The Puliguous Effects of Context Length on Incidental Word Learning

by

Kacey Lauren Wochna Class of 2010

A thesis submitted to the faculty of Wesleyan University in partial fulfillment of the requirements for the Degree of Bachelor of Arts with Departmental Honors in Psychology

Abstract

Most of the words in a person's vocabulary are acquired during the course of natural reading, through a process called incidental word learning. Incidental word learning is mediated by a number of factors concerning word, context, and reader characteristics. The current study investigated the effects of one such factor, context length, using a combination of word learning and eye-tracking methodologies. Participants read novel words embedded in either sentence or paragraph contexts while their eyes were tracked. Eye movement measures, primarily gaze duration and total time, demonstrated that novel words had longer reading times than real words, and that words in sentence contexts had longer reading times than words in paragraphs. The increased processing devoted to words in sentences may result in an increased rate of learning for novel words. This effect was not verified by the posttest vocabulary assessment, which did not find significant word learning. The low rate of learning observed in this study is attributed to the absence of any benefit from partial word knowledge, information that readers possess about words that are unknown or largely unknown. Research has shown that words are easier to learn if they are partially known. This study argues that partial word knowledge is more influential and pervasive than previously considered, and that as a result, the rates of learning reported by many incidental word learning studies may be inflated.

The Puliguous Effects of Context Length on Incidental Word Learning

Despite decades of research on vocabulary size, there is little consensus over the number of words in the English language or the average number of words in a person's vocabulary at any given age. Nagy and Anderson (1984) offer the largest estimates to date, identifying over 88,000 distinct words in the language and reporting the vocabulary of an average high school student to be between 25,000 and 50,000 words. The heterogeneity of these estimates reflects variation in how word knowledge is tested, the source used to represent the corpus of words, and how *word* is defined; Nagy and Anderson, for example, define a distinct word as a *word family*, a group of morphologically related words whose meanings can be easily determined if one word from the family is known. Children, even those who do relatively little reading, encounter between 16,000 and 24,000 unknown words every year (Nagy, Anderson, & Herman, 1987). Based on the estimate that the average high school senior knows 40,000 words, a child's vocabulary must grow by approximately 3,000 words a year (Nagy, Herman, & Anderson, 1985; Nagy & Herman, 1984).

Direct Vocabulary Instruction

How are these new words learned? When children first acquire language, words are learned through oral conversation. However, typical oral conversation uses only the most common words from the available corpus — even the conversations of college graduates contain fewer rare words than the average children's book (Beck & McKeown, 2007; Cunningham & Stanovich, 1991; Hayes & Ahrens, 1988). Another potential source of learning is direct vocabulary instruction, an activity that takes

place in most primary and many secondary schools (Scott, Jamieson-Noel, & Asselin, 2003). A classic example of classroom vocabulary instruction is the keyword lesson, a method in which a set of words to be learned is selected from a story that the class is reading. The words identified as keywords are usually semantically unfamiliar to students, important for comprehension of the story, and useful in a variety of contexts (Beck, McKeown, McCaslin, & Burkes, 1979). A wide range of exercises are used to teach keywords, with varying efficacy (e.g., Beck, Perfetti, & McKeown, 1982; Biemiller & Boote, 2006; Jenkins, Stein, & Wysocki, 1984; Scott et al., 2003). Instructional methods that expose students to the semantic, orthographic, phonologic, syntactic, and contextual features of words have been found to be positively correlated with vocabulary growth, and more effective than methods that focus on only a single feature (for details on two such methods, Rich Instruction and Anchored Word Instruction, see Beck & McKeown (2007) and Juel, Biancarosa, Coker, & Deffes (2003); for a general discussion, see Stahl & Fairbanks, 1986).

Although direct vocabulary instruction can result in vocabulary growth, it is evidently not the sole source of vocabulary learning. Observational studies and reviews of reading programs have indicated that in most classrooms, very little time is devoted to vocabulary instruction (Beck et al., 1979; Durkin, 1979; Juel et al., 2003, Scott et al., 2003). When instruction does take place, it is often one-dimensional and superficial, concentrating primarily on the rote memorization of definitions (Scott et al., 2003). While instructional methods could be improved in many cases, even rigorous vocabulary programs only cover a few hundred words a year, an amount insufficient to account for the vocabulary growth observed during the school years

(Beck et al. 1982; Nagy & Herman, 1984). Additionally, vocabulary knowledge continues to grow throughout adulthood, when instances of direct vocabulary instruction are generally thought to be infrequent. If direct instruction is inadequate, how is the remaining portion of a person's vocabulary learned?

Incidental vs. Intentional Word Learning

It appears that a substantial amount of vocabulary is learned during natural reading (e.g., Nagy & Herman 1984, 1987). Although it had long been assumed that people acquire new words in this manner, research on this method of learning, which is commonly called *incidental word learning*, has accumulated in only the past thirty years. For the purposes of this study, I will use the definition of incidental word learning asserted by Swanborn and de Glopper (1999): the derivation and learning of new word meanings during familiar reading circumstances. Explicitly, incidental word learning requires deriving the meanings of novel words through inferences made possible by predictive contexts, and retaining the meanings of those words in memory. These activities are done in order to comprehend the text at hand, rather than with any intent to learn new words.

Incidental word learning is often compared with word learning that occurs through the intentional derivation of meaning from context. For clarity, I will refer to this paradigm as *intentional word learning*. In the literature, this is often (though inconsistently) labeled derivational learning or learning from context, terms that could equally apply to incidental learning; 'intentional' emphasizes the conscious nature of the task in contrast with incidental learning. In studies of intentional word

learning, participants are instructed to either (a) derive the meanings of unknown words and provide definitions while the supporting context is still in front of them, or (b) use contextual information to actively try to learn unknown words, which requires remembering the meanings for later testing. Meaning derivation is also necessary for incidental learning, and research on the two processes has been highly interrelated.

Intentional word learning is more conducive to experimental study, and therefore has been more thoroughly researched. Findings from this research are sometimes extrapolated to incidental learning. The true extent of their similarity, however, has yet to be determined. Incidental word learning is generally considered to be a much more challenging process for readers (e.g., Jenkins & Dixon, 1983; Nagy et al., 1987). Direct experimental comparison has found that words are more easily acquired intentionally than incidentally (Konopak at al., 1987). Instruction in word meaning derivation may, with difficulty, be able to improve intentional learning (Fukkink (2002) and Fukkink and de Glopper (1998) found improvement; Kranzer (1988) and Carnine, Kameenui, and Coyle (1984) did not). Incidental learning, on the other hand, does not improve with instruction in intentional derivation (Fukkink, 2002; Kranzer, 1988). Though no benefit of instruction has been observed, studies have observed that the ability to derive meaning from context partially predicts the ability to learn incidentally (Fukkink, 2002; Kranzer, 1988). Conceptually, it seems unlikely that intentional and incidental word learning are entirely disassociated. Careful comparison may be worthwhile for understanding both processes, individually and in relation to each other.

Influences on Incidental and Intentional Word Learning

Sternberg and Powell (1983) and Jenkins and Dixon (1983) describe a number of mediating variables that might affect word learning from context. These hypothesized variables occur at the level of word, context, and participant. Today, most of these variables have been examined in the context of both intentional and incidental learning. While not all have been uniformly supported as influential factors, I will borrow their organization of the potential manipulations in word learning studies.

Word-level variables. Word-level variables concern characteristics of the unknown words to be learned, and include concreteness, conceptual difficulty, and part of speech. Concrete words (e.g., dog and table) are more easily learned than abstract words (e.g., love and justice) both intentionally (Fukkink, Blok, & de Glopper, 2001) and incidentally (Schwanenflugel, Stahl, & McFalls, 1997). Similarly, words with conceptually simple meanings (i.e., words that are exact synonyms for other words) are easier to learn than words with abstruse meanings both intentionally (Daneman & Green, 1986) and incidentally (Nagy et al., 1987). The results for part of speech have been inconsistent. Very young children are much quicker to acquire nouns than other parts of speech, as are adults in simulations of child-like learning (see Piccin & Waxman (2007) for an in-depth discussion of this 'noun advantage' and of the Human Simulation Paradigm). Despite these findings, studies of intentional learning have rarely found an effect of part of speech. Under certain conditions, an inhibition of noun-learning compared with verbs and adjectives has been seen (Bolger, Balass, Landen, & Perfetti, 2008). Schwanenflugel et al. (1997) likewise

found that children were worse at incidentally learning nouns than verbs or adjectives. They conjectured that this may occur because as children learn lower-frequency words, the heuristics they originally used to learn nouns (i.e., nouns refer to distinct whole objects) fail, requiring an adjustment in strategy.

Context-level variables. Context-level variables concern characteristics of the texts in which the unknown words are embedded, and include informativeness (constraint), conceptual difficulty, number of word occurrences, and variability. When context is more informative or constrained to predict the meaning of the word (sometimes called 'considerate' context), words are easier to learn through intentional derivation (Bolger et al., 2008; Daneman & Green, 1986; Fukkink, 2002). Likewise, when misleading contexts are introduced, learning becomes more difficult (Frishkoff, Collins-Thompson, Perfetti, & Callan, 2008). The effects of context informativeness have been much less unanimously reported for incidental learning. Herman, Anderson, Pearson, and Nagy (1987) found a facilitative effect of informativeness (as did Elley (1989), for contexts that were heard rather than read), but others have failed to find any significant learning benefit from increased context informativeness (Nagy et al., 1987; Schwanenflugel et al., 1997).

The counterintuitive finding that informativeness does not affect incidental word learning has yet to be fully explained. However, Nagy et al. (1987) suggest that the effect of informativeness found in other studies may actually be confounded with the effect of the conceptual difficulty of the context. They found this factor to significantly influence incidental learning: contexts containing simple concepts result in more learning than conceptually difficult contexts. The impact of context

conceptual difficulty may also explain the finding that intentional derivation is easier if a synonym for the unknown word is provided in the context than if derivation requires making an inference (Carnine et al., 1984).

Much debate surrounds the question of how many times a word must be read in order to learn its meaning. Under some conditions, minimal exposure (a single occurrence) may be sufficient to result in a measurable degree of word learning (e.g., enough knowledge of meaning to select the correct definition from a set) for both intentional (Bolger et al., 2008) and incidental learning (Nagy et al., 1987; Nagy et al., 1985). However, the number of words learned increases and the understanding of known meanings improves with multiple exposures during both intentional (Beck et al., 1982; McKeown, Beck, Omanson, & Perfetti, 1983) and incidental learning (Biemiller & Boote, 2006; Jenkins et al., 1984).

When unknown words occur multiple times in intentional learning tasks, learning is improved if the words appear in a variety of different contexts rather than in the same context multiple times (Bolger et al. 2008). An earlier study by Konopak et al. (1987) failed to find facilitative effects of multiple exposures or context variability. However, number of occurrences and variability are generally manipulated across contexts. When, as in the Konopak study, these variables are manipulated within a single context, number of occurrences and variability may be confounded with informativeness (Konopak et al. noted that when a word in their study occurred multiple times, the context in each instance was generally less informative than the context for a word which occurred only once). Beck et al. (1982) suggest that multiple encounters in varied contexts lead to a greater number of

retrieval cues, an outcome that should likewise benefit incidental learning. This variable has yet to be adequately tested for incidental learning, however, because as Swanborn and de Glopper (1999) explain, repeated exposure to unknown words across contexts may artificially draw attention to those words.

Participant-level variables. Participant-level variables concern individual differences, primarily in reading skill. In the following studies, the classification of skilled and less skilled was generally made according to a standardized measure of comprehension or a battery of verbal ability assessments. Skilled readers learn more words than less skilled readers both intentionally (Bolger et al., 2008; Freebody & Anderson, 1983; Frishkoff et al., 2008; Kranzer, 1988) and incidentally (Herman et al., 1987; Jenkins et al., 1984; Kranzer, 1988, Swanborn & de Glopper, 2002). Vocabulary learning in turn improves the comprehension of contexts containing both taught and untaught words. Because of this relationship, the performance of skilled and less skilled readers becomes even more disparate over time (Beck et al., 1982; Jenkins et al., 1984; McKeown et al., 1983; Stahl & Fairbanks, 1986). It has been theorized that skilled and less skilled readers employ different strategies when deriving word meanings from context (McKeown, 1985). Using ERP recordings, Frishkoff, Perfetti, and Westbury (2009) found different patterns of brain activation in skilled and less skilled readers during intentional derivation. Curtis (1987) suggests that qualitative, if not strategic, differences in word knowledge may persist even for words that are well-known to both skilled and less skilled readers: when asked to define words, skilled readers tend to use abstract descriptions while less skilled readers tend to give a contextualized response.

The differing-strategies theory is supported by the fact that the effects of some contextual variables differ according to reading skill, including reading purpose, proximity of informative context to the unknown word, and spacing of context reading. Skilled readers learn more words incidentally when instructed to read for the purpose of answering comprehension questions, but less skilled readers do not (Swanborn & de Glopper, 2002). Less skilled readers are better at intentionally deriving word meaning when the essential informative context is in close proximity to the word to be learned than when the context is separated, while proximity of context does not affect skilled readers (Cain, Oakhill, & Lemmon, 2004). When reading multiple contexts containing a single unknown word, skilled readers learn more intentionally if the contexts are interspersed with other contexts (spaced practice) than if they are read in immediate succession (massed practice); the learning of less skilled readers is not influenced by spacing (Frishkoff et al., 2008). However, it has also been argued that inequalities in both comprehension and word learning ability can be explained by differences in global cognitive functions such as memory. Daneman & Green (1986), proponents of this theory, found that reading span, a measure of working memory, is positively correlated with the ability to deriving meaning intentionally.

Studying Incidental Word Learning

The incongruities in this body of research reinforce the claim made by Nagy and colleagues that incidental word learning must be studied independently from intentional derivation. If, as these researchers argue, incidental word learning is the

greatest source of vocabulary learning, it is important to know the extent to which readers actually benefit from it, as well as how it can be manipulated. Motivated by this question, Nagy et al. (1985) first attempted to calculate the probability of learning words incidentally. In a later study, they proposed a 5% chance of learning a word from a single exposure (Nagy et al., 1987). Noting that other studies produced anomalous effect sizes, Swanborn and de Glopper (1999) conducted a meta-analysis of 15 incidental word learning studies (20 experiments in all). Their criteria for inclusion in the meta-analysis were: learning occurred in the native language (L1); no mention of learning was made and no attention was drawn to vocabulary; texts (authentic or constructed) were not artificially transparent; and target words appeared in only one text (multiple appearances in the same text was permissible).

In comparing these studies, Swanborn and de Glopper (1999) sought to compute an average probability of learning and to explore the effects of the many dimensions on which the studies differed. They report an average probability of .15. Of the numerous word-, context-, and participant-level factors they assessed, only the following significantly predicted learning: pre-test sensitization (more was learned when the unknown targets words appeared on a pre-test of word knowledge); grade level (older children learned more); reading ability (skilled readers learned more); whether the study gave credit for partial word knowledge (if so, more learning was reported), and text-to-target ratio (when the ratio of known words to unknown words was larger, more was learned). Although many commonly manipulated variables were analyzed, a significant portion of the variation in effect size remained unexplained.

One of Swanborn and de Glopper's (1999) significant predictors, credit for partial word knowledge, invokes a contentious issue for both models of word learning and the methodology of word learning studies. As suggested by evidence that word learning and meaning understanding improve with multiple exposures, word learning, whether intentional or incidental, is widely considered to be an incremental process (e.g., Bolger et al., 2008; Durso & Shore, 1991; Frishkoff et al., 2008; Fukkink et al., 2002, Nagy et al., 1987). According to models of incremental learning, word knowledge can be represented by a spectrum ranging from fully unknown to fully known, but the nature of movement along the spectrum is unclear. Often, readers are familiar with or possess some prior knowledge about words whose meanings are unknown or largely unknown. These intermediate levels of word knowledge are collectively termed partial word knowledge. Studies such as those mentioned above have found that readers are more likely to learn the meaning of a partially known word than the meanings of a novel, unknown word. This advantage will be generally referred to as the partial word knowledge effect. Not all studies have found a difference in learning rate between words that are completely novel and words that are partially known (Daneman & Green, 1986; Schwanenflugel et al., 1997). Regardless, support for incremental learning and partial word knowledge is strong, and many incidental word learning studies have not fully taken these factors into account.

Eye-Tracking: A New Perspective on Word Learning

The theorized spectrum of word knowledge has become a topic of interest to researchers beyond those using the traditional intentional and incidental learning paradigms. Through new perspectives and techniques, these researchers have begun to contribute previously unavailable findings to the field. One potentially fertile source of information on word knowledge is through the study of eye movements. Eye movement patterns are studied through the use of eye-tracking devices, which can record people's eye movements as they read. Because eye-tracking allows participants to read naturally, reading processes can be explored with much greater ecological validity than other experimental tasks. Eye movement recording offers a way to examine on-line reading processes, including fixation patterns and time courses (for more information on eye movements and reading processes, see Rayner & Juhasz, 2006). Eye-tracking may be profitable for the word learning paradigms because eye movements can provide information about how reading is affected by the presence of unknown words that cannot be measured by a post-test assessment of learning. Although very few word learning studies to date have used eye-tracking methodology, they have been highly informative for the current research.

Chaffin, Morris, and Seely (2001) used eye-tracking to study the spectrum of word familiarity. Their experiments were based on the widely-replicated finding that readers spend more time looking at low-frequency words than high-frequency words, where frequency is measured by how often a word occurs in print (e.g., Inhoff & Rayner, 1986; Rayner & Duffy, 1986). Chaffin et al. sought to determine whether the same pattern would hold for words of low- and high-familiarity (as measured by

subjective rankings), and extend to completely unfamiliar novel words (represented by orthographically standard pseudowords). They found that low-familiar words received more initial processing and more rereading, or later processing, than high-familiar words; unsurprisingly, so did unfamiliar words (see Juhasz & Rayner, 2003, for a discussion of word frequency and subjective familiarity effects). The more unexpected finding was that there were no initial processing differences between low-familiar words and unfamiliar novel words. Novel words did receive more later processing than low-familiar words, as measured by total reading time and frequency of regressions. Chaffin et al. concluded that when novel words are orthographically standard and pronounceable, they are encoded using the same initial strategy as that used with low-familiar words.

A later study by Williams and Morris (2004) was unable to replicate this finding. They hypothesized that the low-familiar words used by Chaffin et al. (2001) were so unfamiliar that they were essentially novel. Familiarity ratings collected for those words support this explanation. They found more initial and later processing for novel words than low-familiar words, although they likewise conclude that normal processing strategies are used for novel words. Williams and Morris, unlike Chaffin et al., included an additional post-test for knowledge of word meanings. They reasoned that while eye-tracking demonstrates on-line processing, it produces no way of measuring whether or not word meanings were successfully acquired, this assessment being the aim of most word learning research. Their post-test results offered the somewhat inexplicable finding that novel words whose meanings were correctly selected from a two-option multiple choice test had less initial and more

later processing than novel words whose meanings were not correctly selected. They suggest that this reflects efficiency in determining that novel words are unfamiliar.

The Current Study

The current study explores the effects of a previously untested source of variation among word learning studies: context length. The possible influence of context length is especially important to consider if eye-tracking is to continue being used as a methodology for studying word learning. The collection of word learning studies reviewed earlier used contexts of varying lengths, from single sentences to pages-long passages. Swanborn and de Glopper's (1999) meta-analysis examined text-to-target ratio, and found it to be a significant predictor of word learning. However, text-to-target ratio is not necessarily an equivalent measure to context length. First, all of the studies included in the analysis used passage contexts, and thus the ratios (which range from 37 to 150 known words to one unknown word) are not as disparate as those that could exist between sentences and longer, paragraph-based contexts. It is typical in eye-tracking to use individual words or single sentences as stimuli (this is partly a function of goal, and partly a function of technological limitations). This was the case in the two eye-tracking studies described above: Chaffin et al. (2001) used sentence pairs; Williams and Morris (2004) used single sentences. Second, the text-to-target ratios were calculated across all contexts used during the study rather than per individual context. This obscures possible differences between reading a novel word in a 5-sentence long context and reading a novel word in a single sentence context along with four unrelated sentences. The latter condition

often occurs in eye-tracking studies, which include a considerable number of nonexperimental "filler" sentences to mask the purpose of the experiment.

Some findings from previous word learning research support a learning advantage for longer contexts. Assuming that it is not experimentally controlled, overall informativeness is likely to be greater in longer contexts than single sentence contexts: although the text surrounding the novel word may not be any more constrained for meaning, there is potential for a greater number of context clues. While a benefit of context informativeness has not been consistently found for incidental learning, it is not unreasonable to suspect that it has some effect given its strength as a predictor of intentional learning. Additionally, longer contexts may demonstrate the facilitative effects found with larger text-to-target ratios.

It is also possible, however, that shorter contexts may be more conducive to learning. Freebody and Anderson (1983) examined how increasing the overall number of unknown or difficult words affected passage comprehension, and discovered that it took a substantial proportion of difficult words to significantly detract from comprehension. They proposed that, because participants still understood text in which many words were unfamiliar, novel words in a passage might be quickly passed over or even skipped completely. Their finding may reflect the influence of another of Sternberg and Powell's (1983) mediating variables: how important a word is to understanding the context in which it is embedded. It has been theorized that an unknown word of less importance to the overall meaning of the text will present less of a detriment to text comprehension than an unknown word of higher importance (Jenkins & Dixon, 1983; Stahl & Fairbanks, 1986; Sternberg &

Powell, 1983). Although a significant effect of importance for context meaning has not yet been experimentally produced, it cannot be dismissed as a potential variable (Schwanenflugel et al., 1997). When context is limited to a single sentence, each word in the sentence may take on a greater degree of importance for overall meaning, as readers will not have the aid of surrounding context to compensate for an unknown word. This increase in importance may lead to an increase in the attention given to unknown words, and an increase in the amount of effort devoted to deriving their meaning.

The hypothesis that shorter contexts may result in more word learning may be supported by evidence from eye-tracking. Radach, Huestegge, and Reilly (2008), in a study of eye-tracking as a methodology, compared eye movement records for single sentences and six-sentence paragraphs in German. They found that gaze duration (an initial processing measure that gives the sum of all fixations during the first reading of a word) on the target words was longer when the words were in single sentences than when they were in paragraphs. They also found that word frequency effect sizes differed significantly between sentences and paragraphs. The difference in gaze duration between high- and low-frequency words was reduced when the words were in paragraphs. On the other hand, total reading time, a later processing measure, was longer for words in paragraphs than sentences. Radach et al. argue that this pattern reflects different reading strategies based on text length: short texts are read thoroughly on the first pass, while longer texts are read with a quick first pass followed by extensive rereading.

The initial increase in processing for words in sentences may result in improved word learning. Although it is possible that any relative benefit that this increase confers is then negated by the increased rereading of words in paragraphs, I suspect that shorter contexts do ultimately facilitate learning. The increased initial processing for words in sentences is also theoretically consistent with evidence from word learning that supports an advantage for sentences. The current study aims to test this theory by recording participants' eye movements as they read unknown words in sentence and paragraph contexts. Knowledge of the unknown words' meanings (and thus the amount of word learning) will be formally assessed by a vocabulary test administered after the stimuli are read. The eye movement records will provide important information about how reading processes are affected by the presence of unknown words. Because Radach et al. is the only study to date to explore how material length affects eye movement patterns, this study also serves as an attempt to replicate their findings that reading processes differ between sentences and paragraphs.

The basic methodology for this study is modeled after that of the experiments conducted by Williams and Morris (2004). It consists of an eye-tracking reading portion and a post-test portion, following their rationale that a test of word meaning knowledge is necessary to assess whether learning has taken place. It also uses orthographically standard pseudowords to represent novel, unknown words. Nagy et al. (1987) argue that using pseudowords matched to real word meanings underestimates the difficulty of real word learning, which often requires new concept learning as well. However, using pseudowords guarantees zero familiarity, a level of

word knowledge that may have been difficult to obtain with the highly educated adult population tested in this study.

A number of alterations were made to the method to increase the difficulty of the learning task, and to better reflect the conditions of both other word learning studies and actual adult reading situations. The pseudowords used by both Chaffin et al. (2001) and Williams and Morris were matched with targets of the noun class and were usually located in the subject or direct object position of the sentence's independent clause. They were therefore critically important to meaning of the context and very salient. The current study uses pseudowords paired with adjective class targets. Adjectives are generally less central to sentence comprehension. Additionally, it seemed highly probable that adult readers, like those used in this study, would be more likely to encounter unknown adjectives than unknown nouns in their normal reading. One piece of evidence for this was obtained by examining a book containing difficult vocabulary words commonly tested by the GRE® (Graduate Record Examinations) General Test, a standardized exam designed for college students and graduates: the part of speech distribution was 48% adjectives, 29% nouns, and 23% verbs (Kaplan, 2008).

In order to be comparable with traditional word learning studies, the current study complies with Swanborn and de Glopper's (1999) meta-analysis requirements: learning occurred in L1; no mention of learning was made and no attention was drawn to vocabulary; target words appeared in only one text; and texts were not artificially transparent. In addition to using adjectives instead of nouns, the current study uses pre-existing contexts rather than contexts constructed solely for the

purpose of the study. Although the contexts had to be slightly modified, authenticity was preserved to the greatest possible extent. A different design is used for the posttest vocabulary assessment: the multiple-choice test provides five possible meanings for every unknown word rather than two, to better parallel the procedures of incidental word learning studies. These measures were undertaken to help ensure that this examination of the effects of context length on eye movement patterns and incidental word learning would be methodologically sound by the standards of both experimental paradigms.

Method

Participants

Fifty-six students at Wesleyan University participated in the experiment. Participants were recruited using flyers, email notifications, and a posting on the University blog. They were paid \$7 for their participation. All participants were native English speakers, and had normal or corrected-to-normal vision.

Materials

Words. Twenty low-frequency words of the adjective case were selected to compose the real (known) word set. The average frequency of the set was 4.30 words per million, with a range of 0 to 11.17, as collected from the CELEX database (Baayen, Piepenbrock, & van Rijn, 1995). Despite the low frequencies of the words, they were assessed to be words whose meanings would be familiar to the majority of college students. The words ranged from 8 to 10 characters in length, with an average of 9.4 characters.

Mean type bigram frequency was also collected for each word in order to assess orthographic regularity. Bigram frequency refers to the frequency with which two letters appear next to each other in specific positions within a word (e.g., how often c and a occur in the first and second letter position, respectively, as in cat and calm). Type frequency measures this as the number of words that contain that letter pair in that position (another measurement, token frequency, also factors in the word frequency of each word containing the pair). Mean type bigram frequency reports the average across every pair of adjacent letters in a word (e.g., in calm, c-a, a-l, and l-

m). Mean type bigram frequencies for the real word set ranged from 33.43 to 72.25, with an average of 47.08, as measured using N-Watch software (Davis, 2005).

Each real word from the set was paired with an orthographically standard pseudoword. The words pairs were matched exactly for length and approximately for mean type bigram frequency (the average for the pseudowords was 46.83). The pseudowords were created by the experimenter to have identifiably adjectival endings, but to otherwise look orthographically distinct from any real English word (e.g., *iffergent* and *puliguous*). To determine whether the pseudowords were sufficiently distinct, each pseudoword was searched for on the internet using the search engine GoogleTM: in order to be used in the experiment, a pseudoword had to return zero results, including suggested spelling corrections and foreign language web pages. Additionally, the pseudowords contained no obvious Greek or Latin roots. These measures were taken to assure that the pseudowords would be completely unknown to all participants, as these twenty pseudowords composed the set of novel words to be learned.

Contexts. In order for each participant to read all real and novel target words without exposing the nature of the experiment, it was necessary to create two contextual frames for each word pair. Contexts were collected using REAP, a software package designed to retrieve texts available on the internet based on specified constraints (Collins-Thompson & Callan, 2004; reap.cs.cmu.edu/). Each real target was used as a search query, specifying 10th grade as the minimum reading level and 200 as the minimum word length. Texts were selected based on overall quality of the writing and content, ability to be shortened into a cohesive passage,

informativeness for the meaning of the target word, and topic (in order to select two different topics for each word, but also to achieve a variety of topics across all targets).

Once two texts were selected for each target, the sentence within each text containing the target was edited as necessary. In order to fit on a single line on the display, sentences were edited to approximately 80 characters in length (range of 77-80, average 78.78). The sentences were altered so that the target occurred in the middle of the sentence (at least two words from the beginning and end of the sentence). The sentences were also edited to assure that they were still comprehensible and informative about the meaning of the target word when isolated from the larger paragraph context. This set of isolated sentences comprised the stimuli for the sentence portion of the study.

To create the stimuli for the paragraph portion, the edited sentences from the sentence portion were replaced in their original contexts. The longer texts were then altered to fit the sentences, and to conform to the constraints of the experiment. Each text was shortened to fill nine lines on the display screen, an average of 128 words per paragraph. Changes in content were made as necessary to preserve the coherence of the shortened passages. Misspellings and grammatical errors were corrected, and proper names were sometimes changed. Minor rephrasing and restructuring were sometimes necessary to assure that the target word appeared in the center of the paragraph, at least three lines from the top and bottom of the paragraph, and at least three words from the beginning and end of the line. These restrictions were applied to reduce data loss due to the limits of the eye-tracker (which records less accurately at

the extreme edges of the display) and the inaccuracy of saccades during return sweeps. These paragraphs comprised the stimuli for the paragraph portion of the study. Table 1 gives examples of both context form conditions (a complete list of the sentence stimuli is provided in the Appendix).

Ratings. Eighteen participants (none of whom took part in any other portion of the study) rated the sentence contexts on how well they fit the real word target on a 7-point scale (with higher numbers indicating a better fit). Sentences were split into two lists, so that each participant only saw one sentence for each word. All sentences exceeded an average rating of 5, and the average rating across all sentences was 6.10.

The sentences were also normed for predictability. Eighteen participants (none of whom took part in any other portion of the study) were given the beginning of the sentence contexts, up to where the target word occurred. For each sentence, they were instructed to provide a word that could fit as the next word in the sentence. Sentences were split into two lists, so that each participant only saw one sentence for each word. The target word was correctly predicted only 6% of the time (one sentence elicited two correct answers, the remaining had one or less).

Informativeness and constraint for meaning of the target were permitted to vary in order to maintain textual authenticity. Nineteen participants (none of whom took part in any other portion of the study) were given the entire sentence contexts in a cloze test, with the target word deleted and replaced with a blank. They were instructed to fill in the blank with a word that could reasonably complete the sentence. Sentences were again split into two lists, so that each participant only saw one sentence for each word. Because of the low frequencies of the real targets,

synonyms for the targets were also scored as correct responses. Participants correctly completed the sentence 52% of the time. The most informative/constraining sentences elicited correct responses from 89% of the participants (3 sentences); the least informative/constraining sentence was correctly completed by 0% of the participants. It should be noted that despite their low frequencies, 80% of the target words appeared among the participant responses, indicating that their meanings are familiar to the population, as estimated by the experimenter.

The sentences also differed in where the bulk of the informative context was located in relation to the target word. The informative context occurred primarily before the target in half of the sentences, and after it in the other half.

Post-Task Assessment. A pen and paper multiple-choice vocabulary test was created to assess whether the meanings of the novel words were learned during the eye-tracking portion of the study. For the vocabulary test, five choices were provided for each novel word. The one correct choice was a higher frequency (average 21.89 words per million) synonym for the novel word's corresponding real word (e.g., perfect for the real target impeccable). The remaining four choices were randomly selected words, also of higher average frequency than those used in the real word set (average 23.53 words per million). The incorrect choices were distinct in meaning from the target and from the ideas expressed in that target's contexts. All of the word choices were adjectives.

Design

Each participant read all 20 real words and all 20 novel words during the eye-tracking portion of the study. Likewise, all 40 sentence contexts were read, 20 in sentence form, and 20 in paragraph form. For each participant, the words from each real/novel word pair appeared in different contexts (e.g., the real word with context 1, and the novel word with context 2), and in different contextual forms (e.g., the real word in sentence form, and the novel word in paragraph form). The assignment of words to contexts and contexts to forms was counterbalanced such that across all participants, each word was read in both contexts in both forms.

Apparatus

Eye movements were recorded by an EyeLink 1000 video-based eye-tracking device from SR Research Ltd. Participants were seated approximately 83 cm away from a ViewSonic CRT monitor where the sentences and paragraphs were displayed. Stimuli were displayed in 14 pt. Courier New font. At this distance, approximately 3.62 characters equaled 1 degree of visual angle. Stimuli were displayed using the EyeTrack software package (http://www.psych.umass.edu/eyelab/software/). Eye positions were sampled every 1 ms, and a chin rest and forehead rest were used to reduce head movements. Although viewing was binocular, eye movements were recorded from only the right eye.

Procedure

Participants were run one at a time, and the experiment lasted approximately 45 minutes. During the eye-tracking session, participants sat facing the computer monitor using a chin rest and forehead rest to reduce head movement. At the beginning of each portion of the experiment, the eye-tracker was calibrated. During the calibration, participants were instructed to fixate circles that appeared on the screen: a 3-point calibration (on a single horizontal line) was conducted for the sentence portion, and a 9-point calibration (over the entire screen) was conducted for the paragraph portion. If the calibration was deemed inaccurate, a new calibration routine was conducted. Calibration accuracy was monitored during the course of the experiment by the experimenter, and a new calibration was performed whenever accuracy was deemed inadequate. During the paragraph portion, calibrations were performed approximately every five paragraphs.

Prior to each sentence or passage, participants had a one-point calibration check. If the calibration was deemed accurate, a sentence or paragraph appeared on the computer screen. Participants were given verbal instructions at the start of the experiment to silently read each sentence or paragraph for comprehension. When finished, participants pressed a button on a control pad, causing the sentence or passage to disappear and be replaced by another one-point calibration check. Yes-or-no comprehension questions were presented visually after 25% of the sentences and 100% of the paragraphs. Participants responded by pressing one of two buttons on the control pad. They did not receive feedback on their accuracy. The average response accuracy was 89% for the sentences and 86% for the paragraphs. The order of the

sentence and paragraph portions was counterbalanced, and after one portion was completed, participants were given a short break before beginning the next portion.

After participants completed both parts of the eye-tracking session, the post-test assessment was administered (participants were not warned about the post-test prior to the eye-tracking session). They were instructed to select the best synonym for each novel word from five possible choices. Participants were then asked to rate their level of confidence for each answer given on a 0-2 scale (0 = not at all confident, 1 = somewhat confident, 2 = confident).

Results

Dependent Measures

Six dependent eye-tracking measures were recorded and analyzed: first fixation duration, gaze duration, go-past time, total time, regressions out, and regressions in. First fixation duration is the length of the initial fixation on a word during its first-pass, irrespective of the number of fixations the word receives. Gaze duration is sum of all fixations during the initial reading of a word, before the reader leaves the word. Go-past time is the sum of all fixations during the initial reading, including any regressions to earlier material and any rereading fixations before the eyes move to the right of the word. Total time is the cumulative amount of time spent reading a word over the course of a trial. Regressions out is the percentage of cases when the reader left a word to return to an earlier word. Regressions in is the percentage of time the reader returned to a word after having left it. First fixation duration, gaze duration, go-past time, and total time are measured in milliseconds; regressions out and regressions are recorded as percentages (0% or 100%). In addition, two dependent measures from the post-test assessment were analyzed: accuracy (correct or incorrect) and confidence level (0-2).

Data Loss

Data were discarded from six participants who were unable to complete the experiment due to persistent calibration failures. Trials were excluded from analysis if a track loss occurred on the pre-target, target, or post-target word or if inaccurate tracking on the vertical axis rendered the data indecipherable (for paragraph trials

only). Participants from whom 30% of paragraph trials or 20% of sentence trials were excluded were entirely removed from analysis. From the remaining 32 participants, 9.7% of the data were lost due to track losses or indecipherability. In addition, outliers (times 2.5 standard deviations above or below the mean) were trimmed from the duration measures. This resulted in the removal of 2.9% of the first fixation data, 2.3% of the gaze duration data, 2.2% of the go-past time data, and 3% of the total time data.

Design

The eye movement data were analyzed using a 2 (word type: real or novel) \times 2 (context type: paragraph or sentence) analysis of variance (ANOVA). Error variance was computed over participants (FI) and over items (F2). Average accuracy and average confidence level on the post-test assessment were analyzed using t-tests.

Eye movement Analyses

Mean viewing times and percentages for each of the four cells in the design are reported in Table 2. First fixation duration, gaze duration, and total time were all analyzed for the target (real or novel) word only. First fixation durations were significantly longer for novel words than for real words, FI(1, 31) = 9.66, p < .005, F2(1, 39) = 9.62, p < .005. The main effect of context type and the interaction between word type and context type were not significant (all ps > .1).

Likewise, there was a significant main effect of word type for gaze duration on target, FI(1, 31) = 57.60, p < .001, F2(1, 39) = 199.99, p < .001, with novel

words having longer gaze durations than real words. Gaze duration also showed a significant effect of context type, FI(1, 31) = 6.54, p < .05, F2(1, 39) = 5.56, p < .05. Gaze durations for both word types were longer in isolated sentences than in paragraphs. The interaction between word type and context type was not significant (both ps > .1).

Novel words had significantly longer total times than real words, F1 (1, 31) = 102.25, p < .001, F2 (1, 39) = 286.02, p < .001. Total time was also significantly longer for sentences than for paragraphs, F1 (1, 31) = 5.28, p < .05, F2 (1, 39) = 5.21, p < .05. The interaction was not significant (both ps > .1).

Regressions out and regressions in were analyzed for both the target (real or novel) word and the post-target word (the word immediately following the target) in order to assess late or spillover processing effects. For regressions out of the target, there were no significant main effects for word type or context type, and no significant interaction (all ps > .1). Regressions out of the post-target, however, showed a significant effect of word type, FI(1, 31) = 25.32, p < .001, F2(1, 29) = 32.22, p < .001. When the target was a novel word, readers were more likely to regress out of the post-target. The interaction between word type and context was significant by participants, FI(1, 31) = 11.42, p < .005, but this effect was not significant by items, F2(1, 29) = 2.51, p = .124. Post hoc t-tests showed that when the target was a real word, regressions out of the post-target did not differ significantly between sentences and paragraphs (p > .1). When the target was novel, however, regressions out of the post-target were more likely in paragraphs than

sentences, t(31) = 2.64, p < .05. The main effect of context was not significant in either analysis (both ps > .1).

For regressions into the target, there was a significant effect of word type, FI (1, 31) = 43.05, p < .001, F2 (1, 39) = 65.10, p < .001. Regressions into the target were more likely when the target was novel. The main effect of context and the interaction between word type and context type were not significant (all ps > .1). Regressions into the post-target showed a significant main effect of context by items, F2 (1, 34) = 5.32, p < .05, but the effect was only marginal by participants, FI (1, 31) = 3.03, p = .092. Regressions into the post-target word were more likely in sentences than paragraphs. The main effect of word type and the interaction were not significant in either analysis (all ps > .1).

Go-past times were analyzed for the target, but also for the region between the post-target word and the end of the target sentence. For go-past time on the target, there was a significant main effect of word type, FI (1, 31) = 50.27, p < .001, F2 (1, 39) = 125.49, p < .001, with longer go-past times for novel words. The main effect of context and the interaction were not significant (all ps > .1). Go-past times on the end of the sentence were longer when the target was novel, FI (1, 31) = 40.96, p < .001, F2 (1, 39) = 13.66, p < .001. There was also a significant effect of context, FI (1, 31) = 27.37, p < .001, F2 (1, 39) = 54.05, p < .001. Go-past times on the end of the sentence were longer when the sentences were in isolation than when they were embedded in paragraphs. The interaction between word type and context type was not significant (both ps > .1).

Accuracy and Confidence Level Analyses

Means for accuracy and confidence level are reported in Table 3. Overall, accuracy on the post-test did not differ from chance (chance was 20%; p > .1). The accuracy on novel words read in sentences was not significantly different from the accuracy on novel words read in paragraphs, and neither was different from chance (all ps > .1).

Although accuracy was not different than chance, the average confidence level was greater than zero, t(31) = 5.95, p < .001. In addition, participants reported significantly higher confidence in their correct responses than their incorrect responses, t(31) = 2.21, p < .05. Confidence levels for words read in sentences versus words read in paragraphs did not differ significantly, p > .1.

Discussion

Eye-tracking, typically used to study aspects of reading such as word recognition and syntax processing, is a relatively new method in the field of word learning research. Previous attempts to use eye-tracking to explore incidental word learning have not been entirely comparable to more traditional word learning studies. The current research, however, combines the essential features of both paradigms (e.g., control of confounding variables from eye-tracking, textual authenticity from word learning), and establishes the potential of this methodology for future research. The results of this experiment suggest that context length affects eye movement patterns. As a consequence, the rate of word learning may differ depending on whether the context surrounding unknown words is short or long. Though there is evidence that word learning occurred, the rate was not substantial enough to be measured by the post-test assessment.

Context Length Affects Eye Movements

It is clear from the present research that readers were sensitive to the occurrence of unknown words in text, regardless of context length. While Williams and Morris (2004) reported this finding for unknown words in sentences, the current study demonstrates that it also applies to unknown words in paragraph contexts. Freebody and Anderson (1983) proposed that because it is not necessary to understand every word in a passage to have adequate reading comprehension, unknown words in paragraph contexts may be skipped over or only briefly attended to. This does not appear to be the case. Readers in this experiment showed increased

reading times for novel words, indicating that they recognized that the words were unfamiliar or unknown and, rather than skipping over them, devoted more attention to processing them. Furthermore, readers were more likely to regress back in the text after reading unknown words, suggesting that they were attempting to derive their meanings from the available context on-line.

Although increasing the length of the context did not result in the neglect of unknown words, other differences between sentence and paragraph reading were exhibited. Readers had longer gaze durations on the target words, both real and novel, when they were embedded in isolated sentences than when they were in paragraphs. Readers also had longer total times on targets in sentences. No interaction was observed between word type and context length, but this does not invalidate the hypothesis that shorter contexts facilitate word learning. Although context length may affect the reading of real and novel words to the same degree, more reading time is still devoted to novel words in sentences than to novel words in paragraphs. Subsequently, if increased reading time improves word learning, learning will occur at a greater rate when unknown words are in single sentence contexts.

These results are partially contradictory with previous research on sentence and paragraph reading. Radach et al. (2008) conducted the only other examination of the effects of context length on eye movement patterns. Consistent with the current study, they found that initial processing, as measured by gaze duration, is increased when words are in sentence contexts. However, they found a reverse effect of context on later processing. I did not replicate their finding that total time was longer for words in paragraph contexts.

There are a number of methodological differences between the two studies, and it is conceivable that these contributed to the divergence of their findings. Radach et al.'s (2008) study was conducted in German; the current study was in English. In Radach et al., each constituent sentence within a paragraph was presented on its own line, rather than immediately following another the way a paragraph would typically appear, as in this experiment. These studies also used targets of different parts of speech. The Radach et al. study used exclusively nouns for its target words. As previously discussed, nouns tend to have greater importance for the overall meaning of the context, and therefore may have received more rereading time than the adjectives used in the current study.

It seems unlikely that any of these factors would have resulted in the reversal of effect between Radach et al. (2008) and the current study. It is more feasible that the results deviated because the target words in these two studies differed considerably in word frequency. Radach et al. used equal numbers of high-, medium-, and low-frequency words. They found an important interaction between word frequency and context format: the size of the frequency effect (the difference in reading times between low-frequency and high-frequency words, with low-frequency words having longer times) was greater when the words were in sentences than in paragraphs. This interaction was demonstrated for both gaze duration and total time. Radach et al.'s main finding for total time may have been driven by an increase in rereading for high- and medium-frequency words in paragraphs. The current study was not subject to this potential skew because it only used low-frequency words (as novel words represent extreme low-frequency).

I am unable to confirm that this pattern exists, however, as Radach et al. (2008) do not report separate total time averages for their three categories of word frequency, and thus this explanation is purely speculative. Although they conflict with Radach et al., there is no reason to suspect that the results of the current study are in error. Words embedded in sentence contexts seem to receive more initial and later processing than words embedded in paragraph contexts. Further investigation is required to resolve the discrepancy between these studies.

Rate of Incidental Word Learning

Readers' eye movement patterns appeared to reflect attempts to derive meaning from context, but performance on the post-test vocabulary assessment failed to demonstrate any significant learning. This is not particularly surprising. Williams and Morris (2004), whose methodology best resembles that of the present research, found a learning rate of only 5%, which is consistent with Nagy et al.'s (1987) original estimate but smaller than Swanborn and de Glopper's (1999) calculated rate of 15%. If learning in the present study had occurred at a rate of 5%, one word would have been learned, while an average of four words would have been correctly defined simply due to chance — this effect may have been too small for my analyses to detect. However, I suspect that the rate of learning was in fact less than 5%, given the measures taken to make learning conditions in the present study more difficult (via increased authenticity) than the conditions in Williams and Morris. The targets in the present study were adjectives, which are likely to be less concrete, less conceptually simple, and less important to overall context meaning than nouns. The contexts,

modified from pre-existing text, were less transparently informative. The post-test assessment was more demanding, with five possible choices for each word instead of two. Finally, the context length manipulation cannot be dismissed as a possible contributor to the challenge of the task. Because overall learning was too slight to be measured by the post-test, any effect of context length may likewise have been too subtle to observe.

The difficulty of observing learning through the post-test was anticipated. Multiple-choice vocabulary tests are frequently used to assess incidental word learning because they are thought to require only a moderate amount of word knowledge: single exposure to an unknown word in context is sometimes enough to be able to respond correctly (e.g., Curtis, 1987; Nagy et al., 1987). However, this degree of learning from a single exposure is not a ubiquitous finding. According to theories that propose a spectrum of word knowledge, readers can possess partial knowledge about a word that is insufficient for accurate performance on a vocabulary test. The current study asked participants to provide confidence ratings for their answers on the post-test as a means of exploring partial word knowledge. The confidence ratings revealed that participants felt some degree of confidence in their answers. Although it was not enough to benefit post-test performance, some familiarity was acquired from reading the unknown words in context. More importantly, participants were more confident about words which they correctly defined than those they incorrectly defined. This is notable because a general feeling of confidence could have arisen without any real word knowledge. For instance, if a participant confused a novel word with a different but real word, they would feel

confident but their knowledge (and most likely, their response) would be incorrect.

The relationship between higher confidence levels and correct responses, however, indicates that participants were cognizant of possessing word knowledge and that their knowledge was both accurate and sufficient for selecting the correct definition.

Reader confidence levels suggest that some word learning did occur, but the question remains of why the rate of incidental learning did not approach the 15% or even 5% level reported in prior research, and was in fact too insignificant to be quantified. Word learning is assumed to be a demanding task, and this study was intentionally designed to be more challenging than some previous studies. On the other hand, certain characteristics of the study should have inflated the rate of learning. Grade level and reading ability were two of the factors that Swanborn and de Glopper (1999) found to predict incidental word learning. The readers in the current study were all highly-educated adults, and presumably had the benefit of both higher grade level and greater reading ability than the students tested by the studies included in their meta-analysis (whose grade levels ranged from 4 to 11). Readers in the current study also had the advantage of conceptual simplicity over learners in natural reading situations. The words to be learned were all pseudowords assigned the definitions of real words whose meanings were known to the participants. Though conceptual complexity was increased by using adjectives instead of nouns, learning their meanings did not require learning brand new concepts, as is often the case in real incidental word learning (Nagy et al., 1987).

In spite of this gain in conceptual simplicity, I propose that the use of pseudowords contributed considerably to the below-average learning of the readers in

this study. Pseudowords were used because of the difficulty of finding a sufficient number of unknown words with a highly-educated adult population: by using fabricated words with experimentally assigned meanings, zero prior knowledge was ensured. Although studies that use real words assume zero prior knowledge for their unknown targets, this may be on inappropriate assumption. Studies using real words may receive an unacknowledged benefit of partial word knowledge, one that was not present in the current study.

The Absence of Partial Word Knowledge

The vast majority of word learning studies (including all of the studies in the Swanborn and de Glopper (1999) meta-analysis) use real words as their unknown targets. The procedures they use to distinguish known words from unknown words vary in precision. Occasionally, unknown words are selected based on the judgments of experimenters or teachers, but this is an inexact method and does not measure actual reader knowledge. Instead, most studies assess reader-specific knowledge with vocabulary pre-tests. Pre-testing measures reader knowledge with much greater rigor and validity, but it is also a problematic method.

Swanborn and de Glopper (1999) found that the learning of unknown targets improves when those targets are included in a pre-test of word knowledge, an effect they call pre-test sensitization. It is powerful evidence in support of incremental learning and partial word knowledge theories that readers benefit from prior exposure to unknown words. Pre-test sensitization occurs when words are presented independent of any definition or context, and even when there is a substantial delay

between the administration of the pre-test and the learning task. Any study that uses pre-testing is likely to see this artificial increase in the rate of learning; the current study avoided needing to pre-test by using pseudowords and thus avoided the contribution of pre-testing to partial word knowledge.

Although the process of taking a vocabulary pre-test itself builds partial word knowledge, most vocabulary test designs (e.g., multiple-choice, definition generation, marking words as known or unknown) can only distinguish between extremes in knowledge and fail to account for any kind of partial word knowledge. In scoring a typical multiple-choice test, for instance, correctly defined words are considered known and incorrectly defined words are considered unknown. Shore and Durso (1990) sought to design a pre-test that could discriminate an intermediary level of knowledge. In addition to known and unknown words, their pre-test includes the classification of *frontier words*. A frontier word is only partially known: the reader acknowledges familiarity with the word, but otherwise cannot define it or use it appropriately in context. In Shore and Durso's pre-test, participants are given a list of words (including some pseudowords as foils) and follow ordered instructions: first, participants provide definitions or synonyms; second, they use words in sentences; third, they mark words that look familiar to them; fourth, they mark words as real or unreal. Words that are correctly defined or used in a sentence are considered known. Words marked as familiar are considered frontier. Real words marked as unreal words are considered unknown.

There are, however, levels of knowledge that even Shore and Durso's (1990) pre-test is not sensitive enough to measure. In a series of experiments using words

from the Shore and Durso pre-test, Durso and Shore (1991) found that even when participants could not identify that a word existed in the English language (the criterion for being unknown), they could accurately perform tasks which required some knowledge of meaning. They could distinguish between correct and incorrect use of the unknown word in a sentence as long as the uses were sufficiently distinct (e.g., for the target happy: The happy child played in the park vs. She put the vase on the happy table); between a word semantically related to the unknown word and a semantically unrelated word (e.g., glad vs. purple); and between a word that could sensibly appear with the unknown word and a word that could not (e.g., happy child vs. happy table). On some tasks, participants performed as well with unknown words as they did with the familiar frontier words. Durso and Shore also tested whether knowledge about unknown words was based on the words' orthography/phonology or on concepts that could not have been derived from the words' physical characteristics: performance levels were equally high regardless of whether word knowledge was physically or conceptually based. These remarkable findings suggest that readers possess knowledge about words that to all appearances are unknown to them, and that this knowledge can influence task performance even when readers are unaware of it.

The subtle degree of word knowledge revealed in Durso and Shore's (1991) experiments is not trivial. If it is enough knowledge to successfully complete the tasks described above, it is probably enough knowledge to facilitate incidental word learning. Because of pre-testing and the potential impact of even minimal levels of word knowledge, a state of zero knowledge cannot be assumed in word learning

studies that use real words as their unknown targets. In fact, it is safer to assume that readers will possess some partial word knowledge. I do not wish to suggest that this fact negates any findings from the corpus of word learning research, or that the current methodologies necessarily need revision. Although using pseudowords may be the only practical means of guaranteeing zero prior partial knowledge, there are advantages to using real words over pseudowords. Instead, I would argue that word learning studies need to recognize the influence of partial word knowledge, and that this is especially important if studies using real words and pseudowords are to be compared.

Limitations and Future Directions

The essential limitation of the current study is that the rate of learning was too low to determine whether incidental word learning was affected by context length.

The eye movement patterns demonstrated that words receive more processing when they occur in single sentence context than in paragraphs contexts. This difference may result in differences in learning. Without corroborating evidence from the posttest assessment, however, this relationship cannot be concluded.

Aspects of the experimental design, many of which have been previously discussed, undoubtedly contributed to the inhibition of learning. The unknown targets were adjectives, and thus tended to be abstract, fairly conceptually abstruse, and relatively unimportant to the overall meaning of the context. The informativeness of the contexts, which were modified from pre-existing texts, was not manipulated, and thus the contexts were neither artificially nor uniformly informative. Some were

highly constrained for the meaning of the targets; others not at all. This was done to maintain authenticity, as natural texts often have nondirective or even misdirective contexts that make it difficult or impossible to derive meaning from context (Beck et al., 1979). However, the decision not to control for informativeness most likely increased both the difficulty of the task and the amount of variation in the results. It would have been challenging to achieve high levels of contextual constraint with adjective targets; if future studies increase informativeness in order to facilitate learning, it may be necessary to use a different part of speech or, preferably, an equal distribution of parts of speech.

The most fruitful direction for future research on incidental word learning is the investigation of partial word knowledge. If the post-test assessment in the current study was more sensitive to minimal levels of knowledge, it is very possible that learning would have been measurable. Partial word knowledge was not entirely neglected: the confidence ratings were collected as a supplement to the multiple-choice questions. The ratings indicated that correct answers were not purely guesswork, and thus implied a degree of learning that was unobservable from the multiple-choice test. Tasks specifically designed to evaluate minimal levels of word knowledge, such as those used by Durso and Shore (1991), would be even more effective at illuminating partial word learning. In a study such as this one, where readers receive single exposures to words for which zero prior knowledge is possessed, the knowledge acquired is probably too slight to be detected by more traditional tests of word knowledge.

Conclusion

The field of eye movement research has generally assumed that findings for short, sentence-long texts are extendable to longer texts. However, this study confirms that eye movement patterns, and the processes controlling them, are affected by context length. Reading times are longer for words in sentences than words in paragraphs. This disparity is important for researchers to consider. The results from experiments conducted using sentences as stimuli cannot necessarily be extrapolated to paragraph stimuli. The current study did not find an interaction between context length and word type. However, other variables and processes (including incidental word learning) may be differentially affected by context length. Similarly, researchers must consider the extent to which studies using sentences relate to real reading circumstances, where texts are more likely to be paragraphs than sentences.

It has also been assumed that pseudowords are no more difficult (if not less difficult) to learn than unknown real words, an. In fact, pseudowords may be more difficult to learn due to a complete absence of partial word knowledge. The results of this study suggest that word learning is incremental, and that possessing partial knowledge (even in very small quantities) of a word's meaning promotes learning. The existence of partial word knowledge is apparent and advantageous, and is also quite pervasive. Any word learning study that uses real words may see inflated rates of learning due to the benefits of partial word knowledge. Although it interferes with the ability of experiments to measure word learning with real targets, partial word knowledge may be very valuable in practice.

Partial word knowledge facilitates word learning, whether that learning occurs through direct instruction or incidentally during the course of natural reading. Partial word knowledge can be developed through increased exposure to text, especially text that includes low-frequency words. Because exposure may be beneficial even if it is only superficial (i.e., purely orthographic), educators and learners can easily capitalize on this effect. Word learning is a complex and effortful task. By creating more opportunities for gaining partial word knowledge, the vocabulary sizes of children and adults could be substantially increased.

References

- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). <u>The CELEX Lexical</u>

 <u>Database. Release 2</u> (CD-ROM), Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.
- Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology*, 98(1), 44-62.
- Beck, I., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal*, 107(3), 251-271.
- Beck, I. L., McKeown, M. G., McCaslin, E. S., & Burkes, A. M. (1979). *Instructional dimensions that may affect reading comprehension: Examples from two commercial reading programs*. (DHEW Publication No. LRDC-1979/20).

 Pittsburgh, PA: Learning Research and Development Center. (ERIC Document Reproduction Service No. ED197322)
- Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74(4), 506-521.
- Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes*, 45, 122-159.
- Cain, K., Oakhill, K., & Lemmon, K. (2004). Individual differences in the inference

- of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, 96(4), 671-681.
- Carnine, D., Kameenui, E. J., & Coyle, G. (1984). Utilization of contextual information in determining the meaning of unfamiliar words. *Reading Research Quarterly*, 19(2), 188-204.
- 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(1), 225-235.
- Collins-Thompson, K., & Callan, J. (2004). Information retrieval for language tutoring: An overview of the REAP project. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 544-545). Sheffield, UK: Association for Computing Machinery.
- Curtis, M. E. (1987). Vocabulary testing and vocabulary instruction. In M. G.

 McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19-35). Hillsdale, NJ: Erlbaum.
- Cunninghamn, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator*, 22(1-2), 8-15.
- Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, 25(1), 1-18.
- Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. Behavior Research Methods, *37*, 65-70.

- Durkin, D. (1979). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, *14*(4), 481-533.
- Durso, F. T., & Shore, W. J. (1991). Partial knowledge of word meanings. *Journal of Experimental Psychology: General*, 120(2), 190-202.
- Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading**Research Quarterly, 24(2), 174-187.
- Freedbody, P., & Anderson, R. C. (1983). Effects on text comprehension of differing proportions and locations of difficult vocabulary. *Journal of Reading Behavior*, *15*(3), 19-39.
- Frishkoff, G. A., Collins-Thompson, K., Perfetti, C. A., & Callan, J. (2008).

 Measuring incremental changes in word knowledge: Experimental validation and implications for learning and assessment. *Behavior Research Methods*, 40(4), 907-925.
- Frishkoff, G. A., Perfetti, C. A., & Westbury, C. (2009). ERP measures of partial semantic knowledge: Left temporal indices of skill differences and lexical quality. *Biological Psychology*, 80, 130-147.
- Fukkink, R.G. (2002). Effects of instruction on deriving word meaning from context and incidental word learning. *L1 Educational Studies in Language and Literature*, 2, 37-57.
- Fukkink, R. G., Blok, H., & de Glopper, K. (2001). Deriving word meaning from written context: A multicomponential skill. *Language Learning*, *51*(3), 477-496.
- Fukkink, R. G., & de Glopper, K. (1998). Effects of instruction in deriving word

- meaning from context: A meta-analysis. *Review of Educational Research*, 68(4), 450-469.
- Hayes, D. P. & Ahrens, M. G. (1988). Vocabulary simplification for children: A special case of 'motherese'? *Journal of Child Language*, *15*, 395-410.
- Herman, P. A., Anderson, R. C., Pearson, P. D., & Nagy, W. E. (1987). Incidental acquisition of word meaning from exposition with varied text features.

 *Reading Research Quarterly, 22(3), 263-284.
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. Perception & Psychophysics, 40, 431-439.
- Jenkins, J. R., & Dixon, R. (1983). Vocabulary learning. *Contemporary Educational Psychology*, 8, 237-260.
- Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, 21(4), 767-787.
- Juel, C., Biancarosa, G., Coker, D., & Deffes, R. (2003). Walking with Rosie: A cautionary tale of early reading instruction. *Educational Leadership*, 60(7), 12-18.
- Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixations during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1312-1318.
- Kaplan. (2008). GRE® exam vocabulary prep. New York, NY: Kaplan Publishing.
- Konopak, B. C., Sheard, C., Longman, D., Lyman, B., Slaton, E., Atkinson, R., &

- Thames, D. (1987). Incidental versus intentional word learning from context. *Reading Psychology*, 8(1), 7-21.
- Kranzer, K. G. (1988). A study of the effects of instruction on incidental word learning and on the ability to derive word meanings from context (Doctoral dissertation, University of Delaware, 1988). *Dissertation Abstracts International*, 50(4-A), 910.
- McKeown, M. G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly*, 20(4), 482-496.
- McKeown, M. G., Beck, I. L., Omanson, R. C., & Perfetti, C. A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Reading Behavior*, 15(1), 3-18.
- Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, *16*(3), 304-330.
- Nagy, W. E., & Herman, P. A. (1984). *Limitations of vocabulary instruction*. (NIE Rep. No. 326). Urbana, IL: Center for the Study of Reading. (ERIC Document Reproduction Service No. ED248498)
- Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge:

 Implications for acquisitions and instruction. In M. G. McKeown & M. E.

 Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19-35). Hillsdale, NJ:

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

- Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *International Reading Association*, 20(2), 233-253.
- Piccin, T. B., & Waxman, S. R. (2007). Why nouns trump verbs in word learning: new evidence from children and adults in the Human Simulation Paradigm.

 Language Learning and Development, 3(4), 295-323.
- Radach, R., Huestegge, L., Reilley, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research*, 72, 675-688.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*, 191–201.
- Rayner, K., & Juhasz, B.J. (2006). Reading processes in adults. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (2nd ed., Vol. 10, pp. 373-378).

 Oxford: Elsevier.
- Schwanenflugel, P. J., Stahl, S. A., & McFalls, E. L. (1997). Partial word knowledge and vocabulary growth during reading comprehension. *Journal of Literacy Research*, 29, 531-553.
- Scott, J. A., Jamieson-Noel, D., & Asselin, M. (2003). Vocabulary instruction throughout the day in twenty-three Canadian upper-elementary classrooms. *The Elementary School Journal*, 103(3), 269-286.
- Shore, W. J., & Durso, F. T. (1990). Partial knowledge in vocabulary acquisition:

 General constraints and specific detail. *Journal of Educational Psychology*,
 82(2), 315-318.
- Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A

- model-based meta-analysis. Review of Educational Research, 56(1), 71-110.
- Sternberg, R. J., & Powell, J. S. (1983). Comprehending verbal comprehension.

 American Psychologist, 38, 878-893.
- Swanborn, M. S. L., & de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Education Research*, 69(3), 261-285.
- Swanborn, M. S. L., & de Glopper, K. (2002). Impact of reading purpose on incidental word learning from context. *Language Learning*, 52(1), 95-117.
- Williams, R. S., & Morris, R. K. (2004). Eye movements, words familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, *16*(1/2), 312-339.

Appendix

Stimuli

The real words and pseudowords used as targets during the eye-tracking portion of the experiment are presented with their corresponding sentence contexts. Participants viewed both contexts for each target pair: one containing the real target and one containing the novel target. The sentences below are displayed in real word form, with the real word targets italicized (italics were not used during the experiment). In novel word form, the novel word was inserted in place of the real word. Additionally, one context for each pair was viewed in the form below (as a single sentence) and one was viewed within a longer paragraph context. The paragraph contexts are not included.

eclectic/elusibal

The style was *eclectic*, influenced by colonial and neo-classical architecture. They will perform an *eclectic* set of dance music, spanning four decades of hits.

vigilant/beragual

Engineers advise building owners to be more *vigilant* when inspecting for cracks. However, countries were warned to stay *vigilant* for recurrence of the illness.

lethargic/dramponic

By the end of the week, the previously *lethargic* students had become energetic. Your energy level may dip and you may feel *lethargic* or depressed in the winter.

whimsical/aporostic

The fanciful paintings portray the *whimsical* characteristics of everyday images. Anime stories range from *whimsical* to grim, providing plotlines for every age.

plausible/puliguous

The court agreed that his story was not *plausible*, as it contradicted the facts. There are some *plausible* scenarios, but no one knows for certain what happened.

frivolous/feluteral

Once considered integral, it became a *frivolous* program with no clear purpose. People would omit some *frivolous* trips, carpool more, and use public transport.

vivacious/teffelant

Sources say the star's *vivacious* personality will add vibrancy to the car ads. Thankfully, her once sickly daughter is now a *vivacious*, healthy five-year-old.

malleable/greminous

Children's minds are *malleable* and can be shaped by nearly any strong influence. Gold is very *malleable* and can be easily shaped for jewelry or industrial use.

notorious/iffergent

The auto-rickshaw drivers, *notorious* for bad behavior, are rarely disciplined. Leaf blowers have become *notorious* for noisily stirring up dangerous allergens.

reputable/letervate

Only use internet resources that are as reliable as a *reputable* print resource. It is safest to use a *reputable* travel firm that provides an experienced guide.

beneficial/viermanile

Legalization supporters are passionate about the *beneficial* uses of marijuana. The antioxidants found in chocolate may have *beneficial* cardiovascular effects.

legitimate/argescient

Spam emails can harm the *legitimate* fundraising organizations they impersonate. The counterfeit products look similar to the *legitimate* EPA-registered products.

malevolent/trempliant

They think they are being haunted by a *malevolent* force and try to videotape it. We are warned about *malevolent* viruses embedded within innocent-looking emails.

lackluster/pelluracal

If your hardwood floors look *lackluster*, have them professionally refurbished. Give your students unique assignments instead of *lackluster*, routine projects.

ceremonial/ombrigible

The Yup'ik masks were used for *ceremonial* purposes, to view the unseen world. Taking the birds for *ceremonial* purposes is imbued with religious significance.

meticulous/faripident

A careful worker, Mendel was *meticulous* in recording and quantifying results. The third phase involves the *meticulous* restoration of this historic gallery.

accessible/emprovious

The material kept in archives must be *accessible* to everyone and well preserved. They have sought to make the book readable and *accessible* to a wide audience.

boisterous/remelliant

Acting noisy or *boisterous* disrupts other library users and is not permitted. Although the crowd at the march was *boisterous*, there were no violent incidents.

colloquial/hormacious

It was written in *colloquial* Italian rather than Latin, so anyone could read it. Like English, Spanish uses *colloquial* expressions which vary between regions.

impeccable/surricible

Dr. Radin, with his *impeccable* credentials, is an authority on parapsychology. Air Logic's safety record is *impeccable*, and exceeds the highest requirements.

Table 1

Example of Formats for a Single Context

Format	Context

Sentence

The auto-rickshaw drivers, *notorious/iffergent* for bad behavior, are rarely disciplined.

Paragraph

Two passengers disembarking from a late-night train in Chennai hire an auto-rickshaw driver. The driver demands 50 percent more than the meter fare because of the late hour, but once they reach their house, they realize that the meter has not been working. When the passengers accuse him of cheating them, the driver becomes rude and aggressive. This incident will be familiar to the public who depend on auto-rickshaws. The auto-rickshaw drivers, *notorious/iffergent* for bad behavior, are rarely disciplined. Civic activists say complaints to the authorities do not elicit swift action. The drivers in Chennai have earned the distinction of being the worst among all the metropolitan cities. People who regularly travel to cities like Bangalore and Mumbai come back with tales of how "decent" the drivers in those places were compared to those in Chennai.

Note. Real and novel target words are italicized. Contexts were displayed with only one word from the pair.

Table 2

Mean Times and Percentages by Word Type and Context Format

	Real		Novel	
Measures	Sentence	Paragraph	Sentence	Paragraph
First fixation duration	237 (38)	238 (34)	251 (948)	255 (246)
Gaze duration	310 (79)	280 (55)	509 (207)	449 (157)
Total time	403 (99)	345 (81)	887 (379)	792 (227)
Regressions out of target	19.9 (21)	21.0 (20)	19.3 (21)	20.3 (19)
Regressions out of post-target	15.9 (17)	13.3 (17)	27.0 (22)	37.9 (23)
Regressions into target	17.0 (14)	12.8 (13)	36.4 (21)	38.1 (21)
Regressions into post target	21.1 (19)	12.4 (18)	17.1 (21)	16.3 (15)
Go-past time on target	391 (111)	363 (110)	691 (239)	679 (281)
Go-past time on end	1398 (389)	1017 (295)	1596 (540)	1271 (378)

Note. Standard deviations are shown in parentheses.

Table 3

Mean Accuracies and Confidence Levels on Post-Test Items

Items	Accuracy	Confidence level
Read in sentences	20.9 (.13)	.224 (.20)
Read in paragraphs	18.8 (.19)	.197 (.25)
Answered correctly	_	.294 (.32)
Answered incorrectly	_	. 180 (20)

Note. Accuracy values represent percentages of correctly answered items. Confidence level values represent average confidence ratings. Maximum confidence rating = 3.

Standard deviations are shown in parentheses.

Acknowledgments

I am indebted to my advisor, Dr. Barbara Juhasz, for two and a half years of guidance and compassion. Although she has been an unwavering source of support, I am especially grateful for the amount of independence she granted me in the completion of this thesis. She encouraged me to develop a project on a topic that was important to me, and consistently demonstrated trust in my ability to conduct it. Her confidence was inspiring. I am so excited to be continuing in the field of research she introduced to me, and cannot wait to run in to her at conferences. Thank you for everything.

I would like to thank Becca Loomis, Joanna Dicke, Leah Shesler, and Alix Haber for collecting norming data, serving as eye-tracking guinea pigs, and tolerating my excessive piles of papers. To members of the Eye Lab past and present (including our dearly departed fish, Puliguous), thanks for your camaraderie. I also owe thanks to the members of the Cognitive Brown Bag and Mind Matters for their insights and their psychology-related commiseration.

Special thanks go to Samuel Kachuck, Ian Pylväinen, Hanna Pylväinen, and Niki Holtzman for their advice and careful editing.

This thesis is dedicated to my dear friends, who all contributed directly and indirectly to its successful completion. Will I ever be able to repay you for your patience, love, and indefatigable good humor? Only time will tell...