full picture of CWL in all its complexity. For example, although number of encounters with a word is a powerful predictor of learning, a word may not be learned even after many encounters if a reader's vocabulary is insufficient and if the text is not well understood (Elgort & Warren, 2014; Jenkins, Stein, & Wysocki, 1984; van Daalen-Kapteijns, Elshout-Mohr, & de Glopper, 2001; Zahar, Cobb, & Spada, 2001). Moreover, the use of single-sentence contexts (rather than a continuous text) may affect the choice of reading strategies and perception of word salience, both of which affect the learning of meaning in CWL (Huckin & Coady, 1999).

The present study takes a more naturalistic approach to researching L2 CWL. Dutch-English bilinguals read a continuous English expository text (about 12,000 words) from a nonfiction book, while their eye movements were recorded. Eye movements of the reader are closely coupled to ongoing language processing (Rayner, 2009), reflecting both lower-level lexical processing and word-to-text integration. The learning of *target* (lower-frequency) words that occurred in the text multiple times, was compared with *control* (high-frequency) words in the same text. To assess quality of word knowledge gained from CWL, participants' eye movements were also recorded while reading semantically neutral (i.e., weakly constraining) sentences with the target and control words, after reading the main text.

CONTEXTUAL WORD LEARNING

Two large crowdsourcing studies (Brysbaert, Stevens, Mandera, & Keuleers, 2016; Keuleers, Stevens, Mandera, & Brysbaert, 2015) showed that new word learning is a pervasive phenomenon, continuing beyond the age of 70. For both children and adults, learning a new word through reading involves establishing an orthographic (and corresponding phonological) representation and a semantic representation (Lindsay & Gaskell, 2013; Qiao & Forster, 2012; Share, 2004; Tamminen & Gaskell, 2013), and binding formal and semantic sources of information in such a way that word identity is reliably retrieved from an orthographic input (Perfetti, 2007).

According to the instance-based theory of CWL from reading (Bolger, Balass, Landen, & Perfetti, 2008; Reichle & Perfetti, 2003), each contextual encounter with a word results in a memory trace of this word plus its context. After multiple encounters with the same word in new diverse contexts, aspects of memory traces that overlap across encounters are reinforced, while memory traces that are not supported in the later encounters are weakened (see Nadel & Moscovitch, 1997, for an alternative, multiple trace theory of memory consolidation). Eventually a core meaning of a word is abstracted from multiple contextual encounters and is incorporated into a word's representation, in such a way that when the word's form is presented (with or without context) the reader can access its core meaning fluently, with minimal effort. Words that co-occur with the new word play an important part in the establishment and consolidation of its lexical-semantic representation, suggesting that existing vocabulary knowledge of the reader affects the speed and trajectory of CWL (Perfetti, Wloktó, & Hart, 2005).

This is one of the main reasons why the efficiency of reading as a means of L2 vocabulary acquisition is still being debated (Cobb, 2007; Laufer, 2005), especially when reading nonadapted L2 texts, in which many words may be only partially known (Nation, 2006). Although there is no question that it is possible to learn words through reading, only fairly modest gains in the knowledge of meaning are shown in L2 studies, even by less conservative estimates, such as multiple-choice tests. Zahar et al. (2001) reported word learning gains between 4% and 10%; Pitts, White, and Krashen (1989) found 6.4 % and 8.1% gains, while Horst, Cobb, and Meara (1998), who used a reading-plus-listening treatment, reported learning gains of about 20% (using multiple-choice tests to measure learning). Waring and Takaki (2003) reported an average 40% gain on a multiple-choice meaning test and 18% on a word translation test, at immediate posttest; and a 24% and 3.6% gain, respectively, after three months (but gains were higher on a form recognition test: 61% on the immediate and 34% on the delayed posttest). Importantly, learning gains vary depending on the choice of study design, reading texts and target words, participant populations, and measures used to evaluate learning (Pellicer-Sánchez, 2015).

One of the most powerful predictors of CWL is number of encounters with the word (Anderson, 1989; Ellis, 2002; Hulstijn, 2001; Waring & Takaki, 2003). A precise number of encounters needed for learning will inevitably vary between studies because of text, word, and individual reader variables affecting CWL (Huckin & Coady, 1999; Schmitt, 2008). Studies that use supportive single-sentence contexts, for example, report that as few as one to four encounters may be sufficient for learning word meanings (Elgort, Perfetti, Rickles, & Stafura, 2015; Mestres-Missé, Rodrigues-Fornells, & Münte, 2007; Webb, 2007). By contrast, at least six encounters are needed for a word to be learned from reading paragraphs or short texts (Rott, 1999; Vidal, 2011). In reading long continuous texts, between 8 and 12 encounters are needed for more robust CWL (Horst et al., 1998; Pellicer-Sánchez & Schmitt, 2010; Waring & Takaki, 2003).

Reading studies have also established that CWL is affected by context properties. Better learning of meaning is observed when a novel word is encountered in supportive contexts that make correct meaning inferences more likely, compared to neutral or misleading contexts (Batterink & Neville, 2011; Borovsky, Elman, & Kutas, 2012; Frishkoff, Perfetti, & Collins-Thompson, 2011; Mestres-Missé et al., 2007; Webb, 2008). In addition, varied (rather than repeated) contexts tend to result in better

knowledge of meaning because they establish richer semantic associations and help reject false meaning inferences (Bolger et al., 2008; Ferreira & Ellis, 2016; Rodríguez-Fornells, Cunillera, Mestres-Missé, & de Diego-Balaguer, 2009). However, most of these studies were conducted under highly controlled laboratory experimental conditions, mostly with L1 populations, and not enough is yet known about effects of contextual constraint and diversity on L2 CWL during reading long connected texts.

CWL MEASURES

Word knowledge is commonly measured by off-line tests of form and meaning recognition and retrieval. Evidence of learning provided by these tests varies from the most basic yes/no tests that require the learner to recognize “the form of a word and that it is a word rather than a meaningless jumble of symbols” (Daller, Milton, & Treffers-Daller, 2007, p. 4), to multiple-choice recognition tests that probe form-meaning mapping, to more demanding retrieval and recall tests. Among these off-line tests, the meaning generation task that measures ability to retrieve a meaning for a given written word form targets the kind of word knowledge needed in reading. However, such tests often concern off-line, explicit knowledge that can be expressed verbally, although other, more implicit tasks are available as well (e.g., the degree of priming by the newly learned stimuli, as reported by Elgort & Warren, 2014, among others). Therefore, behavioral online measures of word learning, such as eye movements are a welcome addition.

Psycholinguistic and neurolinguistic studies have an established tradition of combining online-off-line and explicit-implicit measures of word knowledge, especially when the goal is to verify theoretical or computational models of word recognition and processing (e.g., Godfroid & Spino, 2015; McLaughlin, Osterhout, & Kim, 2004; McRae, de Sa, & Seidenberg, 1997; Mestres-Missé et al., 2007). McLaughlin et al. (2004), for example, showed that ERPs (event-related brain potentials) were more sensitive indicators of L2 (French) word learning than off-line lexical (word/nonword) decisions. In L2 CWL, Elgort and Warren (2014) used an off-line meaning generation task to measure the development of off-line, explicit knowledge and an online semantic priming task to measure implicit knowledge of meaning. This approach revealed that these two knowledge types were affected by different reader variables: Meaning generation scores were predicted by L2 vocabulary size, while semantic priming by age of L2 acquisition.

EYE MOVEMENTS IN READING

Recording eye movements during reading is a relatively natural noninvasive way of studying reading processes in real time. Van Assche, Drieghe, Duyck, Welvaert, and Hartsuiker (2011, p. 93) argue that “[t]he use of the eyetracking method ... is probably the closest experimental operationalization of natural reading.” A large number of L1 reading studies have been undertaken in the last 20 years, and corpora of eye movements have been compiled (e.g., Cop, Drieghe, & Duyck, 2015; Kennedy & Pynte, 2005).

Eye movements in reading are not smooth; they consist of successions of fixations (when the reader extracts and encodes information from the text) and saccades (when the eyes are relocated to the next fixation point in the text). In normal fluent L1 reading, an

average fixation is 225 to 250 ms and an average saccade length is 7 to 9 letter spaces (Rayner, 2009). However, readers make more fixations and fixate for longer when they experience processing difficulty, for example when texts are more difficult or readers are less skilled (Rayner, 2009). Readers may preprocess text in the direction they are reading and make saccades' backward in the text (*regressions*). Higher regression rates reflect difficulties in word recognition or text comprehension (Rayner & Pollatsek, 2006).

Fixation durations reflect ease/difficulty of word processing (including accessing the meaning) and word-to-text integration. A key marker of processing difficulty in reading is a word's frequency of use in the language, with longer fixations observed on lower-than higher-frequency words (Juhász & Rayner, 2006; Kliegl, Grabner, Rolfs, & Engbert, 2004). Together with word length, frequency predicts nearly all of the variance that is in common between lexical decision times and word gaze durations in text reading (Kuperman, Drieghe, Keuleers, & Brysbaert, 2013). According to Rayner and Pollatsek (2006, p. 621), "the frequency effect typically ranges from about 20 to 40 ms in first-fixation duration and from 30 to 90 ms in gaze duration" for native speakers. Beyond frequency (and word length), variables that have been shown to influence fixation times, especially for low-frequency words, include word familiarity (Chaffin, Morris, & Seely, 2001; Williams & Morris, 2004); age of acquisition, that is, when a reader first encountered a word (Juhász & Rayner, 2006); and contextual diversity, that is, whether a word occurs widely across documents and text samples (contexts) or in relatively few contexts (Plummer, Perea, & Rayner, 2014). Fixation times and the likelihood of word skipping also vary depending on how easy it is to predict a word in context; in high-constraining contexts shorter fixations and higher word-skipping rates are observed (Reichle & Drieghe, 2013).

To establish a more detailed and accurate picture of the complex online language processing associated with reading, multiple eye-movement measures should be used (Rayner, 2009). When a word is the unit of analysis, most common eye-movement measures are *first-fixation duration* and *gaze duration*. First-fixation duration refers to the duration of the first fixation on a word (provided that the word was not skipped); gaze duration is the sum of all fixations on a word prior to moving to another word. Word-to-text integration measures include *go-past time*, which is the time from first fixating on the word until a fixation is made to the right of it (including regressions back in the text). The go-past time (also referred to as regression path duration) reflects "the time it takes upon reading the target word on first pass until it is successfully integrated with the on-going context" (Rayner & Pollatsek, 2006, p. 620). Another commonly reported word-to-text integration measure is the total time on a word, that is, *total reading time*—the sum of the durations of all fixations on a word including regressions. Two further eye-movement measures—the probability of making additional fixations on the word (following the initial fixation) and the probability of regressing back to the word—may reflect both lexical access and text comprehension processes.

CONTEXTUAL WORD LEARNING AND EYE-MOVEMENT RESEARCH

L1 reading studies (mostly with single-sentence contexts) show longer initial fixations and total reading times on novel words compared to familiar words, and more regressions back to novel words (Chaffin et al., 2001; Williams & Morris, 2004). Chaffin (1997)

observed longer first-fixation duration and gaze duration on novel words compared with familiar words, but not compared with less familiar words, suggesting that word familiarity affects lexical access. However, when followed by semantically supportive contexts, the rereading patterns on the novel words were characterized by more regressions back to these words (regressions-in) and longer processing times compared to both less and more familiar known words (Chaffin, 1997; Chaffin et al., 2001). Higher regression rates and longer reading times suggest that readers make an effort to derive novel word meanings and integrate them into context. Furthermore, larger word length effects for novel than familiar words (Lowell & Morris, 2014) imply that readers spend time on encoding unfamiliar letter strings as potential words. Therefore, it is reasonable to conjecture that unfamiliar orthographically legal letter-strings are processed as lexical items during reading (as if they were the lowest of low-frequency words).

There are not many L1 studies that directly investigate CWL with adult learners using eye-tracking. One such study by Joseph, Wonnacott, Forbes, and Nation (2014) found that repeated contextual exposure to novel words in meaning-defining sentence contexts reduced reading times on these words across all eye-movement measures (i.e., first-fixation, single-fixation, gaze duration, and total times). Joseph et al.'s results are in line with Rayner, Raney, and Pollatsek (1995) who found that the frequency effect is attenuated when words are repeated in short texts (after the third encounter with the word).

In the L2 context, there are only a few studies so far, in which eye movements were recorded during the reading of continuous texts (Balling, 2013; Cop et al., 2015; Godfroid, Boers, & Housen, 2013; Pellicer-Sánchez, 2015). Godfroid et al. (2013) recorded Dutch-English bilinguals' eye movements during reading short L2 (English) texts (newspaper and magazine clippings) adapted to the participants' L2 proficiency levels. The main goal of this study was to investigate whether attention paid to unfamiliar words during L2 reading resulted in better learning. In a counterbalanced manner, participants in this study were exposed to passages containing either control known words or nonwords that were matched with known words for word length in letters and number of syllables. In line with the results of L1 studies with novel words, Godfroid et al. (2013) found that L2 readers fixated on the nonwords for longer than on known words (as measured by first-fixation duration, gaze duration, second pass time, and total reading time). This finding confirms that novel words are processed as low-frequency words in the reading of connected L2 texts.

In another, yet unpublished study, Godfroid et al. (2016) tracked the eye movements of native and nonnative English speakers while they were reading the first five chapters (9,000 words) of an English novel that contained unknown Dari target words. The number of exposures to the target words ranged from 1 to 23. Reading times for the target words decreased in a curvilinear fashion with exposure, for L1 and L2 readers. Both number of exposures and total reading time contributed to vocabulary learning, as assessed in three vocabulary tests after text reading. Learning was 13% in an explicit meaning recall task, 30% in a meaning recognition task, and 33% in a form recognition task. There were no clear differences between the L1 and L2 readers in word learning.

Cop and colleagues shed new light on bilingual reading processes (Cop et al., 2015). In their study, unbalanced bilingual university students read a novel by Agatha Christie (56,000 words) in their L1 (Dutch) and L2 (English). Cop et al. (2015) reported that, for the same participants, L2 reading was characterized by longer sentence reading times

(20%), more fixations (21%), shorter saccades (12%), and less word skipping (4.6%), compared to L1 reading. The absence of differences in regression rates between L1 and L2 reading suggests that the readers had no major issues with word recognition or word-to-text integration when reading in their L2.

A study by Pellicer-Sánchez (2015) can be viewed a precursor to the present study and, therefore, will be discussed in detail. In this study, 23 postgraduate university students and postdoctoral researchers (from diverse L1 backgrounds) read a short story (2,300 words) written for the study in the L2 (English), while their eye movements were recorded. The story contained six nonwords (that replaced real high-frequency words) and six matching high-frequency control words that occurred eight times. The nonwords and controls were six-letter, two-syllables-long concrete nouns. The nonwords always occurred in high-constraining sentences that supported the guessing of their meanings from context. Eye-movement measures on the nonwords and controls (i.e., the first-fixation duration, gaze duration, number of fixations, and total reading time) were analyzed to establish how they changed during reading. After the reading, acquisition of the nonwords by the L2 readers was measured using three off-line vocabulary tests: multiple-choice tests of form and meaning recognition and a meaning retrieval test (a meaning generation task).

The results of the off-line tests of word knowledge were surprisingly high in this study, with the average of 85%, 78%, and 61% correct responses on each of the tests (respectively). These learning outcomes are higher compared with incidental word learning gains commonly reported in L2 reading studies. The initial reading times on the nonwords were longer than those on the control words on all eye-movement measures; these measures decreased significantly after eight contextual encounters, both on the nonwords (on the gaze duration, number of fixations, and total reading time, but not on the first-fixation duration) and the control words (on the number of fixations and total reading time). However, the effect of frequency of exposure was stronger for the nonwords than for the controls. For the nonwords, a significant reduction in reading times and number of fixations (compared to the first encounter) was observed on the third or fourth occurrence, while for the control words it occurred between the fifth and seventh occurrence in the text. The difference in the reading times and number of fixations between the nonwords and high-frequency controls disappeared by the eighth encounter. A similar pattern of results was observed for a group of 25 native speakers of English, who completed the same study procedure; but an earlier onset of reduction in reading times on the nonwords was observed for this group. Higher scores on the meaning generation task were observed for the nonwords that had been fixated on longer during the reading procedure. This result is aligned with the finding of L1 (Williams & Morris, 2004) and L2 (Godfroid et al., 2013) reading studies that also associate better CWL outcomes with longer total reading times, providing further evidence that some effort to derive word meanings during reading may be beneficial for the learning of meaning (Elgort & Warren, 2014; Godfroid et al., 2013).

Pellicer-Sánchez (2015) is the first eye-movement study that examined CWL in real time using a continuous L2 text. The findings of this study make an important initial contribution to our understanding of the trajectory of CWL, while setting an agenda for future research into CWL. As outlined in the discussion section of Pellicer-Sánchez (2015, p. 29), the study has a number of limitations. Firstly, the study design is an ideal

CWL scenario: (a) the nonwords were embedded in the text specifically created for the study in such a way that all other words were known to the readers; (b) the nonwords were presented in supportive contexts that made it possible to derive their correct meanings (as shown in the pilot); (c) each nonword was repeated eight times within a reasonably short text span; and (d) all words were concrete nouns signifying easy-to-understand referents. Secondly, the nonwords used in this study replaced high-frequency words, thus participants were learning new labels for previously known L2 words. Moreover, learning orthographically legal nonwords constructed with regular spelling may be easier than learning real low-frequency words that tend to have less regular spelling and orthography-to-phonology mapping. These study design choices may explain why learning gains reported in Pellicer-Sánchez (2015) were relatively high compared to other L2 studies of incidental word learning from reading. Finally, the study was conducted with participants from diverse L1 backgrounds (alphabetic, logographic, and syllabic), masking any effects that the L2-L1 distance may play in CWL.² The present study addresses many of these limitations and extends our understanding of CWL under more naturalistic conditions.

PRESENT STUDY

Our study settings approximate vocabulary learning conditions that are common in adult L2 vocabulary acquisition (outside of the language classroom), characterized by the need to acquire more sophisticated, topic-specific, low-frequency vocabulary to engage with authentic complex L2 input, such as textbooks and journal articles (Nagy & Townsend, 2012). In Belgium, students are expected to read course materials in English from the start of their university study. Because both the complexity and volume of course reading in English increase overtime, students' ability to build up their L2 vocabularies from reading is a critical success factor in their education. By drawing participants from this student population, while limiting their L1 to Dutch, the present study gains environmental validity while focusing on a homogenous L1 population, avoiding the confounds associated with using participants from diverse L1 backgrounds. We used a section from a nonfiction, general-academic book in English, used in some American universities as a supplementary course reading, that is, the type of text the participants would be expected to read in their university study.

RESEARCH QUESTIONS, LEARNING MEASURES, AND PREDICTIONS

The present study investigates the process of CWL in real time and estimates quality of CWL learning outcomes, posing the following questions: (a) How do eye movements on novel L2 words change compared to known L2 words as a result of reading a long continuous text? (b) What is the developmental pace and trajectory for the knowledge of form and meaning? (c) Do CWL gains progress over and above an increased familiarity effect reported in previous studies with repeated contents (e.g., Hyönä & Niemi, 1990)? (d) Are novel word meanings abstracted from contexts, in which they have been originally encountered, after reading a single long text?

To address the first two questions, eye movements of L2 readers reading a continuous expository text were recorded. Using eye-movement measures representing lexical

access (first-fixation durations, gaze duration) and word-to-text integration (go-past time, total reading time, fixations, and regressions rates) we evaluate the number of contextual encounters needed for the processing of an initially unfamiliar word to start approximating that of a known word (hereafter *control word*). Although Pellicer-Sánchez (2015) showed that eight encounters were sufficient for participants to read target nonwords in the same manner as known L2 words (on all eye-movement measures), her study used an “ideal” scenario. We predict that, although measures indicative of word-to-text integration on the target words will decrease in the course of reading (Joseph et al., 2014; Rayner et al., 1995), they are unlikely to become the same as those on high-frequency control words after only eight exposures (or by the end of reading one continuous text, for that matter). After all, lexical representations of high-frequency words have been fine-tuned through multiple contextual encounters, and their meaning representations are quickly and seamlessly activated regardless of the level of contextual support, making word-to-text integration a much easier affair. It is possible, however, that lexical decoding processes (measured by first-fixation duration and gaze duration) would become quite similar on the target and control words. This is because encountering a new word form eight times may be sufficient to establish a stable orthographic representation.

To address questions three and four, a sentence-reading posttest and a meaning generation task were administered after the reading of the main text. The reading posttest was based on Joseph et al. (2014), who presented contextually learned nonwords in semantically neutral sentences that did not provide any information pertaining to the meaning of these nonwords. In informative contexts, readers can use global or local contextual clues to boost emerging semantic representations of recently learned words, but access to meaning in neutral contexts is more difficult. The assumption is that, for words with well-specified, precise formal representations and context-independent meaning representations, there should be no difference in reading times between meaningful and neutral contexts. For words that have only weakly established representations, reading times in neutral contexts should be longer.

The sentence contexts from the online reading posttest were also used in the meaning generation task that assessed participants’ ability to recall word meanings explicitly. The mean accuracy of 20% was reported on a meaning generation task in Elgort and Warren’s (2014) study, in which nonwords were embedded in a longer extract from the same book. We predict that a similar (but slightly higher) knowledge of meaning scores will be observed in the present study (because some of the lower-frequency targets may be partially familiar to the readers), but that the scores will be lower than 61%, reported for the same task in Pellicer-Sánchez (2015).

METHODOLOGY

PROCEDURE

The study was conducted with each participant individually, over two consecutive days. On day one, the participants received information about the study and signed a consent form. Then they read the first part of the main text while their eye movements were being recorded. After this, the participants answered three open-ended comprehension questions about the passage they had read and completed vocabulary tests that measured

their L2 vocabulary size. On day two, the participants read part two of the text while their eye movements were being recorded and answered three open-ended comprehension questions. After this they answered questions about their experience of reading the text, their general reading practices, and their language background. Then the participants completed the final sentence reading test, while their eye movements were being recorded. Finally, they completed a written meaning generation task.

PARTICIPANTS

Forty Dutch-speaking university students (28 females) enrolled in a bachelor or master program at Ghent University participated in the present study, either for course credit or for a monetary reward (€10 per hour). Participants accepted into the study (Table 1) had an estimated higher-intermediate to advanced L2 (English) proficiency, based on their English LexTALE vocabulary scores (Lemhöfer & Broersma, 2012).

Participants' average L2 vocabulary size (in English) was estimated at 8,855 thousand word families,³ using Nation's (2006) vocabulary size test. L2 readers whose vocabulary size is around 8,000 word families should be able to read authentic texts in English with good understanding (Nation, 2006).

Six participants had to be excluded from the analysis of the eye-movement data during reading the main text, due to cases of drift away from the text line.

MATERIALS

TARGET WORDS

Fourteen lower-frequency words (Appendix A) were used as target words in this study. Because lower-frequency words, by definition, are unlikely to occur multiple times in the same text, 12 (out of 14) target words were created by replacing mid-frequency words with lower-frequency synonyms, for example, a mid-frequency word *fine* (meaning “monetary penalty”) was replaced with *mulct* (meaning “fine” or “compulsory payment”). Twelve target words were nouns and two were gerunds, that is, nominalized verbs (i.e., *rigging*, as in “match rigging,” and *diddling*, as in “mechanics of diddling”). Nine high-frequency words that occurred in the text multiple times were used as control

TABLE 1. Participants' characteristics

Characteristic	Mean	St. Dev	Range
Age	19.18	2.45	17–29
Age of L2 (English) acquisition	11.43	3.03	3–14
LexTALE score (English) [score out of 100]	80.62	8.80	66.25–97.50
LexTALE (Dutch) [score out of 100]	90.16	5.77	75–100
Vocabulary size test (English) [score out of 14,000]	8,855	1,329	6,500–11,300
M-C vocabulary test (Dutch) [score out of 75]	50.48	9.20	29–71
Books read a year (in English)	2.25	2.20	0–8
Books read a year (in Dutch)	3.60	4.74	0–25
Hours a week of course-related reading (in English)	5.46	4.61	0–20
Hours a week of course-related reading (in Dutch)	9.25	8.44	0–40

words in the study (Appendix A). They were matched with target words for number of occurrences in the text and number of letters (Table 2).

THE TEXT

The main text used in the study comprised an introduction and chapter one from the book *Freakonomics* by Levitt & Dubner (2006). The lexical coverage of the text was acceptable for the proficiency level of the target population (95.75% of the text was covered by the first 6,000 word families, based on the BNC and COCA corpora, while the average participants' vocabulary size in English was estimated at 8,855 word families; see Table 1). Thus participants were reading an authentic general-academic text at an appropriate difficulty level.

The text contained 12,152 words. The whole text was divided into two parts; the first part (5,673 words) was read on the first day, while the second part (6,479 words) was read on the second day of the study. Four target words occurred only on day one, five occurred only on day two, and five occurred on both days.

APPARATUS

Throughout the two reading sessions of the main text and the final sentence reading posttest, participants' eye movements were recorded using a desktop-mounted EyeLink 1000-Plus system (SR-Research, Mississauga, Canada) with a sampling rate of 1,000 Hz. A chinrest was used to reduce head movements. The camera-to-eye distance was 50 cm. The monitor display resolution was 1280 × 1024 pixels. Although participants read binocularly, only the movements of the right eye were monitored. During the reading procedure, the room was dimly illuminated.

The main text was presented in black, Courier New font size 14, on a light gray background. The lines were triple spaced. The whole text was presented over 100 screens

TABLE 2. Characteristics of the two word types: Target and control words

Characteristic	Target words (n=14)			Control words (n=9)		
	Mean	St.Dev.	Range	Mean	St.Dev.	Range
Occurrences in the text	22.29	17.43	8.00–64.00	27.89	13.03	9.00–51.00
Frequency list ^a	11.45	5.32	5.00–23.00	1.78	0.83	1.00–3.00
Zipf_Subtlex_US ^b	2.66	0.74	1.59–4.28	4.58	0.61	3.86–5.81
Prevalence ^c	1.11	1.01	-1.60–2.16	2.78	0.23	2.5–3.07
Number of letters (word forms)	6.71	1.68	5.00–10.00	6.67	1.73	4.00–9.00

^aBased on Nation's BNC/COCA word-family lists (available from www.victoria.ac.nz/lals/about/staff/paul-nation).

^bZipf values 1–3 represent low-frequency words, 4–7 are high-frequency words (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). Frequency values were unavailable for three target words (*mulct*, *rikishi*, and *honbasha*).

^cWord prevalence, defined as word knowledge in the population (Keuleers et al., 2015), is a strong predictor of word-processing times, especially for low-frequency words. Prevalence values were unavailable for two target words (*rikishi* and *honbasha*).

of a similar word length (155 words maximum). Participants used a button box connected to the computer to move from one screen to the next. They could not go back to reread previous screens. A drift check was performed prior to the presentation of each screen. Before reading the main text (on both days), participants read a practice paragraph; no target words occurred in the practice paragraphs. On each day, the participants took two compulsory breaks during reading (approximated 10 minutes apart) to rest their eyes; on day one, the breaks were after screens number 16 and 32, and on day two, after screens number 67 and 84. These breaks were placed at the logical points in the text, where a new section of the chapter began. A nine-point calibration was executed before the beginning of the reading procedure and at the end of each break. Additional calibrations were performed for each participant, as required, to manage the eye-tracker signal drifting over time (e.g., Pellicer-Sánchez, 2015, excluded data from 14 L2 participants—38%, due to drift).

The target and control words were never presented in initial or final position in a line. This is because initial and final words are less likely to be fixated on, and *return sweeps* (i.e., saccades moving from the end of one line to the next) “typically last longer than the movements that progress along a line, and they also tend to undershoot the intended target” (Rayner & Pollatsek, 2006, p. 614). The target words were never located closer than two words apart in a sentence.

In the reading posttest of word knowledge, neutral sentences (Appendix B) were presented in the middle of the screen, in the order randomized for each participant. The sentences were never longer than one line, and the target and control words never appeared in initial or final position. Participants were instructed to read sentences for meaning and press a button to proceed to the next sentence. A nine-point calibration was performed before the beginning of the reading test, and a drift check was performed prior to the presentation of each sentence. Four practice sentences were used at the beginning of the test.

The meaning generation task was delivered online using Qualtrics (www.qualtrics.com). Each word was presented on a separate screen, and the participants could not go back to change previously completed responses. The words (with their corresponding neutral sentences) were presented in the order randomized for each participant. The participants were instructed to provide an explanation for each word in Dutch or English, or translate the word into Dutch. For each word, the participants were also instructed to indicate whether they had seen or heard the word prior to encountering it in the main text of the present study.⁴ They were not asked whether they knew the meaning of the word, therefore we will refer to this data as *prior familiarity* rather than prior knowledge. Responses were assigned a rating of 1 (correct or mostly correct) or 0 (mostly incorrect, not enough information, or no response) by two independent raters.

ANALYSIS AND RESULTS

EYE MOVEMENTS IN READING THE MAIN TEXT

Visual Inspection of the Eye-Movement Data

First, we visually inspected the eye-movement data for the first 40 observations on the target and control words in the main text (Figure 1), using (a) plots of the means and standard deviations and (b) plots that illustrate key patterns in the data by applying smoothers (i.e., approximating mathematic functions) that capture the impact of the predictive variables

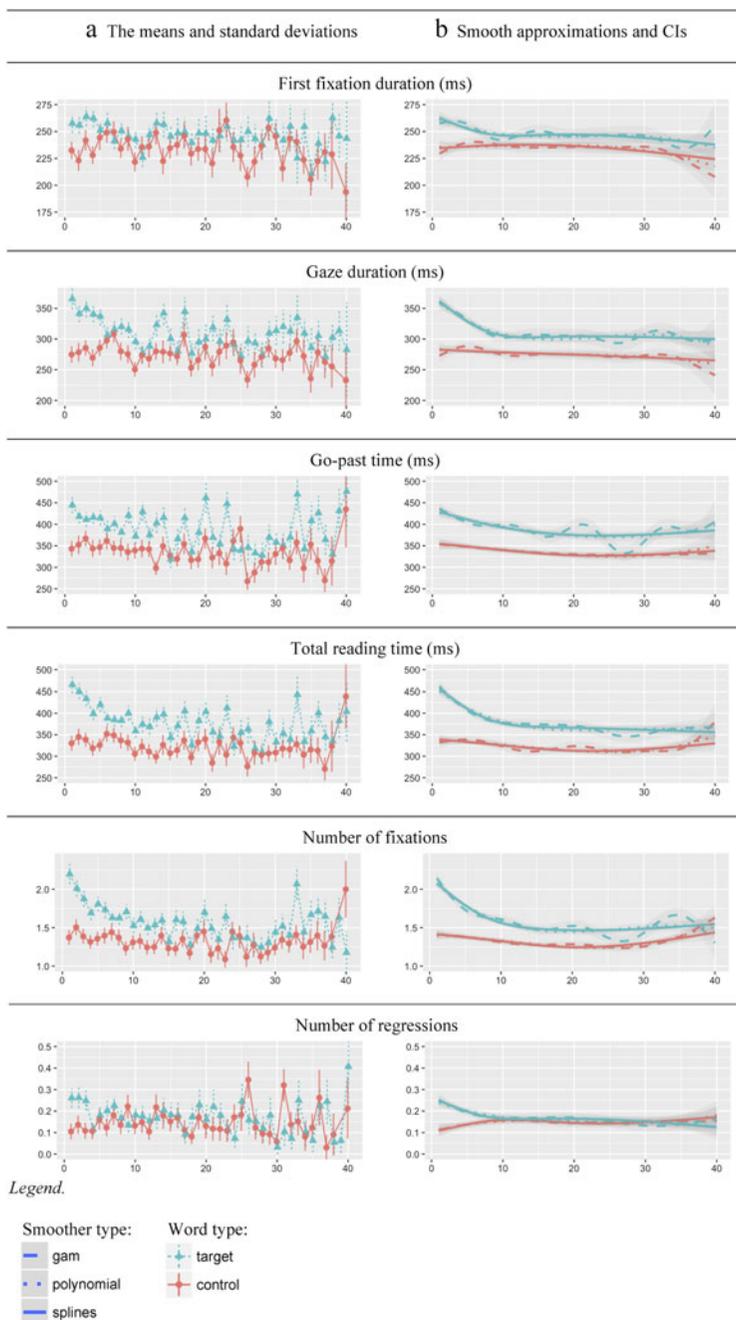


FIGURE 1. Eye-movement measures on the targets and controls in the main text.

Notes: Smoothing methods (functions): *gam*—generalized additive model (linear relationships between variables are not assumed); linear smoothers: *polynomial*—third order polynomial; *splines*—regression splines with three degrees of freedom.

(i.e., order of occurrence and wordtype). The cutoff point (40 occurrences) was chosen because only two targets and one control occurred more than 40 times in the text. The plots show an increase in variability (noise) after the initial 30 occurrences, because only three targets and five controls occurred more than 30 times in the text. Wordskips⁵ and single observations shorter than 100 ms and longer than 1000 ms were discarded (Appendix C).

On the measure most indicative of lexical access, first-fixation duration, there was an overall decrease in the first 5 to 7 observations on the targets, after which data patterns for the targets and controls converged to a difference of some 10 ms (Figure 1b). A sharp decrease in gaze durations continued until just prior to the 10th observation, after which gaze duration remained stable for both word types. A difference in gaze duration of about 25 ms between the targets and controls persisted until the end of reading.

The picture that emerged for the go-past time was somewhat different. On the go-past time, the slope of decrease was less steep than for first fixation and gaze duration, and it continued to decline beyond 20 observations. There was also a decrease in go-past time on the control words, but it was less dramatic than that on the targets. Importantly, a relatively large difference between the two word types (about 75 ms) remained until the end of reading.

A steep decrease in total reading time on the targets occurred within the first 10 observations, and only a minor decrease occurred after this point. Total reading time on the targets remained greater than that on the controls, for which a small gradual decrease was also present. The patterns observed for total reading time were mimicked by those for number of fixations; this is not surprising as total reading time comprises the sum of durations of all fixations on a word.

The number of regressions-in (i.e., regressions back to the target word from later parts of the sentence) showed a fast and dramatic change within the first five occurrences, becoming indistinguishable from regressions-in on the control words.

ANALYSIS OF THE EYE-MOVEMENT DATA FOR THE FIRST EIGHT OCCURRENCES IN THE TEXT

For further analysis of the eye-movement data (Table 3), a cutoff was made at eight occurrences for two reasons: (a) the visual inspection of the data showed a change in the reading patterns on the target words just prior to 10 occurrences in the text; and (b) the cutoff affords a comparison with the results reported in Pellicer-Sánchez (2015) who used eight contextual exposures to the nonwords. Linear mixed-effects models were fitted to the eye-movement data using the R package lme4 (Bates et al., 2016). Participants and words were included in the models as crossed random effects. To avoid inflating Type 1 error rates, all models included the maximal random effects structure justified by the data (Barr, Levy, Scheepers, & Tily, 2013). Order of occurrence in text (*occurrence*) and *word type* (target vs. control) were used as primary interest predictors.⁶ The word-type variable was centered using Helmert contrasts. To investigate whether the differences in the eye-movements on the target and control words reduced overtime, an interaction between occurrence and word type was included in the models. Type III analysis of variance was used as test of effects. For continuous dependent variables (i.e., first-fixation duration, gaze duration, go-past time, and total reading time), we used F tests with the Satterthwaite approximation for degrees of freedom. For the number of fixations and regressions, we used a Poisson model (that uses z and χ^2 distributions) (Table 3).

TABLE 3. Comparison of the eye-movement measures on the target words and higher-frequency controls on the first eight occurrences in the main text

Measure	Occurrence		Word_type		Interaction Occurrence: Word_type	
	Statistic	<i>p</i>	Statistic	<i>p</i>	Statistic	<i>p</i>
First-fixation duration	$F_{(1,33.6)}=0.01$.9400	$F_{(1,38)}=25.94$	9.9e-06	$F_{(1,33.7)}=7.75$.0087
Gaze duration	$F_{(1,21.2)}=2.39$.1369	$F_{(1,21.2)}=16.13$.0006	$F_{(1,21.2)}=9.13$.0064
Go-past time	$F_{(1,23)}=7.12$.0137	$F_{(1,23.6)}=20.88$.0001	$F_{(1,23.9)}=5.10$.0333
Total reading time	$F_{(1,20.9)}=10.20$.0043	$F_{(1,21.4)}=27.6$	3.1e-05	$F_{(1,21.2)}=13.10$.0016
No. of fixations	$\chi^2_{(1)}=11.28$.0008	$\chi^2_{(1)}=23.10$	1.5e-06	$\chi^2_{(1)}=7.79$.0053
No. of regressions	$\chi^2_{(1)}=0.21$.6431	$\chi^2_{(1)}=10.85$.0010	$\chi^2_{(1)}=4.26$.0391

Learning curves are curvilinear (Figure 1), following an exponential function (Heathcote, Brown, & Mewhort, 2000) or a power function (Logan, 2002). The curvilinearity can be captured by polynomials of the second degree or restricted cubic splines with three knots. However, as can be seen in Figure 1, over the first eight trials we are not losing much information by using a straightforward linear regression. As the hypothesis concerns an interaction, this makes the analysis simpler and more robust.

To give an example of the analyses run, we provide the codes used for the gaze durations and the number of fixations:

```
library(lme4)
```

```
gaze <- lmer(GAZE_DURATION ~ OCCURTEXT * WORD_TYPE + (1|BASEWORD)
+ (0+OCCURTEXT|BASEWORD) + (1|PARTICIPANT), data=text8)
```

```
anova(gaze)
```

```
nfix <- glmer(NUMBER_OF_FIXATIONS ~ OCCURTEXT * WORD_TYPE + (1|BASE-
WORD) + (0+OCCURTEXT|BASEWORD) + (1|PARTICIPANT), data=text8,
family=poisson)
```

```
library(car)
```

```
Anova(nfix, type=3)
```

In both analyses, random intercepts of participants and stimuli (BASEWORD) were included, as well as random slopes for occurrence over stimuli.

On all eye-movement measures, there was a significant interaction between occurrence and word type, confirming statistically that the differences in the reading of the target and control words reduced significantly by the eighth occurrence in the text. These results show that both lexical access and word-to-text integration of the target words became more like those of the control words by the eighth occurrence.

POSTTEST: WORD READING IN SEMANTICALLY NEUTRAL SENTENCES

In the reading posttest, the words were presented in semantically neutral sentences. First, we examined whether the processing of the target words was comparable to that of the control words, without the familiar context of the long text, in which they were originally

encountered (Tables 4 and 5).⁷ The analysis showed that the target words were processed reliably more slowly than the controls on most measures, but not on first-fixation duration.

To investigate the quality of knowledge established for the target words, the eye movements in the sentence-reading posttest were compared with those on the last occurrence of each word in the main text (Figure 2). No difference between the eye movements in the text and posttest would indicate that meaning representations were abstracted from the specific text context and could be accessed without contextual support (i.e., in semantically neutral contexts). In this analysis, the reading task (text vs. test) and wordtype (target vs. control) were used as primary interest predictors. To investigate whether the change in reading patterns between the two tasks were different for the targets and controls, an interaction between task and word type was included in the models.

As can be seen in Table 6, the interaction between task and word type was reliable on the total reading time, number of fixations, and number of regressions. The target words (but not controls) had longer reading times (by 49 ms) and more fixations in the posttest than in the main text (Figure 2). The number of regressions back to the target word increased in the posttest compared to the last occurrence in the text; while the trend was in the opposite direction for the controls. There was also a trend toward an interaction between task and word type in the gaze duration analysis ($p = .15$), with gaze duration being 26 ms longer at posttest. No significant difference between participants' reading of the targets in the text and test was observed on the remaining eye-movement measures (i.e., first-fixation duration and go-past time).

Finally, we checked if day of encounter (i.e., day one, day two, or both days) affected the eye-movement measures on the last encounter with the target word in the main text and on the reading posttest. There was an interaction between day of encounter and the difference in the eye movements in the main text and on posttest, for two measures: regressions-in ($\chi^2_{(2)} = 14.90, p < .001$) and go-past time ($\chi^2_{(2)} = 9.67, p < .05$). On the last occurrence of the target word in the main text, there were fewer regressions-in when it was encountered on the same day (either day one or day two) than when it was distributed across two days. However, in neutral contexts on posttest there was no difference between the three distributions (Figure 3a). The go-past time on the last occurrence in the main text was shorter when the target word was encountered on both days than on one day only. However, there was no difference between the distributions on the posttest (Figure 3b).

TABLE 4. Comparison of the eye-movement measures on the target words and higher-frequency controls, in the sentence-reading posttest

Measure	Statistic	<i>p</i>
First-fixation duration	$F_{(1,33.6)} = 2.39$.1300
Gaze duration	$F_{(1,22.3)} = 4.70$.0410
Go-past time	$F_{(1,21.7)} = 9.85$.0048
Total reading time	$F_{(1,23)} = 16.30$.0005
No. of fixations	$\chi^2_{(1)} = 21.12$	4.3e-06
No. of regressions	$\chi^2_{(1)} = 15.20$	9.7e-05

TABLE 5. Estimates and standard errors on the target and control words, in the sentence-reading posttest

Measure	Target		Control	
	Estimate	SE	Estimate	SE
First-fixation duration	254.52	7.13	238.67	9.76
Gaze duration	352.63	21.57	284.84	26.06
Go-past time	392.84	21.30	303.32	24.71
Total reading time	430.17	23.25	305.84	25.09
No. of fixations	1.90	0.06	1.25	0.08
No. of regressions	.21	0.17	0.07	0.28

OFF-LINE MEASURES

PRIOR FAMILIARITY

The mean prior familiarity with the target words was 19%, and it was 99% for the high-frequency controls. For the targets, prior familiarity ranged from 2.5% to 37.5% (but it was 52.5% for the word, *realtor*; Appendix A).

MEANING GENERATION TASK

The mean accuracy of responses to the target words in the meaning generation task was 34% (Appendix A), and it was 99% for the control words (with 98% interrater agreement between two independent ratings). For the target words, the mean accuracy ranged between 0% and 67.5% (but it was 92.5% for the word *diddling*). Target words that were not familiar to the participants prior to reading the text were significantly more likely to get incorrect responses on the meaning generation task, $\chi^2_{(1)} = 13.04, p < .001$ (based on the Pearson Chi-Square Test). Words that were self-rated as unfamiliar had the estimated mean accuracy of 17%, while those self-rated as familiar had the estimated mean accuracy of 48%.

Accuracy of responses to the target words on the meaning generation task was also affected by the distribution of encounters with the targets in the main text across the two days ($\chi^2_{(2)} = 6.30, p < .05$), after accounting for the variation due to the number of target

TABLE 6. Comparison of the eye-movement measures on the target and control words between the last reading in the text and the posttest

Measure	Text vs.test		Word_type		Interaction Text vs. Test: Word_type	
	Statistic	p	statistic	p	statistic	p
First-fixation duration	$F_{(1,27.5)}=0.49$.49	$F_{(1,21.8)}=1.58$.22	$F_{(1,20.1)}=.54$.47
Gaze duration	$F_{(1,26.4)}=0.55$.46	$F_{(1,22.2)}=2.91$.10	$F_{(1,25.5)}=2.20$.15
Go-past time	$F_{(1,28.3)}=0.07$.79	$F_{(1,22.3)}=6.31$.02	$F_{(1,21.7)}=1.37$.25
Total reading time	$F_{(1,28.2)}=1.50$.23	$F_{(1,21.5)}=7.80$.01	$F_{(1,21.6)}=7.33$.01
No. of fixations	$\chi^2_{(1)}=6.09$.01	$\chi^2_{(1)}=9.25$.002	$\chi^2_{(1)}=5.42$.02
No. of regressions	$\chi^2_{(1)}=.14$.71	$\chi^2_{(1)}=8.94$.003	$\chi^2_{(1)}=11.90$.001

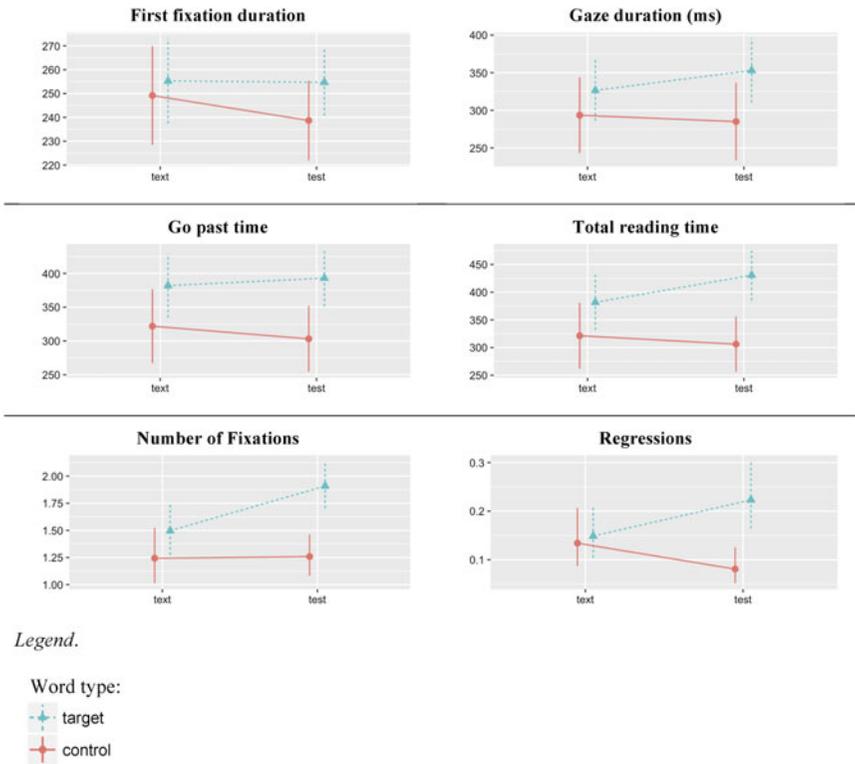


FIGURE 2. Comparison of the eye-movement measures on the target and control words in the main text and the posttest.

occurrences in the text. The targets encountered by the participants on both days resulted in higher accuracy (53%), compared to those that were encountered either on day one (3%) or day two (24%) (Figure 3c). The effect of day remained significant when prior familiarity with the target word was included as a predictor.

DISCUSSION

KEY FINDINGS AND THEIR INTERPRETATION

A rich and complex picture of CWL emerged in the present study. The eye-movement data show a clear change in the manner the target words were read at early and late occurrences in the text, on both lexical access and word-to-text integration measures. The results of the eye-movement analyses on the first eight encounters showed a reliable decrease in the differences between the processing of the target and control words during reading. However, the trajectory of change was different for the two types of measures: after a steep decrease, first-fixation durations, gaze durations, and regressions-in became similar on the newly learned targets and high-frequency controls. Integrating word meanings into context, however, remained more difficult for the target than control words, until the end of reading (Figure 1). This is particularly apparent in the sluggish

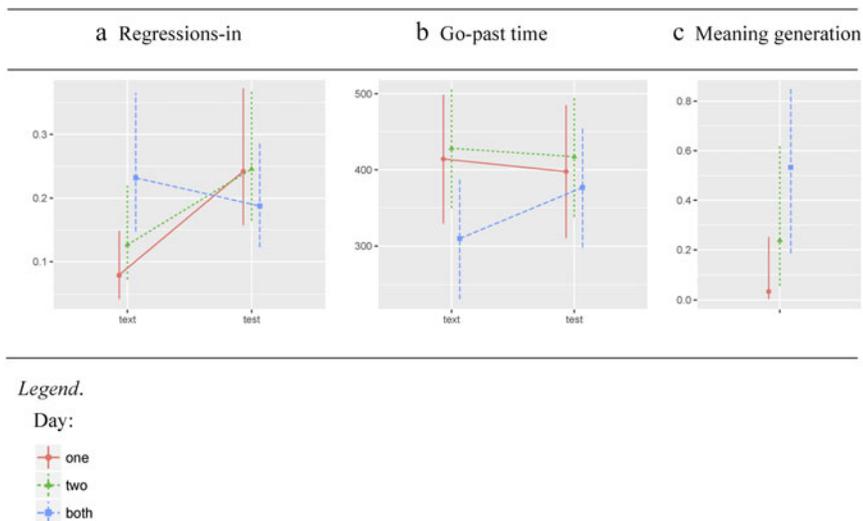


FIGURE 3. Effect of day of encounter with the target word (*day*) on the difference in the eye movements on the last encounter with the target word in the main text and on posttest (a, b); and meaning generation accuracy (c).

patterns of change for the targets observed on the go-past time measure that represents ability to fluently access word meanings and integrate them into the preceding context.

An interesting interpretation of first fixations is provided by the E-Z Reader model of eye-movement control in reading (Reichle, Pollatsek, Fisher, & Rayner, 1998). This model makes a distinction between an orthographic familiarity check and full word identification. The next eye movement in reading is programmed on the basis of an orthographic familiarity check before the word is fully identified and its meaning accessed. Reichle et al. (1998, p. 133, italics added by the present authors) describe the familiarity check: “The process by which a word’s familiarity, *f*, is computed (before lexical access) is likely to be the product of many factors, including frequency of occurrence in printed text, length, age of acquisition, *frequency of usage*, *recency of usage*, and the number and frequency of word neighbors.”

The first-fixation duration results show a relatively fast optimization of orthographic information on which the familiarity check is based, suggesting that these representations are established within the first 5 to 10 encounters with a novel word in reading. This finding is supported by the fast reduction in regressions back to the word (regressions-in). Regressions-in indicate attempts by the reader to derive meaning from new words after receiving additional information from subsequent context (Reichle, Warren, & McConnell, 2009) or to correct wrong inferences that were made. An incorrect inference may occur as a result of misidentification of an unfamiliar word during reading, before its precise orthographic representation has been established. For example, the target word, *goad*, may be misread as a high-frequency orthographic neighbor, *goal*; the target word, *succor*, may be misread as a familiar word, *soccer* (Perfetti, 2007). This scenario is more plausible in L2 reading (Laufer, 2003), because lexical representations of L2 words (even those considered known) are less precise than in L1. An attempt to correct such a

misreading would result in a regression back to the target word. The fact that the patterns of change in regressions-in rates on the targets were similar to those on first-fixation duration supports the misidentification conjecture; once a reasonably stable orthographic representation of a new word is established, no more misreadings occur and regressions-in rates on the targets become the same as those on the controls.

Importantly, the eye-movement measures capturing word meaning and integration in the ongoing text indicate clear differences between the target and the control words up to the very end of text, sometimes after 40 readings of the words (Figure 1). As Table 6 shows, we found significant differences between the target words and the control words on go-past times, total reading times, number of fixations, and number of regressions. In addition, there was a difference in gaze durations, even though it did not reach statistical significance ($p = .10$). To some extent, the higher values for the target words are not surprising, as in eye-movement data frequency effects are still observed for well-known words, which the readers have encountered hundreds of times in their life (Kliegl et al., 2004). What our results show is that such frequency-related differences are also present when a word has been encountered many times *recently*. It takes many readings over prolonged periods of time and probably across different contexts before words reach their full lexically quality (cf. the advantage of the two-day presentation we found).

The analysis of eye movements in reading semantically neutral sentences in the posttest confirmed that the orthographic representations of the target words were established during the reading of the main text, but they were less robust than those of the control words. Lexical access measured by gaze duration was reliably slower (on average by about 68 ms) for the newly learned words compared to the controls, indicating that their lexical-semantic representations were still not as fine-tuned as those of the high-frequency words. On the measures involving word-to-text integration (go-past time and total reading time), the target words were also processed slower than the controls; in addition, they received more fixations and regressions-in, than the controls.

Participants also made more fixations and spent longer reading the newly learned targets in the neutral sentences than on final observations in the main text. This suggests that the familiar context of the main text made it easier for the participants to access meanings of the target words and integrate them into context. This finding indicates that lexical-semantic representations of the target words were still weak and partially contextually bound, even after the participants had encountered them multiple times in one continuous text.

Results of the meaning generation task showed only moderate ability to explicitly recall a word's meaning when its form was presented. As predicted, the overall accuracy (34%) was higher than the 20% observed for nonwords in Elgort and Warren (2014) and the 13% observed for Dari words in Godfroid et al. (2016), but the mean accuracy on the unfamiliar targets (17%) was close to that reported by Elgort and Warren and by Godfroid. The overall accuracy of 34% on the meaning generation task (and even the 48% response accuracy to the targets self-rated as previously familiar) were lower than the 61% observed by Pellicer-Sánchez (2015) for nonwords. This suggests that the CWL rates reported in Pellicer-Sánchez (2015) may be overoptimistic, and may not generalize to more naturalistic CWL conditions. Notably, no reliable relationship between the eye-movement measures on the reading posttest and the meaning generation task were observed in the present study (tested in an analysis not reported in the present study). This suggests a possible dissociation between the development of knowledge supporting word

understanding during reading and knowledge allowing participants to explicitly formulate the meanings of words learned in CWL.

Finally, distribution of the target words in CWL across the two days affected explicit and (to a lesser extent) implicit knowledge (Figure 3). Although the memory trace left from day one was not strong enough for the participants to retrieve the meaning of the target word, it was strongly boosted by the reoccurrence of the word on day two. This effect was beyond the simple recency effect observed for the words encounter on day two only. Furthermore, the distribution of encounters across the two days had an effect on regressions-in and go-past time, on the last encounter with the target in the main text. Fewer regressions back to the target word were observed when it was encountered on the same day, compared to two different days. This suggests that massed distribution may be better for the learning of word form. Conversely, shorter go-past times were observed for the targets that were encountered on both days, than on one of the days. This suggests that encounters across multiple days are more beneficial for the learning of meaning: contextually learned words distributed across both days were integrated into the preceding context in a more fluent manner. This finding is in the same direction as the effect of day of encounter on the participants' ability to recall word meanings explicitly. Still, the effect of day on the two eye-movement measures did not hold on the sentence-reading posttest. This corroborates our previous conclusion that online processing of contextually learned words in neutral contexts was less fine-tuned than in the original text.

A READING COMPREHENSION THEORY PERSPECTIVE

The fine-grained information about contextual L2 word learning obtained in the present study may be further understood in reference to the *Reading Systems Framework* (Perfetti & Stafura, 2014) that situates word-level processes in broader reading comprehension. Within this framework, visual input (supported by existing linguistic and writing system knowledge) triggers word identification (i.e., fluent and accurate recognition of the written form and rapid and reliable retrieval of meaning); and the output of the word identification system contributes to the building of units of meaning (propositions) within the comprehension system. Thus, word-to-text integration processes “reflect a close coupling of word identification with representations of the meaning of the text, mediated by the retrieval and selection of word meanings” (Perfetti & Stafura, 2014, p. 30).

Because, in the present study, orthographic knowledge of the target words was achieved within the first five to seven occurrences in the text, the participants could decode the target words efficiently when they encountered them in new contexts, both in the main text and in the posttest (see results on first-fixation durations). While this decoding is a necessary condition of lexical access, online retrieval of the relevant meaning component of word knowledge is critical for efficient word identification. Longer gaze durations on the newly learned targets than on the controls in the main text and at posttest, and longer gaze durations in semantically neutral contexts than in the main text on the targets suggest that target word identification was not fully optimized, most likely due to weakly established context-dependent meaning representations.

Incomplete or inaccurate input from the online word identification system to the comprehension system may lead to potential coherence breakdowns, necessitating more

laborious active construction processes, instead of memory-driven low-cost word-to-text integration characteristic of fluent reading comprehension. The sluggish change observed for the targets on the go-past time measure in the main text suggests that participants were experiencing difficulties with fluent word-to-text integration even by the 30th occurrence in the text. Higher regression-in rates and longer reading times on the target words in neutral posttest contexts, compared to those in the main text, also point to the need for more active, effortful word-to-text integration processes when the context does not provide sentence- or text-level support for meaning retrieval. Taken together, our results show that L2 reading triggered CWL but, even after multiple encounters, word-to-text integration remained suboptimal for fluent reading comprehension.

CONCLUSIONS AND LIMITATIONS

The present study is a close approximation of the L2 CWL situation in Belgian universities. In this study, Dutch-speaking university students read an introductory expository text in English containing a number of low-frequency words for general understanding. By monitoring eye movements during reading in real time, we charted learning trajectories of different aspects of word knowledge and comprehension. Online processing of these contextually learned words was further examined in isolated neutral sentences, affording a test of context-independent meaning retrieval—a pressure point of reading comprehension.

The overall findings are fairly encouraging; they suggest that unfamiliar words do not tend to be ignored in L2 reading for meaning, at least in an academic setting. Our results also reinforce the point made in incidental word learning research that CWL is a slow and incremental process. The quality of lexical-semantic representations (aka form-meaning mapping) of contextually learned words remained insufficient for fluent word-to-text integration even after many contextual encounters. This means that reading one or two book chapters or journal articles, where a novel word occurs multiple times, does not establish the level of word knowledge needed for low-cost online word-to-text integration. When processing resources are drawn to the word-level comprehension, reading cognitively demanding texts with understanding is more difficult. Therefore, the use of L2 course readings should be carefully planned. For instance, it may be preferable for course texts that introduce new complex concepts and ideas to be read in the students' L1, while L2 texts may be more appropriate as follow-up readings exemplifying or reiterating these ideas.

The present study examined CWL in a specific population—university students—whose approach to reading is likely to be shaped by the context of university study. It may be that, under more relaxed circumstances, readers would be less inclined to learn novel words. Another limitation of this study is related to the density of occurrence of low-frequency words; to investigate CWL over multiple encounters some more frequent words in the original text were replaced with low-frequency synonyms, artificially increasing the density of occurrence of low-frequency words in the text. However, two redeeming factors need to be kept in mind: (a) university courses often introduce new terms that are used more frequently in discipline-specific texts than in general texts; and (b) for less proficient L2 readers, mid-frequency L2 words (that occur more frequently in the language than low-frequency words) are, in fact, main learning targets in CWL. Future studies should explore the effects of density/spacing of encounters on CWL from reading.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <https://doi.org/10.1017/S0272263117000109>.

NOTES

¹*Incidental* means occurring by chance, as a result of another activity, and is often taken to mean the opposite of deliberate (with intention). This distinction is useful when contrasting learning words from reading with and without additional instructions and tasks. However, even when no specific vocabulary learning instructions are given, coming across unfamiliar words during reading may trigger different kinds of processes, from basic visual intake and semantic integration to deliberate attempts to encode form and derive meaning (Paribakht & Wesche, 1999). For this reason, we use an intentionality-agnostic term, CWL.

²Although some of these limitations are addressed in the Godfroid et al. (2016) study, two limitations related to the characteristics of the participants and target items remain. Different (nonspecified) L1s of the study participants add noise to the findings about word learning trajectories, as they are likely to be affected by the native language of the reader. The use of Dari (L3) words in an English (L2) story sets it apart from other studies investigating how reading contributes to L2 word learning.

³Word family is a unit of counting used in word frequency lists organized in levels (or bands) of 1,000 headwords. List one contains the most-frequent 1,000 word families of English, based on the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). The lists are used to measure the vocabulary load of reading texts (Nation, 2006) and estimate receptive vocabulary knowledge of language users (Nation & Beglar, 2007). A word family consists of a headword (e.g., *cancel*) and its inflected forms (*cancelled*, *cancelling*), as well as derivative forms that share a common meaning with the headword (*cancellation*, *cancellations*).

⁴No pretest of word knowledge was conducted prior to the commencement of the reading procedure, to avoid altering participants' perceptions of the words' salience in natural reading.

⁵In excluding word skips (zero fixations) we followed an established practice in the eye-movement research studying reading (e.g., Chaffin et al., 2001). This ensures that the first-fixation durations and gaze durations reported in this article are comparable to those in previously published studies.

⁶The addition of prior familiarity with the word to the models did not change the pattern of significance for the primary-interest predictors on any of the eye-movement measures. Prior familiarity with the word did have a main effect on total reading time (only), at first encounter.

⁷We found no effect of prior familiarity with the word on any of the eye-movement measures in the models comparing the reading of the target and control words at posttest. The R code used for the tests in Table 4 was `gaze <- lmer(GAZE_DURATION ~ WORD_TYPE + (1|BASEWORD) + (WORD_TYPE|PARTICIPANT), data = test)`, showing that the model included random intercepts for basewords and participants, and random slopes for word type over participants.

REFERENCES

- Anderson, J. R. (1989). Human memory: An adaptive perspective. *Psychological Review*, 96, 703–719.
- Balling, L. W. (2013). Reading authentic texts: What counts as cognate? *Bilingualism: Language and Cognition*, 16, 637–653.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., et al. (2016). Package "lme4." Retrieved from <http://kambing.ui.ac.id/cran/web/packages/lme4/lme4.pdf>.
- Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: An event-related potential study. *Journal of Cognitive Neuroscience*, 23, 3181–3196.
- Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Contextual variation and definitions in learning the meaning of words. *Discourse Processes*, 45, 122–159.
- Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development*, 8, 278–302.

- Brysbaert, M., Stevens, M., Mander, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, *7*, 1116.
- Chaffin, R. (1997). Associations to unfamiliar words: Learning the meaning of new words. *Memory and Cognition*, *25*, 203–226.
- 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.
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning and Technology*, *11*, 38–63.
- Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS ONE*, *10*, e0134008. doi:10.1371/journal.pone.0134008.
- Daller, H., Milton, J., & Treffers-Daller, J. (Eds.). (2007). *Modelling and assessing vocabulary knowledge*. Cambridge, UK: Cambridge University Press.
- Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*, 365–414.
- Elgort, I., Perfetti, C. A., Rickles, B., & Stafura, J. Z. (2015). Contextual learning of L2 word meanings. *Language, Cognition and Neuroscience*, *30*, 506–528.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143–188.
- Ferreira, R. A., & Ellis, A. W. (2016). Effects of contextual diversity on semantic decision and reading aloud: Evidence from a word learning study in English as a second language. *Studies in Psychology*, *37*, 162–182.
- Frishkoff, G. A., Perfetti, C. A., & Collins-Thompson, K. (2011). Predicting robust vocabulary growth from measures of incremental learning. *Scientific Studies of Reading*, *15*, 71–91.
- Godfroid, A., & Spino, L. A. (2015). Reconceptualizing reactivity of think-alouds and eye tracking: Absence of evidence is not evidence of absence. *Language Learning*, *65*, 896–928.
- Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye tracking. *Studies in Second Language Acquisition*, *35*, 483–517.
- Godfroid, A., Ahn, J., Cho, I., Ballard, L., Cui, Y., Johnston, S., et al. (2016). *Incidental Vocabulary Learning during Extensive Reading: An Eye-Tracking Study*. Manuscript submitted for publication.
- Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin and Review*, *7*, 185–207.
- 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.
- Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, *21*, 181–193.
- Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge, UK: Cambridge University Press.
- Hyönä, J., & Niemi, P. (1990). Eye movements in repeated movements of a text. *Acta Psychologica*, *73*, 259–280.
- Jenkins, J. R., Stein, M., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, *21*, 767–787.
- Joseph, H. S. L., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition*, *133*, 238–248.
- 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.
- Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, *45*, 153–168.
- Keuleers, M., Stevens, M., Mander, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology*, *68*, 1665–1692.
- 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.

- Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *The Quarterly Journal of Experimental Psychology*, *66*, 563–580.
- 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.
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavioral Research Methods*, *44*, 325–43.
- Levitt, S. D., & Dubner, S. J. (2006). *Freakonomics*. Camberwell, Australia: Penguin.
- Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *39*, 608–622.
- Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*, 376–400.
- Lowell, R., & Morris, R. K. (2014). Word length effects on novel words: Evidence from eye movements. *Attention, Perception, and Psychophysics*, *76*, 179–189.
- McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*, 703–704.
- McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*, 99–130.
- Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex*, *17*, 1858–1866.
- Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*, 217–227.
- Nagy, W. E., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, *47*, 91–108.
- Nation, P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63*, 59–82.
- Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*, 9–13.
- Paribakht, S., & Wesche, M. (1999). Reading and “incidental” L2 vocabulary acquisition: An introspective study of lexical inferring. *Studies in Second Language Acquisition*, *21*, 195–224.
- Pellicer-Sánchez, A. (2015). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*, 97–130.
- Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from reading an authentic novel: Do things fall apart? *Reading in a Foreign Language*, *22*, 31–55.
- Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–383.
- Perfetti, C. A., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*, 22–37.
- Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1281–1292.
- Pitts, M., White, H., & Krashen, S. D. (1989). Acquiring second language vocabulary through reading: A replication of the *Clockwork Orange* study using second language acquirers. *Reading in a Foreign Language*, *5*, 271–276.
- Plummer, P., Perea, M., & Rayner, K. (2014). The influence of contextual diversity on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 275–283.
- Qiao, X., & Forster, K. I. (2012). Novel word lexicalization and the prime lexicality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1064–1074.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*, 1457–1506.
- Rayner, K., & Pollatsek, A. (2006). Eye-movement control in reading. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed.) (pp. 613–658). Amsterdam, The Netherlands: Academic Press.
- 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.

- Reichle, E. D., & Drieghe, D. (2013). Using E-Z Reader to examine word skipping during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1311–1320.
- Reichle, E. D., & Perfetti, C. A. (2003). Morphology in word identification: A word experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading*, *7*, 219–237.
- 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 and Review*, *16*, 1–21.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125–157.
- Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 3711–3735.
- 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*, 589–619.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329–363.
- Share, D. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology*, *87*, 267–298.
- Tamminen, J., & Gaskell, M. G. (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *The Quarterly Journal of Experimental Psychology*, *66*, 1001–1025.
- Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, *64*, 88–107.
- van Daalen-Kapteijns, M., Elshout-Mohr, M., & de Glopper, K. (2001). Deriving the meaning of unknown words from multiple contexts. *Language Learning*, *51*, 145–181.
- Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*, 1176–1190.
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*, 219–258.
- 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. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*, 46–65.
- Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*, 232–245.
- Williams, R. S., & Morris, R. K. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, *16*, 312–339.
- Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review*, *57*, 541–572.