Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Incidental Learning and Long-Term Retention of New Word Meanings From Stories: The Effect of Number of Exposures

Rachael C. Hulme, Daria Barsky, and Jennifer M. Rodd

University College London

This study used a Web-based naturalistic story-reading paradigm to investigate the impact of number of exposures on incidental acquisition and long-term retention of new meanings for known words by native English-speaking adults. Participants read one of four custom written stories in which they encountered novel meanings (e.g., a safe concealed within a piece of furniture) for familiar words (*foam*). These meanings appeared two, four, six, or eight times in the narrative. Results showed reasonably good memory of the new meanings, assessed by cued recall of novel meanings and word forms, after only two exposures, emphasizing the importance of initial encounters. Accuracy in cued recall of novel meanings showed a linear, incremental increase with more exposures. There was no significant forgetting after 1 week, regardless of the

---

This work was supported by a grant from the Economic and Social Research Council (1473923) awarded to the third author. The first author was supported by a doctoral studentship from the Economic and Social Research Council (1473923). We thank Helen Moss and Johan Heemskerk for authoring the stories used as stimulus materials and Rachel Jose for developing the comprehension questions used with the stories. We also thank Eva Poort and Becky Gilbert for advice on the statistical analyses, and we thank Eva Poort for providing helpful comments on an early draft of this article.

This article has been awarded an Open Data badge. All data and analysis scripts are publicly accessible via the Open Science Framework at https://osf.io/ybu6r. Learn more about the Open Practices badges from the Center for Open Science: https://osf.io/tvyxz/wiki.

Correspondence concerning this article should be addressed to Rachael Hulme, Department of Experimental Psychology, University College London, 26 Bedford Way, London, WC1H 0AP, United Kingdom. E-mail: rachael.hulme.14@ucl.ac.uk

DOI: 10.1111/lang.12313

number of exposures during training, demonstrating the efficiency with which adults acquire new word meanings incidentally through reading and retain them over time.

Keywords number of exposures; incidental learning; word meaning; acquisition; homonyms; first language; vocabulary learning; story reading

Introduction

Word learning in the native language (L1) continues throughout the adult lifespan. In addition to frequently learning entirely new words and their meanings, adults must often learn new meanings for words already present in their mental lexicon. As many as 80% of English words are ambiguous in that they have more than one definition (Rodd, Gaskell, & Marslen-Wilson, 2002), and previously unambiguous words often acquire new meanings. This occurs, for example, as language evolves, especially due to changes in technology (e.g., the newer Internet-related meaning of *troll* as a person who posts deliberately antagonizing comments online), or when one learns about a new subject or activity (e.g., the sailing term *boom* for a part of a yacht; Rodd et al., 2012).

Learning new L1 word meanings in everyday life generally takes place incidentally by inferring the new meaning from the surrounding context (Batterink & Neville, 2011) rather than through intentional memorization. Incidental vocabulary learning can be defined as the learning of words and their meanings unintentionally while engaged in another activity, such as reading for comprehension (Hulstijn, 2003), in contrast to intentional learning, which is the deliberate attempt to memorize words and their meanings. For incidental learning from reading, certain factors concerning how new words and their meanings are presented in the text can impact on subsequent learning and retention. One key factor is the number of exposures to new vocabulary items (P. Nation, 2015). The impact of the number of exposures on adults' incidental vocabulary learning from reading has mainly been investigated in the domain of second language (L2) learning (e.g., M. Horst, Cobb, & Meara, 1998; Pellicer-Sánchez & Schmitt, 2010; Rott, 1999; Waring & Takaki, 2003; Webb, 2007). There are relatively fewer studies looking at adults' incidental acquisition of new L1words and their meanings.

Incidental L1 Vocabulary Acquisition From Reading

All the studies on adults' incidental L1 vocabulary learning from reading to date have been concerned with the learning of new word forms. This has either entailed participants learning foreign or nonword labels for already known concepts (e.g., Batterink & Neville, 2011; Mestres-Missé, Càmara, Rodriguez-Fornells, Rotte, & Münte, 2008; Mestres-Missé, Rodriguez-Fornells, & Münte, 2007; Pellicer-Sánchez, 2016; Saragi, Nation, & Meister, 1978; Williams & Morris, 2010) or in a few cases learning new words along with their novel, foreign, or artificial meanings (e.g., Godfroid et al., 2017; Henderson, Devine, Weighall, & Gaskell, 2015).

An early study on L1 vocabulary acquisition from reading was a highly naturalistic study that used an authentic text as the stimulus material (Saragi et al., 1978). In the study, native English-speaking participants read the novel *A Clockwork Orange* by Anthony Burgess, which contains 241 words in the fictional slang register Nadsat that are repeated on average 15 times (range: 1–209). Participants were not aware that they would be tested on their memory of the novel words and were instead told that they would be given a comprehension and literary criticism test. When their memory of 90 novel words was tested several days later in a meaning-to-word matching test, participants showed significant acquisition of the words (76% correct) just from reading the narrative. The researchers also found a significant positive correlation between the number of times a word occurred in the novel and the number of participants who correctly recalled the meaning. Saragi et al. suggested that the minimum number of repetitions required for L1 speakers to learn words incidentally while reading is "somewhere around ten" (p. 76). However, since this early study, research has revealed different factors that contribute to incidental vocabulary learning depending on different properties of the words. Therefore, focusing on a specific threshold to ensure learning is less useful than characterizing the impact of number of exposures under typical incidental learning conditions.

Studies with ecological validity remain highly valued in the study of incidental vocabulary acquisition (Spivey & Cardon, 2015). A new eye-tracking study by Godfroid et al. (2017) investigated participants' incidental learning of 29 Dari (an Afghani dialect of Farsi) words and their meanings while reading part of the novel *A Thousand Splendid Suns* by Khaled Hosseini in English, which was either their L1 or L2. The number of exposures to the Dari words in the text ranged from 1 to 23. In addition to monitoring eye movements during reading, the researchers assessed subsequent vocabulary acquisition through surprise tests of word form recognition, meaning recall, and meaning recognition. There was modest vocabulary learning: Participants reading in their L1 scored 31.4% correct for word form recognition, 32.7% for meaning recognition, and 12.2% for meaning recall. More importantly, number of exposures was the strongest predictor of successful acquisition, more so than the total reading time summed across exposures. The eye movement data revealed a nonlinear decrease in reading times across exposures, with significant cubic and quadratic effects.

Godfroid et al.'s (2017) and Saragi et al.'s (1978) studies demonstrated clear incidental learning in the highly naturalistic context of reading real novels. However, they lacked experimental control over the number of exposures to the target words, which varied greatly in these authentic novels. Crucially, in such highly naturalistic materials the number of exposures may well be correlated or confounded with other properties of the new word meanings, such as how central they are to the story's plot, and some items may be intrinsically easier or harder to learn than others. This therefore emphasizes the need for experimental control of the number of exposures in a within-item design.

In contrast to the previously discussed research, several studies have examined the processing and acquisition of novel L1 words with only a few exposures but in less naturalistic contexts, such as short sentences (Mestres-Missé et al., 2008, 2007; Williams & Morris, 2010). In their eye-tracking study, Williams and Morris measured acquisition of 12 nonwords using a two-choice synonym recognition test after participants had read a single meaningful sentence for each item. Average performance on this simple task was only 62%. Using different online processing measures, Mestres-Missé and colleagues carried out an event-related potential (ERP) study (Mestres-Missé et al., 2007) and a functional magnetic resonance imaging (fMRI) study (Mestres-Missé et al., 2008) to investigate meaning acquisition from context across three exposures with Spanish participants reading in their L1. In the ERP study, they found that after three exposures to 65 items in contiguous sentences, brain potentials to novel words were already indistinguishable from real words. Participants also showed moderate learning on a word pair task: They correctly recognized 69% of new word meanings and correctly rejected 67% of incorrect meanings. The fMRI study revealed similar acquisition rates from three exposures to 50 items (69% correctly identified meanings; 44% correctly rejected meanings). These studies using online measures of reading therefore provided some evidence for inferring and acquiring meanings of novel words from just one or three exposures in sentence contexts. However, the strength of these learning effects and the extent to which they translate into acquisition success remains unclear because these studies used only very simple postreading vocabulary measures, if any at all.

A few studies have combined elements of the more ecologically valid studies with experimental control of the number of exposures to items by using customized stories written or modified specifically for this purpose (e.g., Batterink & Neville, 2011; Henderson et al., 2015; Pellicer-Sánchez, 2016). Batterink and Neville investigated native English speakers' semantic integration of new meanings for 26 nonwords that were derived from context during story reading across 10 exposures. They modified stories to give exactly 10 exposures to the target words and examined semantic integration using the N400 ERP component, a negative component occurring around 400 milliseconds after stimulus onset, whose amplitude varies in inverse relation to readers' expectation of the upcoming word in a sentence (Kutas & Federmeier, 2011). Batterink and Neville found a greater reduction in N400 amplitude, indicating more semantic integration by participants for nonwords embedded in consistently meaningful contexts than for nonwords occurring in inconsistent, meaningless contexts. This reduction was already visible from the participants' second exposure to the words. Acquisition was assessed explicitly through recall and recognition tasks. Participants' accuracy in recognizing the meanings of the novel words was 72.4%, and their accuracy on cued recall of meanings was 63.8%.

Another recent study by Pellicer-Sánchez (2016) used a story that had been written specifically to present stimuli to participants reading in L2 and to a L1 control group. Participants' eye movements were monitored as they encountered the meanings of six nonwords, each appearing eight times throughout the narrative. Pellicer-Sánchez found that accuracy in recognizing the correct spelling for the new words was 91.3% for participants reading in their L1 when they were tested immediately after reading. Their accuracy in recognizing the meanings for those words in a multiple-choice word-to-meaning matching test was 86.6%, and their accuracy in cued recall of the meanings was 65.3%. The eye-tracking data showed that participants reading in their L1 read the novel words significantly faster after only the first encounter, and after eight exposures they read them at a speed similar to real, known words. Longer overall reading times were also associated with higher performance on the vocabulary measures.

These studies have demonstrated incidental learning of new words and their meanings through reading a single text in L1 although with somewhat mixed success. However, vocabulary gains from reading a single text are likely different from incidental learning through more extensive reading. Several studies with L2 learners (M. Horst, 2005; Webb & Chang, 2015) have found larger vocabulary gains from reading multiple different texts than typically found through reading a single text. There are various reasons why the amount of vocabulary learning may be greater from reading multiple texts. For example, within a single text there are smaller intervals between individual exposures whereas multiple texts give more spaced encounters that may be more beneficial for learning (Webb & Chang, 2015). Additionally, words read in multiple texts are likely encountered in more diverse contexts (K. Nation, 2017), which may enable readers to build more stable representations of the meanings of words. Conversely, children learn vocabulary better from being repeatedly read the same storybook than from the same number of exposures across different storybook contexts (J. Horst, Parsons, & Bryan, 2011). Caution must therefore be taken not to overgeneralize from findings of incidental vocabulary learning from reading one individual text to reading in general.

The studies reviewed here varied in ecological validity from the most naturalistic that used authentic novels as the reading material without experimentally controlling the context of exposure (Godfroid et al., 2017; Saragi et al., 1978) to non-naturalistic studies in which participants read individual sentences with only a few exposures to novel words (Mestres-Missé et al., 2008, 2007; Williams & Morris, 2010). Some recent studies have attempted to find a balance between these approaches (Batterink & Neville, 2011; Pellicer-Sánchez, 2016). Several additional differences between these studies could account for variation in acquisition success (e.g., number of items to be learned, measures used to assess learning, and whether participants learned both a novel word form and meaning or a novel word form to describe an already known concept). Number of exposures has consistently been found to be a strong predictor of acquisition success (Godfroid et al., 2017; Saragi et al.). Of the different aspects of vocabulary knowledge, including receptive and productive knowledge of the word form, meaning, and usage (P. Nation, 2001), productive knowledge of word meanings (assessed through cued recall) has been found to be the most difficult to acquire (Batterink & Neville, 2011; Godfroid et al., 2017; Pellicer-Sánchez, 2016) and may therefore require more exposures for successful learning. Little research has investigated the incidental learning of word meanings in isolation from the acquisition of novel word forms, that is, learning new meanings for familiar words.

Learning New Meanings for Familiar Word Forms

Some researchers have suggested that learning new meanings for already known words may be easier than learning entirely new words because learners' attention is not divided between learning a novel word form and mapping a new meaning onto that word (Storkel & Maekawa, 2005; Storkel, Maekawa, & Aschenbrenner, 2013). However, others have argued that learning new meanings for familiar words may be harder due to competition between the old and new meanings (Fang, Perfetti, & Stafura, 2016; Rodd et al., 2012). Furthermore, children are slower to learn words with more than one meaning (Casenhiser, 2005) because they find learning one-to-many mappings between word forms and meanings harder than direct one-to-one mappings. It may also be harder to learn a new meaning for a word with an already well-established meaning than to learn the two meanings simultaneously due to the need to inhibit the more active dominant representation for the preexisting meaning of the word (Dautriche, Chemla, & Christophe, 2016). Fang et al. argued that learning new meanings for known words is a two-phase process in which familiarity with the word form may facilitate initial learning with the first few exposures, but inhibition due to meaning competition comes into play later after subsequent exposures to the newly ambiguous word. Overall, it appears that a greater number of exposures may be required for new meanings for familiar words to reach the same level of learning as for entirely novel words and that memory of these new meanings may be less stable after a long delay.

Another factor that can affect the learning of meanings for familiar words is the relationship of the new meanings to the preexisting meanings of the words. There are two types of semantic ambiguity that can arise in language: polysemy and homonymy. Polysemy is when words have multiple semantically related senses of the same underlying meaning, for example, a computer *virus* is related in function to a medical *virus* (Rodd et al., 2012). Homonymy, on the other hand, is when words have multiple semantically unrelated meanings that arise by chance (e.g., the *bark* of a tree/dog), and it is less common than polysemy (Rodd et al., 2002). Rodd et al. (2012) compared learning new semantically related meanings to learning new semantically unrelated meanings for words. They found that recall of the new meanings for the previously unambiguous words was better for the newly learned polysemous meanings than for the homonyms, which were harder to learn. Participants also responded more quickly to the newly polysemous words than to the newly homonymous words in a lexical decision task. These findings were consistent with those of previous studies showing that, while polysemy facilitates word recognition, homonymy delays recognition due to competition from semantically unrelated meanings (Rodd et al., 2002; Rodd, Gaskell, & Marslen-Wilson, 2004). This effect likely arises because words with multiple related senses have highly overlapping semantic representations that cause them to settle more quickly into the appropriate representation. In contrast, for words with multiple unrelated meanings, the mutually exclusive representations of both meanings are initially activated, with semantic competition between these meanings increasing the time needed for a single meaning to be settled on (Rodd et al., 2004). These same underlying mechanisms may explain why homonyms are harder to learn than polysemes (Rodd et al., 2012). Although rarer in language than polysemy, homonymy poses a unique and interesting challenge to learners because they must acquire a novel word meaning alone and map it onto a known word form without support from the existing representations for that word. This study therefore focused on the learning of homonyms, for which the new meaning was not semantically related to the already known meaning of the word.

The Present Study

The story-reading procedure used in this study involved a combination of the naturalistic elements of the studies using authentic texts (Godfroid et al., 2017; Saragi et al., 1978) and careful within-item experimental control of the number of exposures, similar to methods used by Batterink and Neville (2011) and Pellicer-Sánchez (2016). The homonyms were encountered incidentally within stories that participants read for comprehension with no instruction to memorize the new meanings of the words. This study investigated the effect of the number of exposures on adults' incidental learning and long-term retention of new meanings for familiar words in L1. To the best of our knowledge, this is the first study to use this more naturalistic approach to explore the incidental learning of homonyms because previous studies investigating this phenomenon have used more intentional and less naturalistic learning conditions (Fang & Perfetti, 2017; Fang et al., 2016; Rodd et al., 2012).

Participants encountered new word meanings through reading a single text—one of four short stories that had been specifically written for this experiment. The stories included novel, invented meanings for existing unambiguous English words (e.g., a *foam* is a type of safe concealed within a piece of furniture), with the novel meanings conveyed through the stories' narratives. We manipulated the number of exposures within subjects and within items. Each story contained four words with novel meanings, which were each presented two, four, six, or eight times throughout the text, counterbalanced across participants. We assessed participants' knowledge of the new meanings through cued recall of the new meanings when they were presented with the words and through cued recall of the word forms when they were presented with definitions of the new meanings. We tested participants' memory both immediately (following a short filler task) and 1 week after training. We predicted that participants' accuracy in recalling the novel meanings and identifying which of the meanings paired with each word would be very low for only two exposures but would increase gradually as the number of exposures to the words with their novel meanings increased. We further predicted that participants would experience significant forgetting of the novel meanings after the 1-week delay but that they would have better long-term retention with a greater number of exposures.

Method

Participants

Sixty-four participants took part in the experiment (M_{age} = 31.9 years, SD = 9.2, range: 18–47; 32 female). The participants were recruited through the website Prolific Academic (Damer & Bradley, 2014). All participants were monolingual native speakers of British English who were paid £3 for their participation in the first session of the experiment and £1 for the second session 1 week later. Of the 64 participants who completed the first session, 52 completed the delayed test a week later (81.3%). An additional 18 participants were excluded from the study—11 for not meeting the language background criteria, 6 for getting more than one multiple-choice comprehension question wrong when reading the story, and 1 due to a technical issue.

Materials

Novel Word Meanings

The stimuli consisted of 16 English nouns (see Appendix S1 in the Supporting Information online for the full list of words and their definitions) with only a single meaning in the Wordsmyth dictionary (Parks, Ray, & Bland, 1998). Although all words had only a single dictionary meaning, most had several different related senses of that meaning, that is, they were polysemous but not homonymous (see Appendix S2 in the Supporting Information online for descriptive statistics of the stimuli in each of the stories). Novel concrete noun meanings were chosen to be semantically unrelated to the original meanings of the words,[1] and a pretest confirmed that they were unrelated (see below). Thirteen of the novel meanings were adapted from the stimulus set used by Rodd et al. (2012), and three additional meanings were devised following the same specifications. The new meanings were designed to be semantically diverse and consisted of hypothetical innovations (5), natural phenomena (2), invented objects (2), social phenomena/traditions (5), a technical term (1), and a colloquial term (1). One sentence was written for each of the stimulus words to give a definition of the new meaning, for example, "A foam is a safe that is incorporated into a piece of furniture with a wooden panel concealing the key lock, and each is individually handcrafted so that no intruders are able to recognize the chief use of the furniture." The sentences were matched for length (M = 32.9 words, SD = 3.7). Each new meaning had three distinguishing semantic features to maintain a similar level of complexity for each new concept, for example, for *foam*: "a safe inside a piece of furniture," "has a hidden key lock," and "individually handcrafted to fool intruders." These sentences were given to the authors of the stories to be incorporated into story narratives.

Abbreviated versions of these definition sentences were also written for use in the test task in which participants were asked to recall the word forms that paired with the definitions.

Relatedness Pretest

To ensure that the new word meanings were semantically unrelated to the existing meanings of the words, a pretest was carried out using a separate group of 20 monolingual native British English speakers ($M_{\text{age}} = 30.1$ years, $SD = 10.0$, range: 18–52; 11 females). They rated the relatedness of the novel meanings presented in the definition sentences to the real, existing meanings of the words that they knew. The stimuli for the pretest were the sentences giving definitions of the new meanings, each paired with a semantically unrelated word form. Each of the new meanings was also paired with a semantically related word form from a larger set of items not used in the present study,[2] for example, *slot* for "a safe that is incorporated into a piece of furniture with a wooden panel concealing the key lock, and each is individually handcrafted so that no intruders are able to recognize the chief use of the furniture." Although none of these semantically related form–meaning pairs were used in this study, these provided a frame of reference on a scale from 1 (*highly unrelated*) to 7 (*highly related*). The pretest was split into two versions, with participants pseudorandomly and evenly assigned to one of the two versions so that they saw each new meaning only once, paired with either the semantically unrelated or related word form. There were therefore 10 data points for each meaning rated with its intended unrelated word. The results showed that, as intended, the 16 word form–meaning pairs used in this study were perceived as unrelated to the existing meanings of the words ($M_{\text{rating}} = 1.8$, $SD = 0.3$, range: 1.3–2.6).

Short Stories

Four separate stories were written, each incorporating four of the stimulus words in the context of their new meanings. One of the stories (Story 1: *Pink Candy Dream*) was written by a professional children's author and former psycholinguistics researcher; the other three stories (Story 2: *Prisons*, Story 3: *Reflections upon a Tribe*, and Story 4: *The Island and Elsewhere*) were written by an unpublished student author. The authors were provided with a list of words with their novel meanings (the 16 items included in this study and 16 items not selected by them for inclusion in the stories) grouped broadly into four themes—one for each of the stories. They were asked to choose four of the items in each theme to incorporate into a story, selecting the items that they felt would best fit together into a plausible narrative, with each word to appear eight times so as to provide information about its new meaning through the context. The stories were similar in length (Story 1: 2,307 words, Story 2: 2,320 words, Story 3: 2,446 words, Story 4: 2,330 words) and were designed to be similar in writing style and to be engaging for an adult audience. Each of the stimulus words appeared a total of eight times at naturally distributed positions within one of the four stories, with no stimulus word occurring in more than four consecutive sentences. The number of different words with novel meanings in each of the stories was 0.2% of the total number of words. This is similar to the estimated percentage of novel Nadsat words (0.4%) in *A Clockwork Orange* (Saragi et al., 1978), indicating that the new word meanings were naturally distributed and potentially learnable from the stories. On the first presentation of a stimulus word, sufficient information was given to allow the reader to derive the new meaning from the context from the first exposure, for example, "'Yes,' I murmured, breathing again. 'I *knew* it! It's a foam.' The ornate *chaise longue* was no ordinary piece of furniture but concealed a built-in safe with an intricate key-operated locking system." The amount of information about each new meaning in subsequent exposures varied naturally with the story narratives. None of the stimulus words appeared in any of the stories in the context of its real, existing meaning.

We then modified the short stories to vary the number of exposures to each stimulus word along with its novel meaning. We manipulated the number of exposures by removing some of the occurrences of the stimulus words to leave only two, four, six, or eight occurrences. We achieved this by replacing some of the instances of the stimulus word with words or phrases synonymous to the novel meaning (e.g., *foam* was replaced with *safe* or *hidden safe*) or in a few cases by simply omitting the word where it was not possible to use a synonym in the context of the narrative. This approach ensured that we held constant the amount of semantic content provided for each word regardless of the number of exposures. In all of the exposure conditions, the first and final occurrences of the stimulus word were kept in the story to minimize any primacy or recency effects. In the two-exposures condition, these were the only occurrences. In the four- and six-exposures conditions, the additional occurrences of the stimulus words that we retained were those appropriate to the natural narrative of the stories. In the eight-exposures condition, we retained all of the target words. Each of the four stories contained one stimulus item in each of the four exposure conditions: two, four, six, and eight exposures, so that each participant saw an item in each of the conditions. Additionally, we created four versions of each of the stories so that each stimulus item appeared in each exposure condition across participants.

Design

Each participant read only one of the four stories. The first independent variable of number of exposures to a word with its novel meaning was manipulated within participants and within items. Each participant was trained on four words that appeared in the story two, four, six, and eight times, respectively. To ensure that each stimulus item was seen an even number of times in each exposure condition across participants, we created 16 versions of the experiment (four per story). We assigned participants pseudorandomly and evenly to 1 of the 16 versions of the experiment, with four participants assigned to each version. The second independent variable of time of test (immediate vs. 1 week later) was also a within-subjects variable (based on the 52 participants who completed both sessions). The dependent variables measured were accuracy in cued recall of the novel meanings and cued recall of the word form paired with each novel meaning.

Procedure

We conducted the experiment online using Qualtrics (Qualtrics, 2015) and described the experiment to participants as "a study of different reading styles and the ability to understand texts." We informed participants that they would be reading a short story and answering comprehension questions about what they had read and that this would be followed by a short vocabulary test and then some questions about their personal reading style. They were not made aware that they would encounter novel word meanings in the story nor were they told to try to learn them or that their memory for these novel word meanings would be tested. After the participants had completed the first session of the experiment, they were not informed that they would be invited to complete a delayed test a week later. This was to discourage the use of deliberate memorization techniques by the participants and to discourage rehearsal of the items over the week-long delay.

Each story was divided into five pages of roughly even length and displayed on-screen one page at a time. After each page, a multiple-choice comprehension question appeared on a separate screen asking about details of the story's plot from the preceding page without probing details of the novel word meanings. Participants were instructed to read the story closely and to answer a question about what they had just read after each page. They were not given opportunities to reread previous pages. Participants had to select the one correct answer from four options that appeared in a randomized order. We designed the questions to be very easy for any participant who had fully understood the text. Participants were excluded if they got more than one of the five comprehension questions wrong; as previously stated, six participants were excluded on this basis.

After they had finished reading the story, participants completed the Mill Hill Vocabulary Test (Set A: Multiple Choice), containing 34 items (Buckner et al., 1996; Raven, Raven, & Court, 1998) as a filler task between the training phase and the testing phase. For each test item, participants were required to select one word from a list of six options that most closely matched the meaning of the presented word. None of the stimulus words appeared in the vocabulary test. The purpose of this task was to counteract any recency effects of memory for stimulus items encountered toward the end of the story. Participants were then given a cued recall test of the novel word meanings that they had encountered in the story. Participants saw each of the four stimulus words that had appeared in the story and were asked to recall the appropriate novel meaning and type it into a blank text box. They were encouraged to provide as much detail as possible and to try to answer in full sentences even if they were unsure of their answer. If they could not remember anything about the new meaning for the word, they were instructed to type "don't know." For this test and the subsequent test of cued recall of the word forms, each item was presented one at a time and the order of presentation was randomized separately for each participant.

Participants were next given a cued recall test for the word forms that paired with each novel meaning. Participants were presented with short sentences that defined each of the novel word meanings. For each definition, participants were asked to recall the word that it described and to type it into a blank text box. The definition sentences used for this test were abbreviated versions of the original definition sentences that were provided to the story authors. Although the sensitivity of this second test was expected to be reduced compared to the initial test (due to priming of the word forms during the initial test), it was included to provide a measure of memory that could be used in the event that participants' performance was at floor on the initial test.

After completing both cued recall tests, participants provided demographics details, rated how enjoyable and clear they had found the story on a 7-point scale, and answered questions about their reading style and habits. The primary purpose of these questions was to maintain the cover story that the purpose of the study was to investigate reading styles and comprehension; hence, we did not analyze responses to these questions. Exactly 7 days after the main experiment had been made available to participants, we invited participants to participate in a brief unexpected follow-up to the experiment. Participants began the delayed test a mean of 7 days 0 hours 45 minutes after they had started the first session of the experiment (SD = 1 hour 34 minutes, range: 6 days 21 hours 42 minutes–7 days 5 hours 15 minutes). The delayed test session consisted of the same two cued recall tests in the same order as in the first session, with the order of test items again randomized separately for each participant in both tasks.

Data Analysis

Responses to the items of both cued recall tests were coded for accuracy by one of the experimenters, blind to condition, as either 1 (correctly recalled) or 0 (incorrectly recalled). The responses on the test of cued recall of the novel meanings were leniently coded as correct if at least one correct semantic feature was recalled (e.g., "a safe inside furniture" for *foam*). Any ambiguous or partially correct responses were resolved through discussion with another experimenter. The data were analyzed with logistic mixed-effects models using the lme4 package (Bates, Mächler, Bolker, & Walker, 2016a) and R statistical software (R Core Team, 2017). Four separate models were created: one for each of the two cued recall measures comparing accuracy between Day 1 and Day 8, which included only the participants who completed the tests at both time points ($n = 52$) and one for each of the two cued recall measures for all participants tested on Day 1 only ($n = 64$). These latter analyses aimed to verify that the data from this larger set of participants did not differ from the subset who chose to complete both sessions.

The four models all contained random effects for participants and items (with slopes for exposure condition) and a fixed effect for exposure condition (four levels: two, four, six, or eight exposures). The contrasts for this exposure condition variable were defined using orthogonal polynomial coding, with three separate contrasts to assess potential linear (two: –3, four: –1, six: 1, eight: 3), quadratic (two: 1, four: –1, six: –1, eight: 1), and cubic (two: –1, four: 3, six: –3, eight: 1) trends in the data. We adopted this approach because it is of greater theoretical interest to characterize the overall trend of the impact of number of exposures on acquisition of new meanings for familiar words rather than using conventional contrasts to focus on differences between individual exposure conditions. The two models comparing performance between Day 1 and Day 8 had an additional fixed effect for time, with the contrast defined using deviation coding (Day 1: –0.5 vs. Day 8: 0.5), and a fixed effect for the interaction between time and the number of exposures, which was created by multiplying time by each of the contrasts for exposure condition. These models also included random slopes for time (i.e., Day 1 vs. Day 8) and the interaction between this variable and exposure condition by participants and items.

The first attempted model fit used the maximal random-effects structure, as recommended by Barr, Levy, Scheepers, and Tily (2013), which did not converge.[3] Following this, the models were simplified by removing only the correlations between the random slopes and random intercepts for the random effects by participants and items without removing any of the random slopes. Three of the four models converged at this stage; the model comparing the data from Day 1 and Day 8 for the cued recall of words did not converge. This model was simplified by instead removing the random intercepts by participants and by items, again leaving in all the random slopes, and this time leaving in the correlations between the random slopes, which allowed the model to converge. Therefore, all four analyses were carried out using models with simplifications of the maximal random-effects structure as recommended by Barr et al.

Significance of the main effects and interactions was assessed using likelihood ratio tests by comparing the full model to identical models with only each factor or interaction of interest removed in turn but leaving in any other interactions or main effects involving that factor or interaction, keeping the random-effects structure intact. In the case of a significant effect of number of exposures, an additional analysis was run to determine whether there was a significant linear, cubic, or quadratic trend in the data. This was again assessed through likelihood ratio tests by comparing the full model to models with each of the components removed in turn. All data and analysis scripts for this study are available via the Open Science Framework at https://osf.io/ybu6r.

Results

Cued Recall of Novel Meanings

The data for accuracy in cued recall of the novel meanings comparing performance between Day 1 and Day 8 ($n = 52$) showed a reasonably high level of accuracy even after only two exposures (Day 1: 38.5%, Day 8: 42.3%), appearing to increase in a positive linear trend with an increasing number of exposures (see Figure 1 and Appendix S3 in the Supporting Information online for the descriptive statistics). The data for the delayed test a week later showed the same pattern, and there appeared to be very little change in mean accuracy between these two time points. The analyses showed a significant main effect of number of exposures, $\chi^2(3) = 11.66$, $p = .009$,[4] and no significant effect of time of test, $\chi^2(1) = 0.63$, $p = .429$, therefore showing no evidence of a difference in accuracy between the immediate test and the delayed test a week later. There was also no significant interaction between time and number of exposures, $\chi^2(3) = 1.58$, $p = .664$. The trend analysis revealed that the

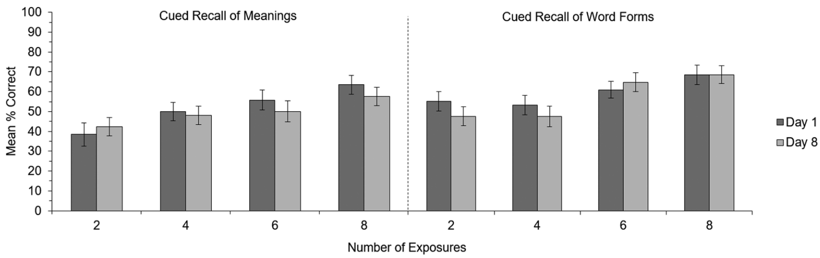

Figure 1 Mean percentage of correct responses across participants for cued recall of novel meanings and cued recall of word forms in each exposure condition when participants were tested on Day 1 (immediately after training) and on the delayed test on Day 8 ($n = 52$). The linear mixed-effects analyses were carried out on the raw binary accuracy data. However, percentage data are displayed in the graphs for ease of interpretation. Error bars show standard error of the means, adjusted for the within-participant design (Cousineau, 2005).

number of exposures had a significant positive linear effect on cued recall of new meanings, $\chi^2(1) = 11.32$, $p < .001$, and no significant quadratic effect, $\chi^2(1) = 0.001$, $p = .973$, nor cubic effect, $\chi^2(1) = 0.15$, $p = .700$.

The data for accuracy in cued recall of the novel meanings for all participants tested on Day 1 ($n = 64$) showed the same pattern as the data comparing performance between Day 1 and Day 8: a reasonably high degree of accuracy after only two exposures, which increased with an increasing number of exposures to the words with their new meanings (see Appendixes S3 and S4 in the Supporting Information online for the descriptive statistics and figure, respectively). The results again showed a significant main effect of number of exposures, $\chi^2(3) = 11.12$, $p = .011$. The trend analysis of the data also revealed a significant positive linear effect of number of exposures on cued recall of new meanings, $\chi^2(1) = 10.47$, $p = .001$, and no significant quadratic effect, $\chi^2(1) = 0.01$, $p = .929$, nor cubic effect, $\chi^2(1) = 0.65$, $p = .421$.

Cued Recall of Word Forms

The accuracy data for cued recall of the word forms that paired with each of the novel meanings comparing Day 1 to Day 8 ($n = 52$) showed that overall accuracy appeared to be higher in this test than in the cued recall of meanings test, although the pattern of the data appeared broadly similar (see Figure 1 and Appendix S3 for the descriptive statistics). These data again showed a high level of accuracy after only two exposures (Day 1: 55.8%, Day 8: 48.1%), with performance increasing gradually with a higher number of exposures. There was again very little change in accuracy between the tests on Day 1 and Day 8 across all exposure conditions. The results showed that the main effect of number of exposures was marginal but nonsignificant for this measure, $\chi^2(3) = 6.82, p = .078$. There was also no significant effect of time of test, $\chi^2(1) = 0.28, p = .599$, and no significant interaction between time and number of exposures, $\chi^2(3) = 0.99, p = .803$. Because the main effect of number of exposures was nonsignificant, any trends in the data were not assessed further. The data for accuracy in cued recall of the word forms for all participants tested on Day 1 ($n = 64$) showed the same pattern (see Appendixes S3 and S4 in the Supporting Information online for the descriptive statistics and figure, respectively). The results again showed no significant main effect of number of exposures, $\chi^2(3) = 3.95, p = .267$, so any trends in the data were not assessed further.

Discussion

The aim of this study was to investigate whether adult readers can learn novel meanings for known words incidentally from stories after encountering very few instances of the novel word meaning and to examine how well these meanings are retained 1 week after exposure. Participants' memory of novel meanings for previously unambiguous words was assessed using tests of cued recall of the novel meanings and of the word forms that paired with definitions of the new meanings. The participants were tested both immediately after training and after a 1-week delay.

Acquisition From Initial Exposures

Although there were substantial individual differences in performance, 38.5% of participants could correctly recall the new meaning for a known word after just two exposures in a single-story context when tested immediately after training. These findings are consistent with some of the studies that used online measures to look at incidental learning of novel words and their meanings (Batterink & Neville, 2011; Mestres-Missé et al., 2007; Pellicer-Sánchez, 2016). Pellicer-Sánchez found that L1 participants read novel words that were embedded in a naturalistic story context significantly faster after only one exposure. The findings are also in line with the ERP studies of Batterink and Neville and Mestres-Missé et al., which both showed evidence of semantic integration after only a few exposures to novel nonword labels for existing meanings.

Conversely, our results are perhaps inconsistent with those found in previous studies using behavioral measures of explicit memory for novel words and their meanings. Both Williams and Morris (2010) and Mestres-Missé et al. (2007, 2008) found much higher accuracy in meaning recognition (66–69%) after only one or three exposures, respectively. However, there are a number of differences between their studies and our study that could account for the lower levels of acquisition that we found. Although participants learned both the forms and meanings of a greater number of words in both prior studies than participants did in our study, they did so from reading in the more constrained context of short sentences. In these previous studies, participants had to acquire a new word form and map it onto a known concept that was easy for them to deduce from the sentences. This is quite different from our study in which participants had to acquire a novel concept from a broader context and map it onto an already known word form. Furthermore, these previous studies used only very simple measures of meaning recognition, which Pellicer-Sánchez (2016) noted is much less difficult to acquire than productive knowledge of word meanings measured through cued recall.

Perhaps the study most comparable to this study in terms of learning conditions and explicit measures of learning is that of Pellicer-Sánchez (2016). Although Pellicer-Sánchez did not measure acquisition after different numbers of exposures, accuracy in cued recall of the meanings for novel words was 65.3% after eight exposures for participants reading in their L1. This is close to the level of meaning recall found in this study after eight exposures (63.5%), suggesting that learning new meanings for familiar words may not be harder than learning new words and their meanings. However, participants in Pellicer-Sánchez's study were trained on more items (six) than in this study (four) and with the same number of exposures to all items. Further research comparing the acquisition of homonyms and nonhomonyms directly within a single study is therefore required.

The Impact of Increasing Exposures

As we had predicted, the number of exposures influenced learning, with a linear increase in performance on cued recall of the new meanings with an increasing number of exposures to stimuli in the written text. The data for cued recall of word forms showed roughly the same trend although no significant main effect of number of exposures was found. This was most likely due to performance on this second task having been enhanced by priming effects from the presentation of the word forms in the prior test of cued recall of the new meanings, although no feedback was provided to participants on either task. The finding of a significant overall effect of number of exposures is consistent with previous studies on incidental learning of word forms and their meanings, where number of exposures was shown to be a strong predictor of learning (Godfroid et al., 2017; Pellicer-Sánchez, 2016).

More importantly, in this study, the trend analyses for the significant effects of number of exposures on cued recall of the new meanings showed that, within the exposure range tested here, recall accuracy increased linearly as the number of exposures increased. At the immediate test, recall accuracy was reasonably good—38.5% after only two exposures. However, the percentage increase in recall accuracy for each subsequent increase of two exposures was not nearly as high as that attained for the first two exposures. There was a steady incremental increase of 8.3% on average with each additional two exposures up to a maximum of 63.5% accuracy with eight exposures. The large difference between recall accuracy for the initial two exposures and the much smaller average increase for each subsequent two exposures suggests that the first one or two exposures are especially important for the acquisition of homonyms. The findings of previous eye-tracking studies (Godfroid et al., 2017; Pellicer-Sánchez, 2016) have suggested that this may be because more time is spent reading and processing words during the initial exposures.

These results suggest that the initial few exposures have a disproportionally large impact on learning, while subsequent exposures all have a similar, lower level of impact. The positive linear pattern in the data likely arises due to a gradual dilution of the contribution of the initial exposures with an increasing total number of exposures (although see Bisson, van Heuven, Conklin, & Tunney, 2014, for an alternative explanation of similar findings). However, had we tested larger numbers of exposures, it is likely that learning gains would eventually have plateaued, similar to the pattern seen in the eye-tracking and ERP studies (Batterink & Neville, 2011; Mestres-Missé et al., 2007; Pellicer-Sánchez, 2016), where processing of novel words became indistinguishable from processing of known words after a few exposures. Based on previous research comparing the learning of homonyms to polysemes (Rodd et al., 2012), we would predict that the incidental learning of new semantically related meanings for known words would be even easier than learning new semantically unrelated meanings as in this study. The initial exposures may have an even greater impact on the learning of polysemes due to support from the existing representations for the word's meaning. Learning gains would also likely plateau after fewer exposures than for learning homonyms.

The learning gains seen in this study are specific to the reading of a single text as opposed to reading multiple texts. Some studies of L2 learning have found higher levels of vocabulary acquisition from more extensive reading (M. Horst, 2005; Webb & Chang, 2015) than usually reported in studies of learning through a single text. This may be due to several contributing factors, such as increased spacing between encounters and greater contextual diversity of individual exposures (K. Nation, 2017). The stimuli in this study were highly contextually constrained within the stories. It is likely that incidental learning of homonyms would be more successful if encounters were distributed across separate stories. Further research is required to explore learning new meanings for familiar words through reading multiple texts. This would help build a clear picture of how adults typically learn L1 vocabulary.

Long-Term Retention

Perhaps most surprisingly, in contrast to the predictions, participants showed no significant forgetting of the new meanings at a retest 1 week later as shown by scores on both measures, and long-term retention was not differentially affected by the number of exposures. None of the previously discussed studies assessed long-term retention for participants reading text in their L1 (Batterink & Neville, 2011; Godfroid et al., 2017; Mestres-Missé et al., 2008, 2007; Pellicer-Sánchez, 2016; Saragi et al., 1978). However, Pellicer-Sánchez retested some of her study's group of proficient L2 learners following a 2-week delay. She also found no significant forgetting between the immediate and delayed tests on measures of meaning recall, meaning recognition, and form recognition.

In contrast, in another study in which intermediate L2 learners read a level-appropriate English novel, Waring and Takaki (2003) found that memory for novel words decreased in general after 1 week and had drastically decayed after 3 months. Contrary to our study, they also found that words with a greater number of exposures were more resistant to forgetting over time. However, there were considerable differences in the learning conditions of these previous studies (Pellicer-Sánchez, 2016; Waring & Takaki, 2003) in which participants read and learned new words in their L2 because participants' general L2 vocabulary knowledge undoubtedly had an impact on their acquisition success. The differences between these studies and our study, in which participants read and learned new meanings in their L1, therefore make direct comparisons difficult.

A possible explanation for the maintained levels of recall accuracy seen over the course of 1 week includes the testing effect (e.g., Roediger & Karpicke, 2006). This describes the phenomenon whereby the inclusion of a memory test immediately following training can facilitate long-term retention due to extra retrieval practice giving a boost to learning even in the absence of any feedback on performance. In our study, the immediate tests could (even in the absence of feedback) have boosted performance on the delayed test. However, as was the case in Pellicer-Sánchez's (2016) study, our participants did not encounter the stimuli between the two test sessions, and they were not aware of the retest beforehand so had no cause to rehearse the stimuli during the preceding week. The results are therefore still a good indication of the long-term retention of new meanings for familiar words 1 week after incidental acquisition. Future studies should take into account the additional impact of an immediate test on long-term retention, for example, by testing only some of the items immediately following training.

Another potentially important factor for the preservation of memory of the new meanings for familiar words 1 week later is offline consolidation during sleep. Sleep has previously been shown to play an important role in learning new spoken word forms (for a review, see Davis & Gaskell, 2009). Although it was not possible in our study to tell at what point consolidation occurred, it is clear from participants' long-term knowledge of the new word meanings that some lexical configuration had taken place, that is, information about the words' new meanings and usage had been correctly obtained and retained (Leach & Samuel, 2007). Nor did our study show at what point lexical engagement occurred, for example, at what point the new meanings were able to compete with existing meanings for access (Leach & Samuel, 2007). Future research could look at the more fine-grained acquisition of new meanings for known words using an implicit measure to investigate at what point these separate stages of learning occur.

Implications for Future Research

The story-reading paradigm used in this study provided ideal training conditions for us to investigate incidental vocabulary acquisition from reading. The training method has good ecological validity: Adults acquire new meanings for known words incidentally while reading or listening for comprehension, and fantasy and science fiction stories are often a source of novel concepts to be mapped onto existing words (e.g., a *grim* is a large black ghostly dog and omen of death in the Harry Potter series of novels by J. K. Rowling). Because the stories were custom written by authors specifically for use in our study, this allowed for complete experimental control over the number of exposures to the stimuli through the narrative in a within-items design. Most importantly, this allowed for control over potentially correlated or confounding factors, such as the centrality of target items to the story's plot and properties of the words. A limitation, however, is that sufficient information was included to elucidate the new meaning for a word on the first exposure. Although this may happen sometimes in authentic texts, this is often not the case, and the amount of contextual information provided in individual exposures has been shown to influence vocabulary gains for L2 learners (Webb, 2008). However, this was necessary in the design of our study to ensure that the key semantic information was available in all of the exposure conditions. Finally, this paradigm has the potential to be adapted for use in future studies to look at how multiple different factors might influence efficiency of learning and retention of new meanings for familiar words, such as attention, depth of processing, modality of story presentation, contextual diversity, repetition of stories (e.g., M. Horst, 2005; Webb & Chang, 2015), and the role of sleep.

Conclusion

In conclusion, this study extends what has previously been found in the L2 incidental vocabulary learning literature (e.g., Pellicer-Sánchez, 2016) to the learning of new meanings for previously unambiguous words in language users' L1. Some participants (38.5% at the immediate test) were able to successfully learn these meanings after just two exposures to familiar words with their novel meanings in a story context. Subsequent exposures additionally improved participants' performance: Learning increased linearly with an increase in the number of exposures in a cumulative incremental manner. Furthermore, knowledge of new meanings for known words was maintained well over the course of 1 week, regardless of the number of exposures during learning. Overall, these findings demonstrate the remarkable success with which adults learn new meanings for known words incidentally while reading as they do in everyday life, as previously unambiguous words become homonyms.

Final revised version accepted 11 May 2018

Notes

1 The new meanings were created by swapping around pairs of words from a larger stimulus set of semantically related meanings (32 items in total, 16 of which were used in our study). None of the previous semantically related meanings for the words were used in any of the stories.
2 All 32 meanings from the larger set of stimuli were included in the relatedness pretest: The 16 items used in the present study and 16 additional items not included in our study. Rating data are given only for items included in the study.
3 The BOBYQA (Bound Optimization BY Quadratic Approximation) optimizer was used as per recommendations by Bates, Mächler, Bolker, and Walker (2016b) for dealing with model convergence issues.
4 Unfortunately, it was not possible to obtain reliable measures of effect sizes (such as odds ratios and 95% confidence intervals) for the reported statistical contrasts as the linear mixed-effects model included a factor with more than two levels.

References

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. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2016a). lme4: Linear mixed-effects models using "Eigen" and S4. R package (version 1.1-12) [Computer software]. Retrieved April 18, 2016, from https://cran.r-project.org/web/packages/lme4/index.html

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2016b). lme4: Linear mixed-effects models using "Eigen" and S4 [Software manual]. Retrieved April 18, 2016, from https://cran.r-project.org/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. https://doi.org/10.1162/jocn_a_00013

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

Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., et al. (1996). Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, *93*, 14878–14883. https://doi.org/10.1073/pnas.93.25.14878

Casenhiser, D. M. (2005). Children's resistance to homonymy: An experimental study of pseudohomonyms. *Journal of Child Language*, *32*, 319–343. https://doi.org/10.1017/S0305000904006749

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Damer, E., & Bradley, P. (2014). Prolific academic [Computer software]. Retrieved November 24, 2015, from https://www.prolific.ac

Dautriche, I., Chemla, E., & Christophe, A. (2016). Word learning: Homophony and the distribution of learning exemplars. *Language Learning and Development*, *12*, 231–251. https://doi.org/10.1080/15475441.2015.1127163

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences*, *364*, 3773–3800. https://doi.org/10.1098/rstb.2009.0111

Fang, X., & Perfetti, C. A. (2017). Perturbation of old knowledge precedes integration of new knowledge. *Neuropsychologia*, *99*, 270–278. https://doi.org/10.1016/j.neuropsychologia.2017.03.015

Fang, X., Perfetti, C., & Stafura, J. (2016). Learning new meanings for known words: Biphasic effects of prior knowledge. *Language, Cognition and Neuroscience*, *32*, 637–649. https://doi.org/10.1080/23273798.2016.1252050

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., et al. (2017). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, *21*, 563–584. https://doi.org/10.1017/S1366728917000219

Henderson, L. M., Devine, K., Weighall, A., & Gaskell, G. (2015). When the daffodat flew to the intergalactic zoo: Off-line consolidation is critical for word learning from stories. *Developmental Psychology*, *51*, 406–417. https://doi.org/10.1037/a0038786

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology*, *2*, 1–11. https://doi.org/10.3389/fpsyg.2011.00017

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

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

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Malden, MA: Blackwell. https://doi.org/10.1002/9780470756492.ch12

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*, 306–353. https://doi.org/10.1016/j.cogpsych.2007.01.001

Mestres-Missé, A., Càmara, E., Rodriguez-Fornells, A., Rotte, M., & Münte, T. F. (2008). Functional neuroanatomy of meaning acquisition from context. *Journal of Cognitive Neuroscience*, *20*, 2153–2166. https://doi.org/10.1162/jocn.2008.20150

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex*, *17*, 1858–1866. https://doi.org/10.1093/cercor/bhl094

Nation, K. (2017). Nurturing a lexical legacy: Reading experience is critical for the development of word reading skill. *Npj Science of Learning*, *2*, 3. https://doi.org/10.1038/s41539-017-0004-7

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

Nation, P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language*, *27*, 136–145.

Parks, R., Ray, J., & Bland, S. (1998). *Wordsmyth English dictionary-thesaurus* [Electronic version]. Chicago: University of Chicago Press. Retrieved October 24, 2014, from https://www.wordsmyth.net

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

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.

Qualtrics. (2015). Qualtrics survey software [Computer software]. Provo, UT: Author. Retrieved November 5, 2015, from https://www.qualtrics.com

R Core Team. (2017). R: A language and environment for statistical computing (Version 3.3.3) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved March 20, 2017, from https://www.r-project.org

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales. Section 5: The Mill Hill vocabulary scale*. San Antonio, TX: Harcourt Assessment.

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., et al. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory & Cognition*, *40*, 1095–1108. https://doi.org/10.3758/s13421-012-0209-1

Rodd, J. M., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*, 245–266. https://doi.org/10.1006/jmla.2001.2810

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science*, *28*, 89–104. https://doi.org/10.1016/j.cogsci.2003.08.002

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

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. https://doi.org/10.1017/S0272263199004039

Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, *6*, 72–78. https://doi.org/10.1016/0346-251X(78)90027-1

Spivey, M., & Cardon, C. (2015). Methods for studying adult bilingualism. In J. W. Schwieter (Ed.), *The Cambridge handbook of bilingual processing* (pp. 108–132). Cambridge, UK: Cambridge University Press.

Storkel, H. L., & Maekawa, J. (2005). A comparison of homonym and novel word learning: The role of phonotactic probability and word frequency. *Journal of Child Language*, *32*, 827–853. https://doi.org/10.1017/S0305000905007099

Storkel, H. L., Maekawa, J., & Aschenbrenner, A. J. (2013). The effect of homonymy on learning correctly articulated versus misarticulated words. *Journal of Speech, Language, and Hearing Research*, *56*, 694–707. https://doi.org/10.1044/1092-4388(2012/12-0122)

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. https://doi.org/10.1093/applin/aml048

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*, 232–245. https://doi.org/10.1177/1362168814559800

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

Williams, R., & Morris, R. (2010). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, *16*, 312–339. https://doi.org/10.1080/0954144034000019

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. List of Stimulus Words.
Appendix S2. Descriptive Statistics for Stimulus Sets.
Appendix S3. Descriptive Statistics.
Appendix S4. Additional Figures.