INDIVIDUAL DIFFERENCES IN ORTHOGRAPHIC PROCESSING

INDIVIDUAL DIFFERENCES IN ORTHOGRAPHIC PROCESSING

By KAITLIN FALKAUSKAS, B.SC.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University MASTER OF SCIENCE (2014) Hamilton, Ontario
(Cognitive Science of Language)

TITLE: Individual differences in orthographic processing

AUTHOR: Kaitlin Falkauskas, B.Sc. (McMaster University)

SUPERVISOR: Dr. V. Kuperman

NUMBER OF PAGES: x, 93

Abstract

This study aimed to examine how variable exposure to language statistical patterns affects reading behaviour, specifically, eye-movements during reading. The statistical patterns of language affect how individuals store, produce and comprehend language. When reading, individuals with greater linguistic proficiency typically have been shown to rely less on language statistical information compared to less proficient readers. Based on the Lexical Quality Hypothesis, however, it was hypothesized that spelling bias, a print-specific probabilistic cue, may only be utilized for representations with sufficient strengths of representation - through increased exposure to print in individuals, or through higher frequency of occurrence for individual words, since these individuals, and these words, would be expected to have representations of high quality in the reader's mental lexicon. Undergraduate students with varying amounts of reading experience were presented with sentences containing English noun-noun compound words that varied in spelling bias, i.e. the probability of occurring in text either as spaced (*window sill*) or concatenated (*windowsill*). Linear mixed effect multiple regression models were fitted to the eye-movement data and demonstrated that compound words presented in their more supported format - i.e. the format with the highest bias, were read faster, but that this effect was modulated by reading experience, as measured by a test of exposure to print, as well as by word frequency. Only individuals with the most reading experience, and words with the highest frequencies benefited from this facilitatory effect of bias. This distributional property can thus be used during reading, but only when individuals' lexical representations are of sufficiently high quality. The results of this study thus suggest that future research considering the relationship between linguistic properties and reading must consider individual differences in reading skill and exposure.

Acknowledgements

Primary Supervisory: Dr. Victor Kuperman

Committee Members: Dr. Victor Kuperman, Dr. Anna Moro, and Dr. Catherine Anderson

Reading Lab: Thank you to all of the members of the reading lab for your continuing support, especially Regina Henry.

This work has been supported by a SSHRC CGS Master's scholarship awarded to the author (Reference number 766-2013-0507).

Table of Contents

List of Figures

List of Tables

Declaration of Academic Achievement

The conceptualization of this study was conducted by myself, in collaboration with Dr. Kuperman. I implemented the experiment and tested participants under the supervision of Dr. Kuperman. The statistical analysis was completed by myself and Dr. Kuperman. The contributions to the submitted article will be noted in the preface to that chapter. The remainder of the thesis was written by myself, with comments from Dr. Kuperman, Dr. Moro, and Dr. Anderson.

1.0 Introduction

1.1 Statistics of Language use / Lexical Statistics

Language statistical patterns are evident in the structure of language, spanning from phonology to discourse. These patterns have consequences for the storage of language in the brain, which then in turn affects language production and comprehension (e.g. Arnon & Snider, 2010; Bybee, 2007; Jaeger & Tily, 2011; Jurafsky, 2003; Jurafsky, Bell, Gregory & Raymond, 2001; Seidenberg & MacDonald, 1999). These language statistical patterns manifest themselves as differences in frequencies of occurrence of segments, words, phrases, etc. Given these differences in frequencies of occurrence, at any time during speech or writing, any given segment, word or phrase will have a probability of occurring. The probability of a given linguistic unit can influence how difficult that segment is to process or produce; if a structure is infrequent or unpredictable, it will be difficult to process (Jaeger & Tily, 2011). This has been demonstrated, for example, through the well-attested effect of word frequency (Inhoff & Rayner, 1986; Kliegl, Grabner, Rolfs & Engbert, 2004), in which more frequent words, in general, are processed more quickly than lower frequency words. Words that are more frequent have stronger memory representations, and therefore are more accessible during language production or comprehension, meaning that they are easier to produce or comprehend (Bybee, 2007). Similarly, more frequent or predictable forms are more likely to be phonologically reduced during language production (Jurafsky, 2001). Individuals are sensitive to a variety of language statistical information, which can be seen through research on word learning and artificial languages.

1.2 Insights from Artificial Language Learning

An ideal situation would be one in which researchers would know the exact frequency of every tested linguistic unit, for each participant. Since this is not feasible, experiments have been conducted in which the distributional properties of linguistic units are artificially manipulated through training participants on artificial languages. For example, adults have been shown to be sensitive to the distributional properties of sounds, allowing them to segment the words of an artificial stream of speech (Saffran, 2003; Saffran, Newport & Aslin, 1996).

More recently, Joseph, Wonnacott, Forbes, and Nation (2014) trained participants on novel words in context over five days. Half of these words were seen early, i.e. on the first day, whereas the other half of the words were not introduced until the second day. The authors monitored participants eye-movements as they were exposed to the novel words, as well as presenting the words to participants in a neutral context at test, and giving participants a surprise memory test. They found that fixation durations on the words decreased across days, indicating a learning effect. They also found that there was an effect of the order of acquisition of words, such that words that were first seen on the first day were read faster than those that were first seen on the second day. This demonstrates that readers can learn words through being exposed to them in context, and that this learning influences eye-movement behaviour.

Individuals are sensitive to very fine-grained language statistical information. Vouloumanos (2007) conducted a study that aimed to determine how sensitive individuals are to fine-grained differences in word-object co-occurrence frequencies. Participants were trained on word-object pairs, and later completed a two alternative forced choice task to match a word with an object. The frequency with which the words and objects were presented together during training varied such that some objects were always labeled with the same word, while other objects were given multiple labels. Words could occur with an object either 10, 8, 6, 2, or 1 times. Vouloumanos (2007) found that participants responded more accurately to higher probability pairings, but, importantly that participants responded above chance, even when choosing between a pairing that had occurred twice compared to one that had occurred once. This demonstrates that participants were sensitive to very slight differences in probabilities.

The majority of previous studies on artificial language learning have focused on group effects, without considering individual differences. The following study, however, although considering syntactic level statistics, highlights the effects of individual differences on statistical learning. Misyak, Christiansen and Tomblin (2010) examined how individual differences in statistical learning of non-adjacent dependencies affected how individuals read sentences containing non-adjacent dependencies. Participants were exposed to sets of non-words that fit an *aXb* pattern. The authors were interested in how participants' sensitivity to the relationship between *a* and *b* affected how they read subject- and object-relative clause sentences. The authors found that there were individual differences in participants' ability to learn these dependencies, and that these differences affected how participants read subject- and object-relative sentences. Individuals who were better at statistical learning read the target verbs faster than those who were not as good at learning the non-adjacent dependencies. Additionally, their results showed that lower scoring individuals had greater difficulty with the object-relative compared to subject-relative sentences, while this difference was reduced in higher scoring individuals. These results highlight the role of statistical learning, and how it relates to language processing.

The effects of the distributional properties of language, both natural and artificial, have been well-attested at the group level, however, relatively less focus has been placed on the effects of these patterns on the individual level. To examine the effects of language statistics on individuals, some theoretical background is required. The Lexical Quality Hypothesis provides a framework which allows one to conceptualize the way in which individual differences in exposure to print can be manifested as behavioural differences during reading.

1.3 Lexical Quality Hypothesis

The theoretical framework of the Lexical Quality Hypothesis allows for an examination of the way in which individual experiences are translated into mental representations of linguistic units, and how these mental representations influence language behaviour. The Lexical Quality Hypothesis emphasizes the role of efficient word recognition in skilled reading, and suggests that a “crisp”, high-quality representation of a word must have precise, fully specified information regarding the word's orthography, phonology, morpho-syntax and meaning (Perfetti, 1985; 2007; Perfetti & Hart, 2001; Perfetti & Hart, 2002). The quality of a word’s representation is defined as “the extent to which the reader’s knowledge of a given word represents the word’s form and meaning constituents in a way that is both precise and flexible” (Perfetti, 2007, p. 359). In high quality representations: orthography, phonology and semantics, are tightly bound through automatic mappings between constituents. With respect to orthography, which is the focus of the present paper, high quality lexical representations will include fully specified knowledge of the symbols in the word, including letters, spaces and hyphens, as well as the order in which those symbols occur. A representation with lower quality with respect to orthography may lack information about the symbols in the word and/or their positions (Perfetti, 2007). Having high quality word representations allows individuals to retrieve words efficiently: rapidly and without a great deal of cognitive resources. Lexical quality increases with repeated exposure to a given word, and to printed materials in general (Perfetti, 2007).

A demonstration of the effect of experience on lexical quality is as follows. Perfetti and Hart (2001) had participants make semantic relatedness judgements on pairs of words, some of which were homophones. Skilled readers showed no differences in processing time for homophones compared to non-homophone controls when the homophone presented was of high frequency, as in the pair *whales-cries*, but did show processing difficulty when the homophone was of low frequency, such as in the pair *wails-dolphins*. A later study, reported in Perfetti (2007), demonstrated that this effect reversed after training with the low frequency homophone variant, such that the originally lower frequency variant led to similar behavioural responses as the originally higher frequency variant did. Increasing experience with a given variant led to a change in the relative frequency of the variants, which led to changes observable behaviour when making decisions related to the variants. Words with high lexical quality, e.g. the higher frequency variants used by Perfetti and Hart (2001), are more resistant to form confusion. This has been shown clearly with homophones, which share phonology, but not form or meaning, however similar effects can be hypothesized for variants that share phonology and meaning, but differ in form. This type of alternation will be discussed in section 1.5. Essentially, the variant that is more supported in an individual's mental lexicon enjoys a processing benefit, while the less supported variant will be processed with more difficulty.

Since each individual has a unique experience with any given word, then we can expect that individuals will vary in the lexical quality of their representation for any given word. For example, one individual may have seen a given word more times than another individual, which would result in differences in word frequencies for each of these individuals. The difference in word frequency, across these individuals, as well as differences in reading skills, could then lead to differences in the lexical quality of the representations of that given word for the two individuals. In other words, the differences in frequency could lead to differences in the strength of individuals' word representations. These differences in the strength or quality of these mental representations could then lead to differences in observable reading behaviour across individuals (e.g. MacDonald & Christiansen, 2002). MacDonald and Christiansen (2002) argue that difference in language comprehension are attributable to differences in experience, as well as biological differences in phonological representations. A similar process may occur for any other distributional properties.

These differences in the lexical quality of representations can vary within an individual, due to differences in experience with specific words, as well as between individuals, due to overall experience with printed text. As mentioned, these differences in lexical quality can lead to differences in visual word recognition. Individuals with reduced amount of exposure to print will have representations of lower lexical quality, and consequently, weaker co-activations of its orthography, phonology and meaning, which will, in turn, make reading less efficient and more cognitively demanding (e.g. Perfetti, 2007). Differences in experience are thus expected to be borne out as main effects in studies that examine some form of processing ease or difficulty as a dependent measure. Main effects of experience, as well as reading-related skills have been observed across multiple studies, and will be discussed further in section 1.4.

Increasing experience with any given word would lead to a stronger representation, and more efficient processing of that word, across all individuals. It is therefore expected that high frequency words will be processed similarly across individuals of all skill levels, since these words are expected to be encountered enough to have high quality representations for all individuals. To gain high quality representations of low frequency words, individuals must be exposed to a large sample of written texts (Kuperman & Van Dyke, 2013). This means that readers with less exposure to print are likely to have lower quality representations for these low frequency words, compared to more experienced individuals, causing them to have more processing difficulty when reading these low frequency words. Interactions between the lexical properties of words, such as the one described between reading experience and word frequency, do occur in the literature, and will be discussed.

1.4 Individual Differences in Visual Word Processing

Individual differences in reading experience and skill greatly affect various aspects of reading. Buswell (1922) first demonstrated the variability that exists between individuals with respect to their reading behaviour. He examined three eye-movement measures: average number of fixations per line of text, average fixation length, and average number of fixations, for participants ranging from first grade to college level. He detailed the large amount of variation across individuals at each grade level, noting that some individuals had more fixations, longer fixations, and more regressions than average. Later studies connected individual differences in reading proficiency, as indexed by performance on skill tests, as well as differences in reading experience, as indexed by tests of exposure to print, with these differences in reading behaviour. Individuals with lower scores on a variety of tests have been shown to have longer fixation durations (Everatt & Underwood, 1994; Kuperman and Van Dyke, 2011a; Underwood, Hubbard, & Wilkinson, 1990), an increased number of regressions (Griffin, Walton, & Ives, 1974; Kuperman and Van Dyke, 2011a), were less likely to skip words (Kuperman and Van Dyke, 2011a), and were more likely to fixate on words multiple times (Kuperman and Van Dyke, 2011a) compared to individuals with higher scores. Similar effects have also been observed in individuals with dyslexia (Hawelka, Gagl & Wimmer, 2010). Individuals with lower reading proficiency and experience have also been shown to get less of a preview benefit while reading (Chace, Rayner & Well, 2005). As well, individuals with reduced reading skill and experience have been shown to have smaller perceptual spans (Rayner, Slattery & Bélanger, 2010; Veldre & Andrews, 2014). Veldre and Andrews (2014) presented sentences using a gaze contingent, moving window paradigm, in which only a set number of characters, ranging from three to fifteen, were presented to the right of where a participant fixated. They found that better readers and spellers were facilitated compared to less proficient participants when the given span was eleven characters or greater, and that the higher-scoring individuals were also harmed more by being given a smaller character span. This suggests that individuals with higher reading and spelling scores have larger perceptual spans.

Less skilled readers have also been shown to rely more on phonological information (Jared, Levy & Rayner, 1999). The authors presented participants containing words that either had the correct spelling, an incorrect spelling, or were written as a homophone of the target word. They found that good and poor readers, based on scores on a comprehension test, were differentially sensitive to homophone errors. Skilled readers had longer fixation durations on spelling and homophone errors, whereas less skilled readers had longer fixation durations on spelling errors, but similar fixation durations for homophone errors and correctly spelled target words. This indicates that less skilled readers rely more on phonology compared to skilled readers, since the durations on correctly spelled targets and homophone errors did not differ.

In addition to the main effects of individual skill and experience, several interactions have been observed, demonstrating that the individual differences in skill and exposure to print lead to differential effects of the lexical properties of words on visual word processing. These distributional measures that have been tested include word frequency (Adelman, Sabatos-DeVito, Marquis, & Estes, 2014; Chateau & Jared, 2000; Sears, Siakaluk, Chow, & Buchanan, 2008), word length (Butler & Hains, 1979), sentence context (Hersch & Andrews, 2012) and are robustly established across experimental paradigms. The review of these studies will be broken down by the paradigms used.

1.4.1 Sentence recall

The sentence recall task has been used to demonstrate how individual differences in reading skill interact with language statistical properties. Hersch and Andrews (2012) had participants complete a sentence recall task in which a word pair was presented at some point in the sentence, and participants were asked to choose which word fit best in the sentence, and to say that word during the later sentence recall. In some sentences, the context that biased the target word was presented before the word pair, and in some sentences the biasing context was presented after. The authors found that individuals with higher reading and spelling scores showed less of a difference in accuracy between whether the context was before or after the target pair, compared to less skilled readers and spellers. The authors suggest that these results indicate that more proficient readers and spellers rely less on contextual information when reading.

1.4.2 Naming/Lexical Decision

Similar interactions have also been observed for lexical decision tasks, where participants are asked to decide if a given string is a word or not, when presented with a set of words and non-words. Butler & Hains (1979) used naming and lexical decision tasks and found an interaction between vocabulary size and word length, such that individuals with a larger vocabulary were less affected by word length than individuals with a smaller vocabulary. Adelman, Sabatos-DeVito, Marquis, and Estes (2014) tested the effects of lexical properties on word naming and lexical decision across individuals. They found evidence of individual differences in the sensitivity to: word frequency, word length, non-word length, lexicality, exception words and the location of irregularities (i.e. where the letter that indicated irregularity occurred in each word). Chateau & Jared (2000) found that individuals with greater exposure to print were less sensitive to word frequency in a lexical decision task. Similar results were found by Sears, Siakaluk, Chow, and Buchanan (2008). Yap and colleagues examined the relationship between vocabulary size and sensitivity to various lexical characteristics using data from the English Lexicon Project (Yap, Balota, Sibley & Ratcliff, 2012). The authors analyzed data from speeded pronunciation and lexical decision tasks and found a negative correlation between vocabulary size and a principal component consisting of orthographic and phonological neighbourhood size, as well as another component containing the frequency and semantic characteristics of words. Individuals with higher vocabulary sizes were less affected by these lexical properties.

Overall, these studies have demonstrated that individuals with lower scores on a variety of test: use more contextual information, and are more sensitive to word frequency, word length, non-word length, lexicality, exception words, the location of irregularities, and neighbourhood size. The sentence recall and lexical decision tasks that were used, however, are not naturalistic, and do not allow experimenters to study word processing in natural, sentence contexts.

1.4.3 Event-related potentials (ERPs)

Event-related potentials (ERPs) allow researchers to investigate the electrophysiological activity in response to stimulus events with good temporal resolution. Interactions of skill and language statistical properties have also been found using this methodology. Perfetti, Wlotko and Hart (2005) examined how individual differences in reading skill affect how individuals learn new words. They trained participants on low frequency words that were previously unknown to the participants. Participants then completed semantic decisions on the trained words, as well as untrained low frequency words and familiar, but not recently presented, mid-frequency words. This required participants to decide if a word was semantically related to a second word. They found that behaviourally, skilled readers were more accurate when judging the meanings of the newly trained words. Electrophysiologically, skilled readers showed a greater difference in P600 amplitude for trained words compared to the other types of words, compared to less skilled readers. The authors suggest that this indicates that skilled readers form stronger episodic memory traces for the trained words than less skilled readers do. Balass, Nelson and Perfetti (2010) found similar results, and added that skilled readers were better able to use phonological and orthographic information paired with the word meanings compared to less skilled readers. They also demonstrated that the best performance occurred for paired orthographic and semantic information, but only for skilled readers. This highlights the important role of orthographic information in word learning.

Frishkoff, Perfetti, and Westbury (2009) examined the processing of words with different levels of semantic knowledge, across participants, while considering individual differences in comprehension skill. They had participants perform lexical decisions on low frequency, rare, and very rare words, as well as orthographically non-anomalous pseudowords, while recording the event-related potentials in response to the stimuli. The authors found that individuals with the highest scores responded differently to partially semantically known, or *fronteir*, words, compared to lower scoring individuals. The more skilled comprehenders showed increased activation in the left ventral temporal cortex for these words, and behaviourally, were less likely to treat these words as real words.

Presently, ERP research generally relies on the serial visual presentation of stimuli. This manner of presentation does not reflect naturalistic processing, as participants are allowed a set amount of time to fixate on each word, are not able to move their eyes from word to word, and thus cannot make regressions, or skip words.. The simultaneous recording of ERPs and eye-movements may be a useful tool for studying these types of effects, however this methodology was not in the scope of the present project.

1.4.4 Eye-tracking

Studies have also been conducted using the eye-tracking methodology, in which participants read sentence while their eye-movements are monitored. Ashby, Rayner and Clifton (2005) examined the effect of word frequency on reading times in constrained and non-constrained sentence contexts for individuals with different skill levels. The authors found that in non-constrained contexts, average and highly skilled readers had shorter gaze durations for less frequent words, and that this effect was marginally significantly lower in the highly skilled readers. The group differences were most evident in low frequency words. In the constrained contexts, average readers did not show an effect of word frequency on gaze duration, while highly skilled readers still read higher frequency words faster. Highly skilled and average readers had similar gaze durations on low frequency words. However, in unpredictable contexts, average readers had greater spillover effects for low frequency words, indicating that the processing of these words continued after the eye passed the word. The authors suggest that in unpredictable contexts, average readers do not access unpredictable, low frequency words until after the eye has moved past them. A comparison of both experiments indicated that average readers were highly affected by frequency in non-constrained contexts, but were not affected by frequency in constrained, unpredictable contexts, while the effect of frequency for highly skilled readers did not change across levels of predictability. Overall, average readers seem less able to handle unpredictable contexts compared to highly skilled readers. The authors point to automaticity in highly skilled readers as being a potential reason behind this effect. Contrary to this, Whitford and Titone (2014) found a smaller effect of word frequency during a passage reading task for individuals with lower comprehension scores. The authors did not, however, independently measure comprehension scores, so the results are not conclusive. The authors suggest that their measure of passage comprehension may be indexing reading strategies, rather than reading skill.

Individuals with lower scores on measures of reading skill were found to be more affected by word length than more skill readers, as indexed by a larger correlation between word length and reading times compared to more skilled readers (Kuperman & Van Dyke, 2011a). Similarly, less skilled readers were also more affected by word frequency compared to more skilled readers.

Kuperman and Van Dyke (2011b) compared the effects of whole-word and base frequency of derived words (e.g. the frequency of *trucker*, or the base *truck*) in individuals of different skill levels. Individuals of all skill levels were similarly affected by whole-word frequency - more frequent words were read faster. Differential effects, however, were observed for base frequency. Individuals who scored lowest on a segmentation task showed a facilitatory effect of base frequency. This effect reversed as the segmentation score increased, such that the mid range of scorers were not affected by base frequency, and the highest scorers were hindered by base frequency. Low scoring individuals seem to get a recognition boost from the base words, whereas high scoring individuals may experience competition effects.

Overall, these eye-tracking studies have shown that individuals with lower scores on skill tests are less able to comprehend less predictable sentences, are more sensitive to word length and to word frequency (however, see Whitford & Titone, 2014). The use of eye-tracking as a methodology is useful for this type of research for various reasons. An individual's eye movements when reading are indicative of the cognitive processes occurring (McConkie, Hogaboam, Wolverton, Zola & Lucas, 1979; Rayner, 1978, 1998; Rayner, Sereno, Morris, Schmauder & Clifton, 1989; and references therein). It is well-attested that manipulating the properties of the words being read leads to differences in fixation durations (Rayner & Duffy, 1986). Using eye-tracking allows researchers to gain insight into a reader's cognitive processing on the millisecond level, thus making it a temporally sensitive measurement tool. Another benefit of the eye-tracking methodology is that it allows researchers to capture natural reading behaviour. This allows researchers to capture complex reading behaviour, such as skipped words and regressions, which could not be done using measures such as self-paced reading or Rapid Serial Visual Presentation.

All of the previous studies that have found interactions between reading skill or experience and the distributional properties of language have found effects such that individuals with less skill or experience rely more on the distributional properties, or other sources of information, when reading. Given the observation that lexical quality increases with repeated exposure to a given word and to printed text in general, it is possible that there are distributional patterns that can only affect individuals with representations of sufficiently high quality to encode this information. We would expect therefore that individuals with extensive exposure to print to be preferentially sensitive to these patterns, and that all individuals would be sensitive to these patterns for words of high enough frequency. A distributional pattern that could be expected to cause these differences in responses for highly experienced individuals and high frequency words would be specific to print, and likely not existent in spoken language, since there is more variability in reading experience and proficiency than in spoken language experience and proficiency. Thus, a fluent language speaker who has never seen printed text would be completely naive to such a pattern, compared to an individual with many years of extensive reading experience, whose representations may be of a sufficient quality so as to include this information. The possibility of preferential effects for high quality representations - through experience with specific words or through global exposure to print - has not, to our knowledge, been extensively explored.

1.5 Compound Spacing in English

A potential print-specific pattern that may only be applicable for individuals with large amounts of reading experience or for high frequency words, is the alternation between the orthographic variants of English noun-noun compound words. Written compounds in English can be spelled in one of three spatial formats: spaced (*house plant*), concatenated (*baseball*) or hyphenated (*student-teacher*). Although spelling conventions may dictate the spelling format of compound words, there is variability in what format is actually used (Sepp, 2006; Shie, 2002).

This alternation is ideal for these purposes because it is specific to the printed medium, and therefore allows for an examination of the role of reading experience. Since this alternation is not found in spoken language, participants can be compared based on degree of experience with printed text, rather than spoken language proficiency, since our participant pool is expected to contain fluent speakers. Individuals with greater amounts of exposure to the target compound words, as well as to printed text in general, are expected to be affected more by probability compared to individuals with less exposure to the target words or to print in general. Readers who are less familiar with how a compound is represented in print, or with print in general, are predicted to be naïve to subtle differences in variants' probabilities.

1.6 Isolating Role of Probability (meaning preserving alternations)

Meaning preserving alternations have been used to examine the role of language statistical properties on the storage, production and comprehension of language. One example of a meaning preserving alternation which has been widely studied is the dative alternation, in which a sentence containing a subject, a verb and two objects can either be realized as a prepositional dative structure (*The man gave the book to the boy.*) or a double object structure (*The man gave the boy the book.*).

Bresnan, Cueni, Nikitina & Baayen, (2007) examined the factors that affected the probability of which of these structures would be chosen. They found, through statistical modelling, that a variety of syntactic and semantic factors, such as givenness, animacy, and other variables, reliably predicted which structure would be used (see also Bresnan & Ford, 2010).The structure of sentences can thus be altered to manipulate the probability of a prepositional or double object dative, to examine whether readers are sensitive to these probabilities. This has been done by Tily, Gahl, Arnon, Snider, Kothari & Bresnan (2009) who found that the probability of a given variant affected the duration of the preposition (*to*) in spoken sentences, as well as the likelihood of producing a disfluency. Less probable variants were more likely to have longer prepositions and were more likely to contain disfluencies. Similar results were found by Kuperman and Bresnan (2012).

The probability of the variants also affects how they are read. Bresnan and Ford (2010) examined the role of the probability of the prepositional dative structure on comprehension. The authors presented with dative sentences, while participants made lexical decisions on the words in those sentences. The authors found participants' lexical decision latencies were faster on the word *to* when the prepositional dative was more probable. This indicates that individuals are sensitive to the probabilities of the variants, and that this information influences visual word processing.

Meaning preserving alternations are an ideal tool to study effects of probability because the structure of the variants changes without causing a change in meaning. If the variants have different probabilities of occurrence, then we can attribute differences in processing to the probabilistic differences or the difference in form, rather than differences in meaning, or other factors.

A meaning preserving alternation at the lexical level, which has not been studied in this manner, is the alternation that occurs between spelling formats in English compound words. When considering the spelling alternation in noun-

noun compound words in English, the spelling format is the only difference between the alternative forms, so each set of alternating forms serves as its own control, in that the only aspect that differs is the one which will be tested.

1.7 Previous Studies on Compound Spacing in Multiple Languages

The effect of spelling format on the visual processing of compound words has been studied, cross-linguistically, using the eye-tracking methodology (Bertram, Kuperman, Baayen & Hyönä, 2011; Cherng, 2008; Inhoff, Radach & Heller, 2000; Juhasz, Inhoff & Rayner, 2005). Inhoff et al. (2000) examined the role of adding inter-word spaces on the reading of three-component German compound words, which are typically concatenated, as dictated by the spelling conventions of German. When inter-word spaces were added between the constituents of compound words, naming latencies decreased, and fixation durations were shorter, compared to those to compound words that were concatenated, despite the fact that inter-word spaces are orthographically illegal in German. This suggests an overall benefit in the reading of compound words containing inter-word spaces, however, there were differences in fixation durations when early fixations were compared to later fixations. For compounds presented as spaced, earlier fixations on target words were shorter than later fixations on the words (i.e. first fixations durations were shortest, followed by increasing subsequent fixation durations). This is contrary to the general pattern of decreasing fixation durations for subsequent fixations on a word (Rayner, Sereno & Raney, 1996). The authors suggested that there are two processes

occurring when a space is added to a typically concatenated compound word. First, an early benefit due to facilitated access to the constituents of the compound word, and second, a cost caused by difficulties integrating the meanings of the compounds. When a space is added to a concatenated compound word, it is no longer clear that the constituents of that ccompound are orthographically
, and potentially semantically unified, thus uncertainty may occur regarding the relationship between the constituents of the compound until after all components have been read. Plausibility effects may also play a role if the initial constituent is integrated into the context of the sentence, causing later reanalysis if subsequent constituents of the compound no longer fit within the context (cf. Staub, Rayner, Pollatsek, Hyönä & Majewski, 2007). A similar pattern of effects has also been observed when adding spaces to English concatenated compounds such as presenting *earthquake* as *earth quake* (Juhasz et al., 2005), and when adding hyphens to typically non-spaced compounds (Cherng, 2008; Bertram, Kuperman, Baayen & Hyönä, 2011). Juhasz and colleagues (2005) examined the effect of compound spelling on the visual processing of noun-noun compound words in English, using eye-tracking while reading for typically spaced and typically concatenated compound words. Participants read sentences containing typically spaced and typically concatenated compound words, as well as the same compounds in the opposite formats - typically spaced compounds presented as concatenated and the reverse. They found that first fixations on spaced compound word were shorter than those on concatenated compounds, however gaze

durations, which include all of the fixations in the first pass of reading, were shorter for concatenated compounds compared to spaced compounds. These results demonstrate that the time-course of processing for spaced and concatenated compounds in English is analogous to the effects found by Inhoff and colleagues (2000) for German compound words. Bertram and colleagues (2011) also found that gaze durations were longer on Dutch compound words presented with a hyphen compared to concatenated compounds, but that this effect was confined to the beginning of the experiment. The authors found that the effect was attenuated later in the experiment, likely due to a learning effect. Based on these studies, it is hypothesized that the present study will replicate the characteristic time-course of effects.

The aforementioned studies have focused on compound words that are highly biased, potentially even categorically, towards one spelling variant - spaced, concatenated or hyphenated, however, little is known about the processing of compound words that are intermediately biased to any given spelling format. A study conducted by Kuperman and Bertram (2013) provided the groundwork for examining this question. The authors calculated the probabilities of the three spelling variants, spaced, concatenated, or hyphenated, for noun-noun compounds in English using data from the Wikipedia corpus, as well as characterizing the factors that affect the probability of a compound occurring in a given format. They found that orthographically longer compound words and compounds with a greater semantic association between their constituents were more likely to be spaced than concatenated, compounds with a higher frequency considered across all formats were more likely to be concatenated than spaced, and compounds with an orthographic family (which includes compounds that share either the left or right constituent) whose members were biased towards a given format were more likely to occur in that format. Kuperman and Bertram (2013) also examined the effect of the probability of observing a compound word in its concatenated form on lexical decision latencies to concatenated compounds, and found that compounds with a higher bias towards concatenation had shorter response latencies compared to less biased compound words. This indicates that participants are sensitive to the spelling bias, and that reading a compound word as concatenated, if it often occurs as concatenated, facilitates processing. Due to the composition of the lexical decision database, however, Kuperman and Bertram (2013) only addressed the processing of one of the three spelling variants, and did not consider how the probabilistic biases toward the spelling variants affect individuals with varied levels of reading experience (see section 1.4 for a discussion of the relevance of individual difference). The current body of research on the effects of compound spelling is therefore lacking in two areas: first, the effects of compound words with intermediate propensities toward the spelling formats, and second, an examination of how these differences in probabilities may affect individuals with varying levels of reading experience.

1.8 The Present Study

The present study examined the way in which individual differences in reading experience, as well as the lexical properties of words influences how individuals read noun-noun compound words with alternating spellings. Using the spelling biases for English compound words (Kuperman & Bertram, 2013), compounds were selected across the range of biases, such that the level of support of the presentation format of the compounds would vary, based on participants' experience with those words. These compounds were presented to participants in sentences while their eye-movements were monitored, and participants completed a variety of skill tests. The hypotheses of the present study are as follows. First, that the findings of an early advantage and late cost for spaced compounds will be replicated (Bertram, Kuperman, Baayen & Hyönä, 2011; Cherng, 2008; Inhoff, et al., 2000; Juhasz et al., 2005). Second, in accordance with the results of Kuperman and Bertram (2013), it was expected that words presented in a more biased, or supported format would be read faster overall. Third, based on the Lexical Quality Hypothesis, spelling bias was expected to interact with word frequency, as well as reading experience. More experienced readers were expected to have higher quality lexical representations compared to less skilled readers, and all readers were expected to have higher quality representations for high frequency words. With more exposure, and higher quality representations, individuals' representations will be more likely to encode the bias information, thus, setting up an expectation of how the compound word will be spelled. To summarize, the effect of bias was expected to be stronger in individuals with greater amounts of reading experience, and for higher frequency words. In the following chapter, the submitted article is presented.

2.0 When experience meets language statistics: Individual variability in processing English compound words

2.1 Preface

This article has been submitted to the Journal of Experimental Psychology: Learning, Memory and Cognition. This article was written by myself, with guidance and edits from Dr. Victor Kuperman.

2.2 Abstract

Statistical patterns of language use demonstrably affect language comprehension and language production. This study set out to determine whether the variable amount of exposure to such patterns leads to individual differences in reading behaviour as measured via eye-movements. Previous studies have demonstrated that more proficient readers are less influenced by distributional biases in language (e.g. frequency, predictability, transitional probability) than poor readers. We hypothesized that a probabilistic bias that is characteristic of written but not spoken language would preferentially affect readers with greater exposure to printed materials in general and to the specific pattern engendering the bias. Readers of varying reading experience were presented with sentences including English compound words that can occur in two spelling formats with differing probabilities: concatenated (*windowsill*, used 40% of the time) or spaced (*window sill*, 60%). Linear mixed effects multiple regression models fitted to the eye-movement measures showed that the probabilistic bias towards the presented spelling had a stronger facilitatory effect on compounds that occurred more frequently (in any spelling) and on readers with higher scores on a test of exposure-to-print. Thus, the amount of support towards the compound’s spelling is effectively exploited when reading, but only when the spelling patterns are entrenched in an individual's mental lexicon via overall exposure to print and to compounds with alternating spelling. We argue that research on the interplay of language use and structure is incomplete without proper characterization of how particular individuals, with varying levels of experience and skill, learn these language structures.

Keywords: compound words, morphology, eye-movements, individual differences, learning

2.3 Introduction

There is a consensus that the statistical patterns of language use affect both the representation of linguistic structure – from phonological segments to discourse units – in the brain, and language production and comprehension (see, among many other reviews, Jaeger & Tily, 2011; Jurafsky, 2003; Jurafsky, Bell, Gregory & Raymond, 2001; Seidenberg & MacDonald, 1999). Although the effects of distributional patterns on language processing are well-attested at the aggregate level, the following questions are less explored: (a) how do differences in language experience affect the variation in distributional patterns across individuals, (b) how do differences in experience, as well as cognitive abilities, translate into variable strengths in individuals' mental representations of language units, and finally (c) how do these differences in representations influence individual variability in observable language behaviour (MacDonald & Christiansen, 2002). The present study contributes to the investigation of (a)-(c) by examining individual differences in the visual comprehension of English compound words that allow for alternate spellings (e.g., *girlfriend, girl-friend, girl friend*), each with its own probability of occurrence in written language. In what follows, we motivate our study in view of two largely disjoint theoretical frameworks: one that proposes a mechanism underlying variability between individuals at the word level, and one that highlights the utility of meaning-preserving linguistic alternations in studying probabilistic effects on language.

The Lexical Quality Hypothesis provides a theoretical framework to studies of individual variability. It argues that a “crisp”, high-quality representation of a word entails both precise and full specifications of the word's orthography, phonology and semantics, including the word's semantic and syntactic environments (Perfetti, 1985; 2007; Perfetti & Hart, 2002). It further argues that automatically-activated mappings occur between these three components. The quality of a word's representation is then defined as "the extent to which the reader's knowledge of a given word represents the word's form and meaning constituents" (Perfetti, 2007, p. 359). As theories of statistical learning predict (Hebb, 1949; Rescorla & Wagner, 1972; and specifically for lexical processing, Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011), lexical quality develops both with individuals' repeated exposure to a word and with individuals' increased experience with printed materials in general, which requires discriminating that word from other words. Lexical quality can therefore vary within an individual's lexicon, due to differences in exposure across words, as indexed by word frequency (Perfetti, 2007). On average, lexical quality can also vary across individuals, which can lead to variability in reading (Perfetti, 2007). With respect to orthography, which is the focus of the present paper, lexical quality can range from having fully specified representations of the symbols (letters, spaces, and hyphens) in a given string, as well as the order of those symbols, to having a representation in which not all of the symbols or their positions are known (Perfetti, 2007).

The Lexical Quality Hypothesis, along with models of statistical learning, makes a number of predictions regarding how variability in reading experience may affect behavioural indices of visual word comprehension. First, reduced exposure to print is expected to render reading more effortful overall (Perfetti, 2007), and thus is likely to appear as a main effect in every aspect of reading behaviour that is constrained by processing difficulty. This prediction has been supported by studies that explored the variability in skills and reading performance in non-clinical adult populations. For comparability with the present set of findings, our review emphasizes studies that use eye-tracking as their experimental paradigm. Indeed, since Buswell (1922), eye-tracking studies have reported individual differences in reading behaviour, including differences in fixation durations, number of fixations, and number of regressions. Since then, studies have demonstrated that less proficient (i.e. those with a weaker performance on skill tests) or less experienced readers (i.e. those with fewer years of schooling, or lower scores on exposure-to-print tests) made longer fixations on words, made more regressions, skipped fewer words, and were more likely to fixate on words multiple times, compared to more proficient and experienced readers (see recent reviews in Radach & Kennedy, 2013; Rayner, Pollatsek, Ashby, & Clifton, 2012). Additionally, skilled readers have been found to have a larger perceptual span in reading (cf. Rayner, Slattery, & Bélanger, 2010; Veldre & Andrews, 2014), gain a greater parafoveal preview benefit as compared to less skilled readers (Chace, Rayner, & Well, 2005; Veldre & Andrews, 2014), and rely less on phonological information (Jared, Levy & Rayner, 1999).[1]

[1] While couched in terms of natural reading, which is a complex skill and requires extensive

The Lexical Quality Hypothesis also suggests that limited experience with a particular word would cause an individual to have a weaker representation of that word, and weaker co-activations of its orthography, phonology and semantics. Limited exposure would also result in a narrower set of contexts in which the word is learned. The Lexical Quality Hypothesis would therefore lead to the prediction that, for all individuals, repeated exposure to a word would lead to stronger representations, and consequently, less effortful processing of that word. In the case of extremely frequent words, individuals with different levels of proficiency would be expected to vary minimally in their reading behaviour, as the quality of these lexical representations would be similarly high due to extensive exposure for all readers. In the case of low frequency words, less experienced readers would be expected to be at a greater disadvantage than experienced readers, as gaining a representation of sufficient lexical quality for these lexical items requires a broader sampling of written texts (see Kuperman & Van Dyke, 2013). Thus, several experimental paradigms have robustly established both the facilitatory main effects of distributional measures of word use (e.g. word frequency, n-gram frequency, contextual diversity, or predictability) and the interactions between such measures and individuals' reading experience (for eye-tracking, see Ashby, Rayner & Clifton, 2005; Hawelka et al., 2010; Jared, Levy & Rayner, 1999; Kuperman & Van Dyke, 2011a, 2011b, 2013; Whitford & Titone,

---

language background and specific training, statistical learning has an effect even within short experimental sessions (Vouloumanos, 2008) and with artificial languages (Saffran, Newport & Aslin, 1996; Saffran 2003, and references therein). Similar learning effects can be seen when probabilities are monitored through repeated readings of words or passages (Levy, Abello & Lysynchuk, 1997; Levy, Nicholls & Kohen, 1993).

2014; for other paradigms, see e.g., Adelman, Sabatos-DeVito, Marquis, & Estes, 2014; Butler & Hains, 1979; Chateau & Jared, 2000; Hersch & Andrews, 2012; Sears, Siakaluk, Chow, & Buchanan, 2008; Yap, Balota, Sibley, & Ratcliff, 2012 and references within). The nature of these interactions is such that, while all readers show more difficulty in processing words to which they have had less exposure, proficient or more experienced readers show a relatively small contrast between words that are more and less frequent, predictable, or repeated, as compared to less proficient or experienced readers (but see Whitford & Titone, 2014).

Another interesting implication of the Lexical Quality hypothesis, and one that, to our knowledge, has not yet been extensively explored, is the possible existence of distributional patterns that would only affect individuals with extensive reading experience, i.e., those individuals who have accumulated a sufficient amount of memory traces to encode the patterns. In the case of reading, such patterns are likely to be associated with phenomena that are specific to the printed medium and that are either infrequent or non-existent in spoken language. To take an extreme example, a person who has never seen printed text would not be aware of or affected by the distribution of a specific orthographic pattern, even if he or she were fully fluent as a language speaker. A similar argument can be made about any linguistic phenomenon, in speech or in print, that is so rare as to require a considerable amount of exposure to language or its specific genres: these phenomena would not affect readers with reduced exposure.

To explore the hypothesis of greater applicability of certain probabilistic patterns to proficient readers, we used the spelling alternation observed for English noun-noun compounds. Written compounds in English can be spelled in one of three spatial formats: spaced (*house plant*), concatenated (*baseball*) or hyphenated (*student-teacher*). Although the spelling conventions of English may dictate the spelling format of compound words that supposedly should be used, there is variability in the formats used in actual writing (Sepp, 2006; Shie, 2002). A study by Kuperman and Bertram (2013) extracted noun-noun compounds from the Wikipedia corpus and found 2,306 compound words that alternated between two or all three spelling variants. For example, the word *lunchroom* appears as concatenated 70% of the time, and spaced (*lunch room*), 30% of the time. The spelling alternation in compounds is optimal for our purposes for two reasons. First, it exemplifies a meaning-preserving alternation, in which each linguistic variant is associated with a probability of realization, yet the meanings of all variants are near-identical[2]. Meaning-preserving alternations are often invoked in studies of how language statistics affects the storage, production, and recognition of variants with differing probabilities. For instance, the probability of a subject, a verb and two objects to be realized as a double object dative structure (*The man gave the boy the book.*) or as a prepositional dative (*The man gave the book to the boy.*) is contingent on multiple semantic and formal properties of the syntactic

[2] In a minority of compounds, a difference in spelling translates into a difference in meaning (*dishwasher* is a device, *dish washer* is a person employed to clean dishes). Such compounds were excluded from consideration in Kuperman and Bertram (2013) and in the present paper.

constituents (Bresnan, Cueni, Nikitina & Baayen, 2007; Bresnan & Ford, 2010) and demonstrably affects spoken production of dative constructions (Tily, Gahl, Arnon, Snider, Kothari & Bresnan, 2009; Kuperman & Bresnan, 2012) and their comprehension in reading (e.g., Tily, Hemforth, Arnon, Shuval, Snider, & Wasow, 2008). Other examples of meaning-preserving alternations which use variable ordering of syntactic constituents include the genitive alternation, and particle verbs (Bresnan, et al., 2007; Gries, 2003; Lohse, Hawkins & Wasow, 2004; Wasow, 2002). Since all of the variants have the same meaning, any behavioural differences in the comprehension or production of these variants can be attributed to their differences in probabilities or the differences in their form, and not to their differences in semantics, associated perceptual, sensorimotor or emotional experience, or world knowledge. When considering the spelling alternation between noun-noun compound words in English, the spelling format is the only difference between the alternating forms, so each set of alternating forms serves as its own control, in that the only aspect that differs is the one which will be tested. Second, the spelling alternation of interest is print-specific and therefore non-existent in spoken language. Readers with more intensive exposure to compounds with alternating spellings, as well as to printed materials in general, are therefore expected to be preferentially affected. Readers who are less familiar with how a compound is represented in print, or with print in general, are predicted to be effectively naïve to subtle differences in variants' probabilities.

The connection between the spelling format of a compound and its visual processing has been explored in several eye-tracking studies (Bertram, Kuperman, Baayen & Hyönä, 2011; Cherng, 2008; Inhoff, Radach & Heller, 2000; Juhasz, Inhoff & Rayner, 2005), though without an emphasis on probabilities of spelling variants. Inhoff and colleagues (2000) examined the role of inter-word spaces in the reading of three-component German compound words, which are typically concatenated in accordance with the conventions of German. The authors found that adding inter-word spaces between compounds decreased naming latencies, and resulted in shorter fixations compared to compounds that were presented as concatenated, even though compound spacing is orthographically illegal in German. This suggests that there is an overall benefit in the reading of compound words containing inter-word spaces. There were differences in fixation durations, however, when early fixations were compared to later fixations. When compounds were presented as spaced, earlier fixations on target words were shorter than later fixations on the words (i.e. first fixations durations were shortest, followed by increasing subsequent fixation durations). This is contrary to the general word-reading pattern in which subsequent fixations on words tend to be shorter than the first one (Rayner, Sereno & Raney, 1996). Inhoff and colleagues (2000) suggested that the difference between early and late fixations indicated two different processes occurring when spaces are added to typically concatenated compound words. The authors suggest that the early benefit is due to a facilitation of access to the individual constituents of the compound. Inhoff and colleagues (2000) further suggested that the later costs of compound spacing are due to difficulties in integrating the meanings of the compounds. Since spaced compounds are no longer clearly semantically unified, uncertainty may occur regarding the relationship between the constituents of the compound until after all components have been read. In addition, plausibility effects may occur if the initial constituent is integrated into the context of the sentence, causing later reanalysis if subsequent constituents of the compound no longer fit within the context (cf. Staub, Rayner, Pollatsek, Hyönä & Majewski, 2007). This pattern of effects – an early processing benefit followed by a later penalty – has also been found when typically concatenated compound words in English are presented as spaced such as presenting *earthquake* as *earth quake* (Juhasz et al., 2005), and when a hyphen was added to typically concatenated compounds (Cherng, 2008; Bertram et al., 2011). We expect to replicate the characteristic time-course of the effect of compound spelling in our data.

Importantly, these studies have focused on compound words with a very strong (sometimes, categorical) bias towards one spelling variant, however, little is known about the processing of compound words with intermediate propensities for each spelling variant. The groundwork for addressing this question was done by Kuperman and Bertram (2013) who calculated the probabilities of the spelling variants for alternating English noun-noun compounds based on the Wikipedia corpus, and characterized the factors affecting the probability of each variant, including the compound length, frequency and semantic association between its morphemes (*lunch* and *room*). They also found that the probability of observing a compound word in the concatenated format influenced lexical decision latencies to concatenated compounds: a higher probabilistic bias toward concatenation came with shorter responses. Yet, due to the composition of the lexical decision database, Kuperman and Bertram's (2013) study only addressed the processing of one of the three spelling variants (concatenated), and left out a question of how the probabilistic biases towards spelling variants affect individuals with varied levels of reading experience. Similarly, Marelli, Dinu, Zamparelli, and Baroni (in press) have calculated the semantic transparency of English compounds separately for concatenated and spaced variants, and observed that the measures based on spaced compounds are better predictors of lexical decision latencies to printed compounds: again, the authors only considered lexical decisions to concatenated compounds.

The present study aims to examine how individual differences in reading experience and how the lexical properties of compound words influence the way in which compounds with spelling variants of different probabilities are read. The documented variability in spelling biases allowed us to present compounds in the formats that were fully, partially or not supported by readers' experience with those words. We expected to replicate the early processing advantage and later cost found for spaced compounds in previous studies (Bertram, Kuperman, Baayen & Hyönä, 2011; Cherng, 2008; Inhoff, et al., 2000; Juhasz et al., 2005). With respect to the spelling bias, we expected that words presented in a more supported spelling format would be easier to read overall. Furthermore, we expected that the strength of this effect of probability would be modulated by the frequency of occurrence of the word in question, as well as by an individual's amount of reading experience. Compounds that occur more frequently in any spelling variant provide more exposure to all spelling variants (proportionally to the variant's probability), and provide more opportunity for readers to learn what orthographic alternatives are preferentially associated with a given lexical meaning. Additionally, more experienced readers will have had greater exposure to any and all words, including the critical compounds, as well as complex words that belong to the same morphological family as critical compounds, i.e. share a morpheme with these compounds (e.g. *girlfriend*, *girl scout*, *boyfriend*). Spelling preferences among family members have a strong analogical influence on the orthographic choice in a given compound (Kuperman & Bertram, 2013), and so increased familiarity with whether and how the compound's family members alternate in spelling would contribute to individuals' sensitivity to the distributional bias of that given word. That is, superior exposure to print gives experienced readers more opportunity to include subtle orthographic information such as the compounds' spelling biases into their lexical representations[3]. In sum, we hypothesized that the effect of the probabilistic bias towards the presented

[3] An ideal measure would be overall frequency of occurrence of a compound, and frequencies of specific variants in an individual's mental lexicon. Since these are unavailable, we resort to corpus frequencies as an index of predicted exposure to the word.

spelling of compound words would be stronger in more frequent compounds, and also stronger in individuals with more reading experience.

2.4 Method

2.4.1 Stimuli

Compound words were selected to represent the entire range of biases towards spaced (*girl friend*) or concatenated (*girlfriend*) spelling. For simplicity, we only included compound that alternated between these two variants, so compounds that occurred as hyphenated (*girl-friend*) were excluded. Bias was defined as the proportion of occurrences of the compound in a given format compared to the total occurrences of the compound and ranged from 0 (the compound never occurs in this format) to 1 (the compound always occurs in this format). The estimates of the biases were taken from a corpus study by Kuperman and Bertram (2013) based on the orthographic forms of the compounds in the 1.2-billion token Wikipedia corpus. Compounds that had a frequency of less than 60 occurrences in the Wikipedia corpus were excluded from our sample. The compound words that met the criteria above were then divided into four bins based on spacing bias so that the selected compounds would have evenly distributed biases. Between twenty eight and thirty two compounds were then selected from each of these bins, for a total of 120 compounds. The bias towards spacing of the resulting compounds – i.e., the probability of the compound to appear in the spaced format – ranged from 0.0031 to 0.9969: the bias towards concatenation was calculated as one minus the bias towards spacing. Each constituent of the compound word was only used once in the stimulus set to avoid repetition priming.

Sentence frames were created for each compound word, such that the compound was not one of the first or last two words in the sentence. The target words were always preceded by a neutral context: see *Predictability of Compounds* below. Additionally, we selected compounds such that their first constituents (*tree* in *tree house*) were not plausible continuations of the preceding sentence fragment: see *Plausibility of First Constituent* below. For example, in the sentence: *The carpenter built a tree house for his children to play in*, the first constituent, *tree,* is not a plausible continuation, while *tree house* is. The restriction on having less-plausible first constituents was imposed because previous research has shown that plausibility judgements are performed by readers as early as the first constituent of a noun-noun compound (Staub et al., 2007), thus affecting the integration of words into sentence and the potential syntactic re-analysis of the sentence (see also Cutter, Drieghe, & Liversedge, in press). We obtained ratings of the plausibility of the sentence fragments including the first constituent of the compound, and separately for the fragments including the whole compound. These ratings were subsequently included in the models to account for any differences in plausibility. In addition, all sentences had simple structure in order to reduce syntactic influences on processing difficulty.

Two sentences were made for each target word using the sentence frames described above: see Appendix A for stimuli list. One sentence frame contained a spaced compound word, and the other contained a concatenated compound word: the sentence pair was otherwise identical. Each participant only saw one spelling variant of each compound. This was achieved by creating two lists, one containing spaced variants for half of the compounds and concatenated for the other half, and the other list with the reverse. Half of the compounds from each bin were used in each list, so compounds with different biases were represented in both lists. The two lists were created in order to compare the processing of compound words in both their more or less probabilistically supported orthographic presentations.

2.4.2 The eye-tracking study

Participants. Twenty nine undergraduate students (27 female, 18-23 year old, mean age of 20) from McMaster University completed the eye-tracking study for course credit. All participants were native speakers of English. All participants had normal or corrected-to-normal vision, and did not report a diagnosed reading or learning disability.

Procedure. Participants first completed the offline tests (see below), and then proceeded to the eye-tracking experiment. Participants were seated approximately 60 cm from the computer. The sentences were displayed on a 17 inch monitor with a resolution of 1600 x1200 pixels, and a refresh rate of 60 Hz. Eye movements during sentence reading were recorded with an Eyelink 1000 desk-mounted eye tracker (SR Research, Kanata, Ontario, Canada). The data were collected at a 1000 Hz sampling rate from the participants' dominant eye, or the right eye if the dominant eye was not known. Sentences were presented one at a time in a Courier New, a monospace font, size 20, in black on a white background, and occupied exactly one line on the screen. Each character subtended 0.36 degree of visual angle. A three-point horizontal calibration of the eye tracker and a three-point horizontal accuracy test were performed before the beginning of each experiment, and after any breaks.

The experiment began with a practice block, consisting of ten sentences, in order to familiarize participants with the experiment. Then participants read sentences containing the target compound words presented as spaced or concatenated. Participants were instructed to press a button when they had finished reading the sentence, and the sentences remained on the screen until the button was pressed. Participants read 120 target sentences and 67 fillers, which served as target sentences for a separate experiment. Each sentence trial was preceded by a drift correction, which used a fixation point 20 pixels to the left of the beginning of the sentence, in order to ensure accurate recording of eye movements. Sentences were presented 100 pixels away from the left edge of the screen, and in the middle of the vertical dimension of the screen. Sentences were randomized such that no more than two sentences from the same probability bin appeared sequentially. Comprehension questions followed 20% of target sentences. Participants were shown the sentences and were asked to respond whether they were true or false. Participants pressed the **a** key if the sentence was true and the ' (single quote) key if it was false. 50% of the correct answers were true, and 50% were false.

Dependent Variables. The dependent variables examined were: single fixation duration (the duration of the first and only fixation on the compound), first-of-many fixation duration (duration of the first fixation when other fixations on the compound followed in the first reading pass), refixation probability within the first pass, second fixation duration, gaze duration (the sum of all fixations before leaving the compound for the first time), and total fixation time (the sum of all fixations on the compound). The eye-movement measures were calculated for target words defined either as the entire compound word for the concatenated presentation (*girlfriend*), or as the entire spaced compound, including the space separating its constituents (*girl friend*).

The eye-movement record enables a fine-grained analysis of the time-course of word processing, with first-of-many fixation duration as the initial measure of word decoding and lexical access, second fixation duration and refixation probability as indices of subsequent processing stages, and single fixation duration, gaze duration and total reading time as indices of the cumulative processing effort during the first pass or all passes on the word.

Independent variables. The predictors included individual scores on the skill tests (described below), as well as compound frequency, length, and spelling bias. The frequencies were taken from the Wikipedia-based sample of compounds (Kuperman & Bertram, 2013) and represented the combined frequency of the compounds in *both* spaced and concatenated formats. The length of the target words in characters included the space separating the two constituents of spaced compounds. All spaced compounds were therefore one character longer than their concatenated counterparts. Details of calculating the spelling bias are presented in the Stimuli section. For simplicity, we will refer to “bias” as the bias towards the spelling in which the compound word was actually presented in the sentence: the bias is either the compound’s bias towards spaced presentation or its inverse. Additional predictors were derived from norming studies.

2.4.3 Norming Studies

Predictability of Compounds: 19 undergraduate students from McMaster University completed a study of the cloze predictability of the stimuli. All participants were native speakers of English, and did not take part in any other experiments reported here. Participants were presented with the sentence frames prior to the first constituent of the compound word, and were asked to provide the next word. The cloze predictability of the compound words was calculated by taking a proportion of the responses that matched the target word compared to the total number of responses. Of the 120 sentences, 102 had cloze predictability of zero and another 6 had predictability above zero and below 10%. The consistently low predictability ratings were therefore not considered in the models.

Plausibility of first constituent: An additional 19 undergraduate students from McMaster University completed a study of the plausibility of the sentences up to and including the first constituent of the compound word. All participants were native speakers of English, and none participated in other experiments reported here. Participants were given a scale from 1 to 7 with 1 being completely implausible and 7 being completely plausible. An average plausibility rating was calculated for each compound (range: 1.364-6.727; mean: 3.846; standard deviation: 1.477). These ratings did not produce a significant effect in any of the models, and therefore are not further discussed.

Plausibility of whole compound: 21 undergraduate students from McMaster University completed a study of the plausibility of the sentences including the whole compound word. All participants were native speakers of English, and none participated in other experiments reported here. The compound words were presented in their more supported format. Participants were given a scale from 1 to 7 with 1 being completely implausible and 7 being completely plausible. An average plausibility rating was calculated for each compound (range: 2.263-6.857; mean: 5.665; standard deviation: 0.790) and these ratings were included in the regressions models.

2.4.4 Skill tests

Tests of orthographic segmentation, vocabulary size and reading experience were conducted in order to assess individual variability in reading proficiency and experience.

Segmentation Task. The text for the segmentation task consisted of a passage of text from a Canadian newspaper, The Globe and Mail, with the spaces, punctuation and capitalization removed. The original article was determined to have been written at a grade 9 level, as indicated by the Flesch-Kincaid Grade Level estimated via Microsoft Word 2007 (Flesch, 1948; Kincaid, Fishburne, Rogers & Chissom,1975). Participants were given 60 seconds to segment the text into words by drawing a line in between each word boundary. This test of lexical decomposition was selected in place of an overt test of morphological decomposition, since the latter would have revealed too much about the experimental manipulation.

Vocabulary Size Test. The Vocabulary Size Test gives an indication of the size of individuals' vocabularies, or more specifically their written receptive vocabulary (Nation & Beglar, 2007). This test consists of multiple choice questions with four choices. Each question contains a word in context, and four optional definitions of that word. Participants were asked to select the correct definition of the word. Words decreased in frequency as the test progressed, such that each thousand lemmas in the hypothesized reader's vocabulary was represented by 10 words from that frequency level. Participants completed the test from the 7th 1000 to the 14th 1000, for a total of 80 questions. Participants received a score out of 80 points corresponding to the number of correct answers.

ART/MRT. The Author Recognition Test (ART) and the Magazine Recognition Test (MRT) were used to assess amount of reading experience of participants (Acheson, Wells, & MacDonald, 2008). These tests provide participants with a list of authors or magazines. The two lists contain 50% names of authors or titles of magazines and 50% distracters in the form of non-author names or non-existing magazine titles. Participants were instructed to indicate the names and titles that they were certain were authors, or magazines respectively. The score for these tests was the number of correctly identified authors or magazines minus the number of incorrectly identified authors or magazines.

2.4.5 Statistical Considerations

All continuous predictors were scaled (z-transformed) to allow the predictors to be compared on the same scale: compound frequency was additionally log-transformed prior to the z-transformation. Continuous dependent variables were also log-transformed to attenuate the influence of outliers, as indicated by the Box-Cox power transformation (Box & Cox, 1982). The plots presented below depict back-transformed values of dependent variables (in ms) to ensure interpretability. Table 1 reports descriptive statistics for all dependent and independent variables (before and after transformation).

Mixed-effects multiple regression models were used for this study with participant and word as random effects (Baayen, 2008; Baayen, Davidson & Bates, 2008; Pinheiro & Bates, 2000): package lme4 v 1.1-6 (Bates, Maechler, Bolker , & Walker, 2013) in the R statistical software 3.1.0 (R Core Team, 2014) was used. Only the fixed effects that reached the 5% significance level are reported below, unless stated otherwise. While the full random effect structure was tested, only those random effects were retained which significantly improved the performance of the models. An improvement was indicated by a significantly higher log likelihood estimate of the model when a given random effect was included, compared to when that random effect was not included (all ps < 0.05 using likelihood ratio tests). After fitting a model, we removed outliers if they were outside of the range of -3.0 to 3.0 units of standard deviation away from the residual error of the model: the model was then refitted to a trimmed data set. No model showed a large degree of collinearity, as indicated by medium condition numbers below or equal 14.6. Nonlinearities were explored for all predictors and, where warranted by the increase in the model performance, modeled with the restricted cubic splines function with three knots. The body of the paper reports regression coefficients for simple main effects and interactive terms if predictors in question entered into an interaction, and regression coefficients for main effects of predictors if no interaction was observed. For regression models fitted to continuous dependent variables, we report t-values rather than p-values, as an accurate estimation of the degrees of freedom in mixed-effects models required for calculating p-values is still debated in statistical literature (for an early treatment cf. Bates, 2006). Roughly, $|t| > 2$ is $p < 0.05$.

TABLE 1 APPROXIMATELY HERE

2.5 Results and Discussion

The original data set contained 3465 data points. Trials in which the compound word was not fixated on were removed (38 data points, 1.1%). All trials in which the compound was skipped in the first pass and subsequently fixated on were removed (17 data points, 0.5%). Trials in which the first fixation was shorter than 50 ms were removed (13 data points, 0.4%), as were trials in which compounds were fixated six or more times (23 data points, 0.7%). Distributional outliers were trimmed based on individual participant data. Data points that were greater than three standard deviations away from the participant mean for total reading time were removed (41 data points, 1.2%). The resulting data pool contained 3333 data points. All participants answered 90% or more of the comprehension questions correctly, so no participants were excluded.

Table 2 summarizes effects of the critical predictor (the probabilistic bias towards the presented spelling) and the critical interactions of bias by reading experience (ART score) and bias by compound frequency, as estimated by the regression models fitted to eye-movement measures. Main effects (or simple main effects when part of an interaction) of the presentation format (spaced versus concatenated), reading experience and joint compound frequency are reported as well. Table 2 further reports sample sizes before and after data trimming. Regression models are reported in full in the Appendix B. In what follows we group the findings by the type of predictors.

TABLE 2 APPROXIMATELY HERE

2.5.1 Critical effects

Presented spelling format. All compounds in our data set were presented to readers both as spaced and as concatenated, though each reader saw each compound in only one format. As length was a control predictor in all our models, the effect of presentation format was estimated over and above the effect of one extra character in spaced vs concatenated compounds. Spaced compounds elicited shorter single fixation durations [b= 0.069, SE=0.016, t = 4.39] and second fixation durations [b = 0.092, SE = 0.032, t = 2.87], however they also were more likely to elicit a second fixation [b = -0.741, SE = 0.086, p = < 0.0001]. As a result, spaced compounds came with longer gaze durations [b = -0.051, SE =0.024, t = -2.15] and total reading times [b=-0.035, SE=0.017, t = -2.11]. This pattern is in line with prior findings that spacing facilitates early morphological decomposition into constituents but incurs a cost at later stages, when the meanings of the constituents have to be integrated into a unified semantic representation (Bertram et al., 2011; Cherng, 2008; Inhoff et al., 2000; Juhasz et al., 2005).

Bias towards presented spelling. We expected compounds presented in their more frequently occurring spelling format (spaced or concatenated) to be recognized faster. This expectation was confirmed. An effect of spelling bias was seen at the earliest in gaze duration and also in total reading time. Specifically, spelling bias enters into interactions with ART score, a measure of overall reading experience, and with word frequency, a measure of experience with a particular compound word.

Measures of skill/experience.

Vocabulary Size. Individuals with higher scores on the Vocabulary Size Test showed a processing advantage compared to individuals with lower scores. Individuals with higher scores were less likely to refixate on the target words [b = -0.464, SE = 0.138, p = 0.0008], had shorter second fixations [b = -0.086, SE =

0.034, t = -2.51], had shorter gaze durations [b = -0.142, SE = 0.032, t = -4.51], and had shorter total reading times [b = -0.164, SE = 0.031, t = -5.25]. The estimates were obtained from models that contained all critical and control predictors previously discussed in the Methods (full models not reported, but see Appendix B for similar models containing ART as predictor).

Segmentation. Individuals with higher segmentation scores showed a processing advantage relative to individuals with lower scores. Higher-scoring individuals were less likely to refixate on the target words [b = -0.500, SE = 0.136, p = 0.0003], had shorter second fixations [b = -0.077, SE = 0.034, t = -2.27], had shorter gaze durations [b = -0.131, SE = 0.034768, t = -3.76], and had shorter total fixation times [b = -0.156, SE = 0.031, t = -4.98]. The estimates were obtained from models that additionally contained all critical and control predictors as described in the Methods

ART/MRT: Scores on the ART measured overall exposure to print, or reading experience. Scores on the MRT did not produce significant effects in any model, and will therefore not be discussed. Individuals with higher scores on the ART showed a relative processing advantage over less experienced readers across the entire eye-movement record. They had shorter single fixations [b = -0.065, SE = 0.021, t = -3.07], second fixations [b = -0.080, SE = 0.037, t = -2.19], gaze durations [b = -0.178, SE = 0.024, t = -7.50], and total reading times [b = -0.139, SE = 0.035, t = -4.01] on the target words. They were also less likely to refixate on the target word [b = -0.63578, SE = 0.11549, $p < 0.0001$]. As outlined below, ART scores also interacted with spelling bias, such that more experienced readers were more affected by the bias, in line with our prediction. As is evident from this section, effects of all measures of overall experience with printed material show highly convergent results. Because ART has been argued to be the most direct measure of exposure to print (Acheson et al., 2008; Stanovich & West, 1989), while vocabulary size has been demonstrated to provide inaccurate estimates in the lower-frequency range (Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014), our further discussion of indices of individual variability concentrates on ART.

Compound frequency. The joint frequency of a compound word in all of its formats serves as a proxy for individuals' experience with that specific word. More frequent compounds elicited shorter single fixation durations [$b = -0.086$, $SE = 0.039$, $t = -2.22$] and gaze durations [$b = -0.032$, $SE = 0.009$, $t = -3.34$]. In addition, compound frequency interacted with spelling bias such that higher-frequency compounds showed stronger bias effects (see below), as predicted.

Interactions between lexical properties and skill tests. Importantly, the effect of bias was modulated by compound frequency and by individuals' scores on the ART. Figure 1 depicts the effect of bias on total reading time plotted per quantiles of the ART scores. Figure 2 depicts the effect of bias towards presented spelling on gaze duration plotted for quantiles of joint compound frequency. Both figures show a partial effect of bias as estimated by linear mixed-effects models and adjusted for values corresponding to 10th, 30th, 50th, 70th, and 90th percentiles of ART score or joint frequency distributions, respectively.

A significant interaction was observed between spelling bias and ART scores for total reading time [b =-0.019, SE = 0.007, t = -2.79]. The processing advantage associated with a higher bias towards the presented spelling was essentially restricted to individuals in the top 30% of the ART range and was stronger the more proficient the reader was (cf. the steeper negative slope of the bias effect in the line denoting the highest ART value, i.e. the 90th percentile of ART scorers). The bias effect was attenuated and negligibly small in readers with less exposure to print (see Figure 1). An analogous interaction was also seen between spelling bias and segmentation scores in total reading time [b = -0.017, SE = 0.007, t = -2.47] and between bias and vocabulary size scores in total reading time [b = -0.013, SE = 0.007, t = -1.80] although the interaction between bias and vocabulary size was marginally significant.

FIGURE 1 APPROXIMATELY HERE

An interaction was also seen between bias and joint compound frequency for gaze duration [b = -0.015, SE =0.006, t = -2.48] and total reading time [b = -0.022, SE =0.006, t = -3.83] (see Figure 2). The facilitatory effect of bias was larger for compounds more frequently attested in either spelling format. Effectively, the bias effect was only observed in compounds in the upper half of the frequency range. As compound frequency increases, the negative slope of the bias effect becomes steeper (see Figure 2). No interactions were observed between bias towards spelling and the presentation format of the compound, nor were there three way interactions between ART, compound frequency and bias (all $|t| < 1.5$).

FIGURE 2 APPROXIMATELY HERE

2.5.2 Control variables

Several other lexical and experimental variables were included in our analyses, as they are known to affect the reading times.

Trial number. As the experiment progressed, readers got faster. Second fixation durations [b = -0.025, SE = 0.007, t = -3.38], gaze durations [b = -0.027, SE =0.007, t = -3.74] and total fixation times [b= -0.001, SE = 0.0002, t = -4.11] on the target compound were shorter, and refixations were less likely to occur [b = -0.082, SE = 0.039, p = 0.034] towards the end of the experiment. A significant interaction between the trial number and ART scores was also observed in first-of-many fixation duration. The overall advantage in compound processing for relatively experienced readers was increasingly weaker as the experiment progressed [b = 0.025, SE = 0.007, t = 3.50], suggesting a training effect in less experienced readers. Trial number did not enter into two-way or three way (trial x ART x bias) interactions with any critical predictor (all |t|'s < 1.5).

Word length. Longer target words elicited longer first-of-many fixation durations [b = -0.014, SE = 0.006, t = -2.15], and gaze durations [b = 0.060, SE = 0.010, t = 5.81], and were more likely to be refixated [b = 0.238, SE = 0.040, p < 0.0001].

Whole compound plausibility. Compound words that were more semantically plausible in the context of the beginning of the sentence elicited shorter second fixation durations [b = -0.035, SE = 0.011, t = -3.26], gaze durations [b = -0.032, SE = 0.012, t = -2.67] and total reading times [b = -0.078, SE = 0.013, t = -5.98].

2.6 General Discussion

One under-tested corollary of Perfetti's (1985; 2007) Lexical Quality hypothesis is that those statistical patterns of language use which can only be acquired through extensive reading practice will preferentially disrupt or benefit comprehension in the most proficient readers and for the most commonly occurring items that exemplify those patterns. It is those individuals and those items that are likely to create sufficient learning opportunities for encoding statistical information in linguistic representations in the brain. The present study probed this hypothesis by considering a specifically orthographic phenomenon: spelling alternation in English compound words. Knowledge of the different probabilities associated with the alternating variants can only be acquired through reading, and, given the degree of uncertainty in the spelling choice for many compounds (cf. Kuperman & Bertram, 2013), requires extensive exposure to these compounds and to similar words. Furthermore, since the meanings of compound words in their spaced and concatenated formats are near-identical, the spelling variants served as their own controls, only differing in their probability of occurring in one format or another. This alternation thus enabled us to isolate the effect of language use (i.e. the probabilistic bias) on language comprehension over and above the effects of a variety of formal, semantic and pragmatic dimensions.

The central finding of this paper is a confirmation that probabilistic biases towards one or another spelling variant have a particularly strong influence on individuals with greater exposure to print, measured via ART scores, and for words that occur most frequently in either format, measured via the joint frequency of occurrence of those words. The facilitatory effect of the bias towards the presented spelling was strongest in readers with higher scores on the ART task in total reading time. This effect was also stronger for compounds with the highest joint frequency of occurrence in gaze duration and total reading time, see Table 1. Less experienced readers were overall less sensitive to subtle, print-specific language statistical information, and all readers were less likely to respond to differences in the distributional patterns of lower-frequency compounds. These lower-frequency compounds may not have been encountered enough times to be sufficiently entrenched in participants' orthographic representations to the extent necessary to discriminate the forms from one another. The effect of the bias towards the presented spelling did not arise at the initial stages of word identification (single fixation durations or first-of-many fixation durations), but was only visible in cumulative measures of the first pass (gaze duration) and of all passes (total reading time), i.e. stages roughly associable with the integration of lexical representations within a larger context (Boston, Hale, Kliegl, Patil & Vasishth, 2008).

Our results also support previous cross-linguistic studies on how the spelling of compound words affects their processing; specifically, an early processing advantage, and later cost for compound words presented as spaced (Inhoff, Radach & Heller, 2000; Juhasz, Inhoff & Rayner, 2005) or hyphenated (Bertram, Kuperman, Baayen & Hyönä, 2011; Cherng, 2008). Single fixation durations and second fixation durations on spaced compounds were shorter than those on concatenated compounds, however this pattern reversed for later measures. Spaced compounds were more likely to be refixated, and had longer gaze durations and total fixation times. This is consistent with the view that spaces between the constituents of compound words facilitate the identification of the constituents, but that later integration is harmed since the semantic link between the words is no longer immediately clear.

In the remainder of this section, we discuss the critical effect of the bias towards the presented spelling and the interactions it enters into. It is quite uncommon to observe, as we have, a situation in which individuals with more exposure to print, and thus higher quality lexical representations, are influenced by a given distributional pattern, whereas individuals with less exposure are less so or not at all. In most previous studies on individual differences that we are aware of, individuals with lower proficiency make more use of, and are more strongly affected by the distributional patterns of language than those individuals with greater proficiency. Rather than having little difference between processing times for words with high and low frequency or predictability, as seen in previous studies (Adelman, Sabatos-DeVito, Marquis & Estes, 2014; Ashby, Rayner & Clifton, 2005; Butler & Hains, 1979; Chateau & Jared, 2000; Hawelka et al., 2010; Hersch & Andrews, 2012; Jared, Levy & Rayner, 1999; Kuperman & Van Dyke, 2011a, 2011b, 2013; Sears, Siakaluk, Chow, & Buchanan, 2008; Yap, Balota, Sibley, & Ratcliff, 2012, but see Whitford & Titone, 2014), individuals with more exposure to print showed a larger difference in reading time for high-and low-bias words compared to less experienced readers in our study. The results thus dovetail well with the framework of the Lexical Quality hypothesis (Perfetti, 1985, 2007) and its predictions. As individuals gain increased exposure to print, and to specific, highly frequent words, their lexical representations increase in quality. Our results suggest that information about the amount of support that a spelling alternative receives from distributional patterns in natural language may factor into individuals' lexical representations by way of differential exposure to compounds in their spaced or concatenated forms. Individuals' representations of a word like *grapevine* (bias towards concatenation: 0.82) are more likely to support the concatenated presentation, than the spaced presentation, and thus individuals may be facilitated when they read the word as concatenated, and potentially harmed when they read it at as spaced. Without exposure to enough instances of a word, or without sufficient exposure to print, however, individuals' representations will likely not be sufficiently specified to include this information.

Interestingly, our data make a compelling case that the mental representation of a word must encompass *all* orthographic representations of that word. Relative frequencies of the alternatives impact compound word recognition in a gradient way, which is proportional to the amount of support towards a presented spelling in written language. This notion converges with recent demonstrations that the mental lexicon simultaneously stores multiple pronunciation variants of a word, including a full phonological representation and several acoustically reduced forms (for a review, cf. Ernestus, 2014). While exploring this possibility is beyond the scope of this paper, our finding may have implications for research on the mental storage of common misspellings (*receive* and *recieve*) and orthographic reductions (*kind of* and *kinda*).

A limitation of this study is that the participant pool consisted only of undergraduate students, who are expected to be fairly proficient readers. Although there was a sizeable amount of variation in participants' scores, our participants were likely younger and more proficient than the general population. Typically, a future direction would be to look to individuals with low proficiency, and consequently less experience, as well as to clinical populations. However, given our findings, future studies may want to test individuals on the other end of the spectrum. Individuals with very extensive exposure to print, such as copy editors or English professors, may show even larger effects. We may see a larger benefit for more supported compounds, and more of a cost for less supported compounds.

Our finding that sensitivity to the distributional bias in compound spelling is primarily confined to proficient readers and frequent words raises a methodological question. Given that most psycholinguistic research on non-clinical adult populations, including this study, uses convenience pools of undergraduate students, it is possible that some of the apparently robust probabilistic effects on processing would not be confirmed should the studies be re-run in a sample of the less proficient population at-large. Logically, this caveat would particularly apply to patterns that are prevalent in written, rather than spoken, language, or occur so rarely in naturalistic use that accumulation of necessary statistical information would require years of intensive skilled reading. The corpus analysis by Roland, Dick and Elman (2007) points to a variety of syntactic constructions that fit the bill, i.e. are infrequent in use and used predominantly in written language. Thus, in line with the literature on individual differences in word and sentence processing (Adelman et al., 2014; Daneman & Carpenter, 1980), it is plausible that theories of language comprehension would benefit from a revision of their empirical base against a broader variety of individual skills, both higher and lower than those normally found in university convenience pools.

2.7 References

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, *40*(1), 278-289.

Adelman, J. S., Sabatos-DeVito, M. G., Marquis, S. J., & Estes, Z. (2014). Individual differences in reading aloud: A mega-study, item effects, and some models. *Cognitive psychology*, *68*, 113-160.

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1065-1086.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* New York: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological review*, *118*(3), 438.

Bates, D. (2006, May 19). *[R] lmer, p-values and all that*. Retrieved from: https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.0-4. Accessed online: December 2013.

Bertram, R., Kuperman, V., Baayen, R. H., & Hyönä, J. (2011). The hyphen as a segmentation cue in triconstituent compound processing: It's getting better all the time. *Scandinavian journal of psychology*, *52(6)*, 530-544.

Boston, M., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, *2*(1), 1-12.

Box, G. E. P., & Cox, D. R. (1982). An analysis of transformations revisited, rebutted. *Journal of the American Statistical Association*, *77*(377), 209-210.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation.
In G. Boume, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69-94). Amsterdam: Royal Netherlands Academy of Science.

Bresnan, J., & Ford, M. (2010). Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language*, *86*, 168-213.

Buswell, G. T. (1922). *Fundamental reading habits, a study of their development*. Chicago: Chicago University Press.

Butler, B., & Hains, S. (1979). Individual differences in word recognition latency. *Memory & Cognition*, *7*(2), 68-76.

Chace, K. H., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview: effects of reading skill. *Canadian Journal of Experimental Psychology*, *59*(3), 209-217.

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition*, *28*(1), 143-153.

Cherng, M. (2008). *The role of hyphenation in English compound word processing* (Unpublished bachelor's thesis). Wesleyan College, Middletown, CT.

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2014). Preview benefit in English spaced compounds. *Journal of Experimental Psychology Learning Memory and Cognition* (In Press).

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, *19*(4), 450-466.

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, *142*(2), 27-41.

Flesch, R. (1948). A new readability yardstick. *Journal of applied psychology*, *32*(3), 221-233.

Gries, S. (2003). *Multifactorial analysis in corpus linguistics: A study of particle placement*.
New York: Continuum International Publishing Group Ltd..

Hawelka, S., Gagl, B., & Wimmer, H. (2010). A dual-route perspective on eye movements of dyslexic readers. *Cognition, 115*(3), 367-379.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological approach*. New York: John Wiley & Sons.

Hersch, J., & Andrews, S. (2012). Lexical Quality and Reading Skill: Bottom-Up and Top-Down Contributions to Sentence Processing. *Scientific Studies of Reading,* 16(3), 240-262.

Inhoff, A. W., Radach, R., & Heller, D. (2000). Complex compounds in German: Interword spaces facilitate segmentation but hinder assignment of meaning. *Journal of Memory and Language, 42*(1), 23-50.

Jaeger, T. F., & Tily, H. (2011). On language 'utility': Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science, 2(*3), 323-335.

Jared, D., Levy, B. A., & Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: evidence from proofreading and eye movements. *Journal of Experimental Psychology: General, 128*(3), 219-264.

Juhasz, B. J., Inhoff, A. W., & Rayner, K. (2005). The role of interword spaces in the processing of English compound words. *Language and cognitive processes, 20*(1-2), 291-316.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay & and S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 39-96). Massachusetts: MIT Press.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typological studies in language, 45*, 229-254.

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (Automated readability index, Fog count and Flesch reading ease formula) for navy enlisted personnel*. Millington: Naval Technical Training Command.

Kuperman, V. & Bertram, R. (2013). Moving spaces: Spelling alternation in English noun-noun compounds. *Language and Cognitive Processes, 28*(7), 939-966.

Kuperman, V., & Bresnan, J. (2012). The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language*, *66*(4), 588-611.

Kuperman, V., Van Dyke, J.A. (2011a). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language, 65*(1), 45-73.

Kuperman, V. and Van Dyke, J.A. (2011). Individual differences in visual comprehension of morphological complexity. In L. Carlson, C. Hoelscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 1643-1648). Austin, TX: Cognitive Science Society.

Kuperman, V. & Van Dyke, J.A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance, 39*(3), 802 – 823.

Levy, B. A., Abello, B., & Lysynchuk, L. (1997). Transfer from word training to reading in context: Gains in reading fluency and comprehension. *Learning Disability Quarterly, 20*(3), 173-188.

Levy, B. A., Nicholls, A., & Kohen, D. (1993). Repeated readings: Process benefits for good and poor readers. *Journal of Experimental Child Psychology*, *56*(3), 303-327.

Lohse, B., Hawkins, J., & Wasow, T. (2004). Domain minimization in English verb-particle constructions. *Language, 80*(2), 238-261.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*(1), 35-54.

Marelli, M., Dinu, G., Zamparelli, R., & Baroni, M. (in press). Picking buttercups and eating butter cups: Spelling alternations, semantic relatedness and their consequences for compound processing. *Applied Psycholinguistics*.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Perfetti, C. A. (1985). Reading ability. New York: Oxford University Press.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy*, *11*, 67-86.

Perfetti, C. A., (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading*, *11*(4), 357-383.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.

R Core Team. (2014). *A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing: Vienna, Austria. Accessed at: http:/www.R-project.org

Radach, R., & Kennedy, A. (2013). Eye movements in reading: Some theoretical context. *The Quarterly Journal of Experimental Psychology*, *66*(3), 429-452.

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in cognitive science*, *6*(1), 5-42.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton Jr, C. (2012). *Psychology of reading*. New York: Psychology Press.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: a comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(5), 1188.

Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic bulletin & review*, *17*(6), 834-839.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. Prokasy (Eds.)
Classical conditioning II: Current research and theory (pp. 64-99). New York: Appleton-Century-Crofts.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language, 57*(3), 348-379.

Saffran, J. R. (2003). Statistical language learning mechanisms and constraints. *Current directions in psychological science*, *12*(4), 110-114.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of memory and language*, *35*(4), 606-621.

Sears, C. R., Siakaluk, P. D., Chow, V. C., & Buchanan, L. (2008). Is there an effect of print exposure on the word frequency effect and the neighborhood size effect?. *Journal of Psycholinguistic Research*, *37*(4), 269-291.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive science*, *23*(4), 569-588.

Sepp, M. (2006). *Phonological constraints and free variation in compounding: A corpus study of*
English and Estonian noun compounds. Unpublished doctoral dissertation, City University of New York, NY.

Shie, J. S. (2002). English hyphenated compounds. *Journal of the Da-Jeh University*, *11*, 89-98.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*(4), 402-433.

Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1162-1169.

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*, *1*(2), 147-165.

Tily, H., Hemforth, B., Arnon, I., Shuval, N., Snider, N., & Wasow, T. (2008). Eye movements reflect comprehenders' knowledge of syntactic structure probability. In *The 14th Annual Conference on Architectures and Mechanisms for Language Processing*.

Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. *The Quarterly Journal of Experimental Psychology*, *67*(4), 703-727.

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*(2), 729-742.

Wasow, T. (2002). *Postverbal Behavior*. Stanford: CSLI.

Whitford, V., & Titone, D. (2014). The effects of reading comprehension and launch site on frequency–predictability interactions during paragraph reading. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1151-1165.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 53.

Tables

	Range	Mean	SD	range of transformed values
Compound Frequency	61, 9698	444.000	1217.466	-1.109, 3.936
Length	8, 14	9.700	1.239	-1.374, 3.469
Bias	0.003, 0.997	0.500	0.299	-1.664, 1.658
ART	0, 40	11.080	8.175	-1.355, 3.538
Vocabulary Size Test	47, 72	58.537	6.449	-1.789, 2.088
Segmentation Test	129, 385	249.142	69.806	-1.721, 1.946
Whole compound plausibility	0.00, 0.63	0.020	0.070	-4.307, 1.509
Trial number	11, 198	104.260	54.133	-1.723, 1.732

Table 1: Descriptive statistics for independent variables before and after transformation. All variables were scaled and joint frequency was log transformed and then scaled.

Eye-movement measure	Spelling format (with spaced as the reference level)	Bias	ART	Joint Frequency	ART x Bias	Bias x Joint Frequency
Single Fixation Duration N = 1420 Ntrimmed = 1404	b= 0.069 SE=0.016 t = 4.39 **Spaced compounds are read faster.**	NS	b = -0.065 SE = 0.021 t = -3.07 **Better readers are faster.**	b = -0.086 SE = 0.039 t = -2.22 (restricted cubic splines term 2) **More frequent words are read faster.**	NS	NS
Refixation Probability N = 3333 Ntrimmed = 3324	b = -0.741 SE = 0.086 p < 0.001 **Refixation is more likely for spaced compounds.**	NS	b = -0.636 SE = 0.115 p < 0.001 **Refixation is less likely for better readers.**	NS	NS	NS
Second Fixation Duration N = 1913 Ntrimmed = 1883	b = 0.092 SE = 0.032 t = 2.87 **Spaced compounds are read faster**	NS	b = -0.080 SE = 0.037 t = -2.19 **Better readers are faster.**	NS	NS	NS
Gaze Duration N = 3333 Ntrimmed = 3315	b = -0.051 SE =0.024 t = -2.15 **Spaced compounds are read slower.**	b = -0.012 SE = 0.008 t = -1.48	b = -0.178 SE = 0.024 t = -7.50 **Better readers are faster.**	b = -0.032 SE = 0.009 t = -3.34 **More frequent words are read faster.**	NS	b = -0.015 SE =0.006 t = -2.48 **Greater effect of bias for more frequent words.**
Total Reading Time N = 3333 Ntrimmed = 3316	b=-0.035 SE=0.017 t = -2.11 **Spaced compounds are read slower.**	b = -0.008 SE = 0.008 t = -1.03	b = -0.139 SE = 0.035 t = -4.01 **Better readers are faster.**	b= -0.042 SE = 0.010 t = -4.03 **More frequent words are read faster.**	b =-0.019 SE = 0.007 t = -2.79 **Better readers are more affected by bias.**	b = -0.022 SE =0.006 t = -3.83 **Greater effect of bias for more frequent words.**

Table 2: Summary of the critical effects for each eye-movement measure. The regression coefficient estimate, the standard error, and the t-value ($|t| > 2.0$ roughly translates into $p < 0.05$) are listed for each critical main effect and interaction. NS – not significant at the 5% threshold. In cases of interactions, we reported both simple main effects (e.g. Bias and Joint Frequency), and the interactive terms (Bias x Joint Frequency).

Figures

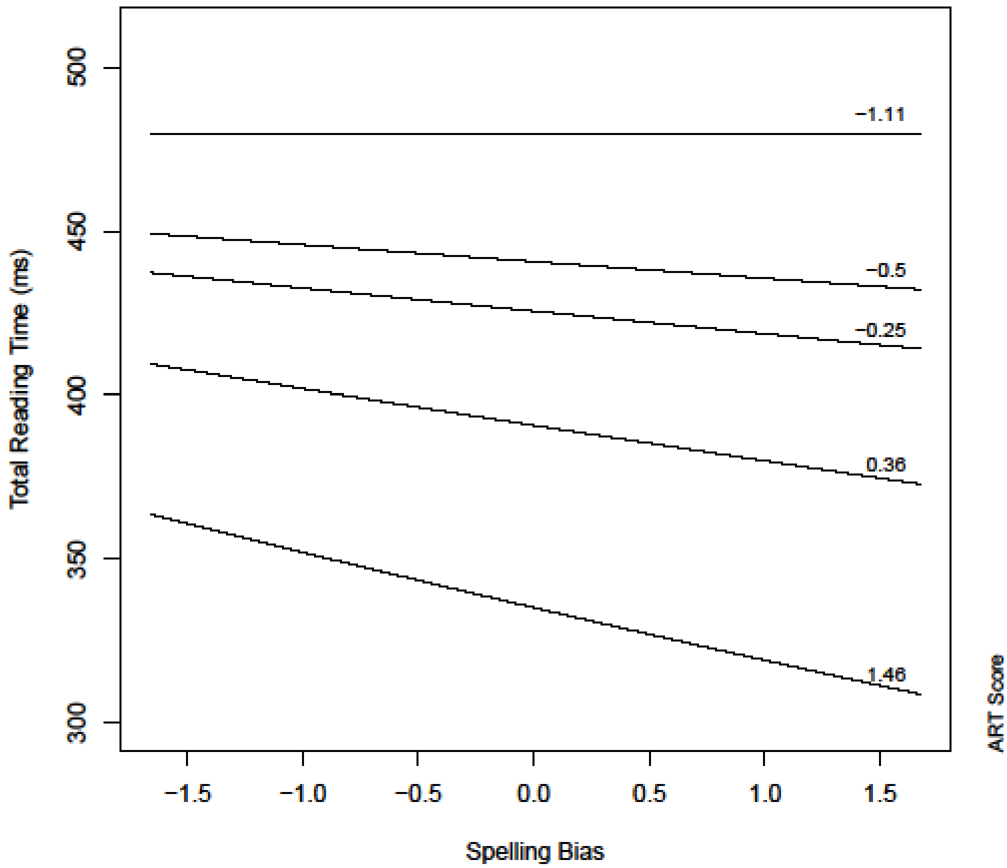

Figure 1: Partial model-estimated effects of the probabilistic bias towards the presented format on total reading time, broken down by percentiles of ART scores. Values of ART are shown on the right edge and stand for the 10th, 30th, 50th, 70th and 90th percentiles of ART scores.

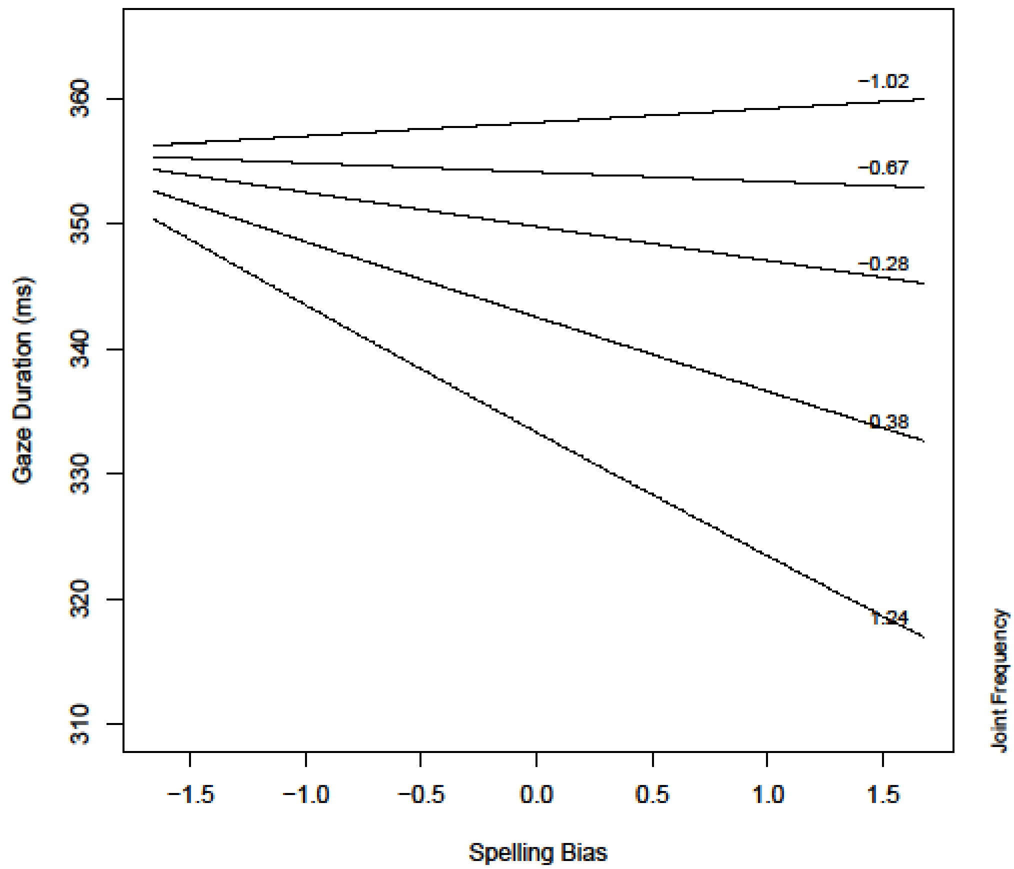

Figure 2: Partial effects of the probabilistic bias towards the presented format on gaze duration, broken down by percentiles of joint frequency. Values of joint frequency are shown on the right edge and stand for the 10th, 30th, 50th, 70th and 90th percentiles of joint frequency.

Appendix A - List of Stimuli

Each stimulus sentence is presented with a target compound in a spaced format, followed by an estimate of the compound's bias towards spacing (i.e. the number of spaced occurrences divided by the total number of the compound's occurrences as spaced or concatenated). The estimates are based on the Wikipedia corpus reported by Kuperman and Bertram (2013).

Gerald carried the **pocket watch** that he got for his birthday. 0.81
The agent interviewed the **cover girl** for an upcoming magazine. 0.85
The employees used a **mine shaft** to reach the bottom of the mine. 0.67
The scientist discovered a new **brain wave** and published a paper about it. 0.59
Lila took the **song list** from the front of the stage. 0.76
Jeremy added a **bank note** to his collection of old bills. 0.15
Melissa made the **milk shake** in her new blender. 0.12
The baby had **club foot** so he needed many surgeries. 0.5
Heather crafted a **bread basket** to give to her mother. 0.19
The doctor determined the **blood type** of his patient. 0.96
Ethan surveyed the **coal field** for geological markers. 0.13
Max recorded a **demo tape** to give to local producers. 0.97
The voters selected a **council member** who had a lot of experience. 0.88
Brian followed the **rock slide** down the side of the mountain. 0.45
Susan bought the **paint brush** for her upcoming art class. 0.23
Lauren looked through the **photo book** with her children. 0.4
Tim focused on one **body part** when he worked out at the gym. 0.95
The team competed in a **quiz bowl** at the provincial level. 0.56
The boys visited the new **skate park** every day after school. 0.57
Cathy sent her son to a **boot camp** because of his poor marks in school. 0.93
The kids had their **play time** after they ate their snack. 0.5
The general studied the **battle fleet** that would be deployed soon. 0.55
Erica attached the **drain pipe** to a bucket so she could collect the rain water. 0.46
Shelby rode the **chair lift** to the top of the hill. 0.18
Sarah painted a **bird cage** that she found in her basement. 0.43
Barbara supplied a **fact sheet** to the group she was leading. 0.65
Allen patrolled a **cell block** as part of his job at the jail. 0.71
Scott installed a **sound card** in his new computer. 0.45
Adam fixed the **mouth piece** of the old telephone. 0.03
The nurse measured the **birth weight** of the new baby. 0.93
Liz added **passion fruit** to enhance the flavour of her souffle. 0.68
Ashley downloaded a new **ring tone** for her phone. 0.29
The manager hired a **stage hand** to help with the upcoming play. 0.27
Joe discovered the **rain shadow** on one side of a mountain. 0.87

Bob grew the **silk worm** for a science project. 0.11
Dustin cleaned the **shot glass** after the bar was closed. 0.73
Amy joined the **ball game** between two local teams. 0.5
Corey begged the **loan shark** for money to pay his bills. 0.77
Sue used the **paper clip** to keep her files together. 0.73
Mark set a **mouse trap** in his apartment. 0.15
Monica sprinkled **corn meal** on the pan before baking her pizza. 0.34
Amanda visited the **sink hole** that had appeared in town. 0.09
The children played in the **court yard** while their parents ate lunch. 0.01
Blake wore a **face mask** when he played hockey. 0.64
The men played until **match point** but then it started to rain. 0.76
Stacy cleaned the **lunch room** because it was messy. 0.3
The knight wore a **chest plate** during the battle. 0.38
Jenna sold her watch to a **pawn shop** because it no longer fit her. 0.61
Derek wore **chain mail** when he played the king for the play at school. 0.52
Hannah remained close with her **school friend** for many years. 0.86
Ben used a **blow torch** to weld the leaky pipes. 0.27
The surgeon cut the **breast bone** in order to operate on his patient. 0.3
Anna stretched the **sheep skin** to make a blanket. 0.1
Richard lives in a **border town** so many of his friends are from the United States. 0.97
The police officer searched the **data base** for the DNA of a suspect. 0.01
Eric stood in the **band shell** and imagined he was playing for an audience. 0.32
Jack called his **class mate** who he had not seen since university. 0.01
The student consulted the **message board** for help with her calculus exercises. 0.88
Amber contacted the **coast guard** after she saw a ship hit the rocks. 0.57
The presentation set a **bench mark** for all others that followed it. 0.02
The cats sat on the **window sill** to watch the birds fly by. 0.4
The contractors created the **waste pile** when they demolished the house. 0.45
The detective took a **finger print** from a mug that the suspect had touched. 0.03
Josh saw the **weather vane** on top of the barn. 0.58
Taylor inspected the **fault line** that ran across the desert. 0.83
Alexis worked at the **help desk** at the local library. 0.65
Paul climbed the **flag pole** outside of his old school. 0.21
Andrea monitored the **heart beat** of a patient in the clinic. 0.18
The mayor ordered a **stop light** for a busy intersection. 0.31
The manager trained the **flight crew** on basic safety measures. 0.95
The men witnessed a **jail break** at the local prison. 0.28
Many women worked on the **home front** during the war. 0.75
Ryan made a **flow chart** to use for his upcoming presentation. 0.34
Jacob cut enough **fuel wood** to last the winter. 0.49
The dog joined a **wolf pack** when it was released into the wild. 0.78

Violet bought the **snow globe** to give to her niece as a souvenir. 0.71
Dylan hit the **goal post** when he was aiming for the net. 0.45
The technician prepared the **film strip** for the new movie. 0.22
Antonio was the **station master** when trains were still popular. 0.54
The teacher took a **head count** of her students at the end of their trip. 0.35
James drew a **floor plan** for the new house. 0.68
Logan explored the **fire hall** while visiting his father at work. 0.74
The troops used the **smoke screen** to hide their entry to the area. 0.52
The physician removed the **gall bladder** of an elderly man. 0.33
Kelly was a **stunt woman** when she was in her twenties. 0.26
Thomas toured the **flour mill** while on vacation. 0.95
Kevin hunted on the **forest land** around his friend's farm. 0.64
George passed through a **toll booth** while driving across the border. 0.75
The carpenter built a **tree house** for his children to play in. 0.32
Julia repaired her **video camera** before she went on vacation. 0.97
The supplies landed in the **drop zone** that had been designated by the charity. 0.91
The volunteers cleared the **flood water** after the hurricane. 0.53
David registered for a **trade show** to see all of the latest cars. 0.86
The divers visited the **wreck site** of the Titanic. 0.85
Tyler collected a **life raft** for everyone on the boat. 0.79
The workers repaired the **stone wall** around the old building. 0.97
Natalie sanded the **door frame** of her new room. 0.79
Courtney attended a **horse race** for the first time last week. 0.98
Aaron felt the **earth quake** that happened two towns away. 0
The farmer harvested the **honey comb** so he could sell the honey. 0.03
The cashier sold a **trench coat** to a tourist who had left his coat at home. 0.57
Abby got **heat stroke** from working outside on a hot day. 0.67
Jennifer reserved a **week night** for doing chores. 0.12
Steve used a **sledge hammer** to demolish the wall. 0.14
Mary crocheted a **table cloth** to match her decorations. 0.25
Sheryl saw a **space craft** that had flown to the moon. 0.01
Rachel wanted a wide **stair case** in her new home. 0.01
Emma grew a **grape vine** in her garden over the summer. 0.18
Andy wore a **track suit** when he went for a run. 0.22
Jared moved the **gear shift** in his new car. 0.53
An albatros has the **wing span** of at least two metres. 0.07
Emily walked across a **draw bridge** that lead to a castle. 0.03
Olivia went to the **drug store** to buy milk and eggs. 0.5
Julian admired the **tile work** at his hotel in Portugal.0.5
Lisa entered her **user name** to access her account. 0.29
Maria's favourite place was the **duck pond** in the middle of her parents' farm. 0.87
Frank got stuck in the **thunder storm** on his way to work. 0.02

The storm caused a **tail wind** so the plane arrived early. 0.31
Justin watched a **sword fight** between two knights on television. 0.55
Peter rode his **sport bike** to the store on Monday. 0.82

Appendix B - Multiple Regression Models

Single Fixation Duration

	Estimate	Standard Error	t value
Intercept	5.339	0.028	188.93
ART	-0.065	0.021	-3.07
Bias	-0.004	0.008	-0.53
Spelling Format (concatenated)	0.069	0.016	4.39
Joint frequency, rcs term 1	0.043	0.027	1.60
Joint frequency, rcs term 2	-0.086	0.039	-2.22

Table 3a: Fixed effects of the multiple regression model fitted to single fixation duration. The R^2 of the model is 0.211 and the standard deviation of the residual is 0.280. All numeric predictors were scaled. “Rcs” stands for restricted cubic splines, fitted with 3 knots. The reference level for Spelling Format is spaced.

Random Effect	Standard Deviation
Compound	0.036
Participant	0.107

Table 3b: Random effects of the multiple regression model fitted to single fixation duration, including random intercepts for compound word and participant.

First-of-many fixation duration

	Estimate	Standard Error	t value
Intercept	5.291	0.023	226.13
Trial Number	0.007	0.006	1.20
ART	-0.036	0.023	-1.52
Spelling Format (concatenated)	-0.002	0.018	-0.10
Bias	-0.003	0.006	-0.44
Word Length	-0.014	0.006	-2.15
Whole Compound Plausibility	-0.012	0.006	-1.91
Trial Number x ART	0.025	0.007	3.50

Table 4a: Fixed effects of the multiple regression model fitted to first of many fixation duration. The R^2 of the model is 0.235 and the standard deviation of the residual is 0.246. All numeric predictors were scaled. The reference level for Spelling Format is spaced.

Random Effect	Standard Deviation	Correlations between by-participant slopes and intercepts
Compound	0.017	
Participant	0.118	
Presentation format by participant	0.070	-0.017

Table 4b: Random effects of the multiple regression model fitted to first of many fixation duration, including random intercepts for compound word and participant by presentation format.

Refixation Probability

	Estimate	Standard Error	z value	p value
Intercept	-1.580	0.419	-3.770	<0.001
ART	-0.636	0.115	-5.505	<0.001
Bias	-0.066	0.039	-1.706	0.088
Spelling Format (concatenated)	-0.741	0.086	-8.665	<0.001
Trial Number	-0.082	0.039	-2.118	0.034
Word Length	0.238	0.040	6.031	<0.001

Table 5a: Fixed effects of the logistic multiple regression model fitted to refixation probability. All numeric predictors were scaled. The reference level for Spelling Format is spaced.

Random Effect	Standard Deviation
Compound	0.210
Participant	0.579

Table 5b: Random effects of the multiple regression model fitted to refixation probability, including random intercepts for compound word and participant.

Second Fixation Duration in the First Pass

	Estimate	Standard Error	t value
Intercept	4.601	0.037	125.37
ART	-0.080	0.037	-2.19
Bias	0.011	0.016	0.67
Spelling Format (concatenated)	0.092	0.032	2.87
Joint Frequency	0.011	0.016	0.67

Table 6a: Fixed effects of the multiple regression model fitted to second fixation duration. The R^2 of the model is 0.171 and the standard deviation of the residual is 0.322. All numeric predictors were scaled. The reference level for Spelling Format is spaced.

Random Effect	Standard Deviation
Compound	0.000
Participant	0.162

Table 6b: Random effects of the multiple regression model fitted to second fixation duration, including random intercepts for compound word and participant. Note the lack of inter-compound variability in intercepts.

Gaze Duration

	Estimate	Standard Error	t value
Intercept	5.795	0.0253	228.88
ART	-0.178	0.024	-7.50
Bias	-0.012	0.008	-1.48
Spelling Format (concatenated)	-0.051	0.024	-2.14
Trial Number	-0.028	0.007	-3.85
Word Length	0.058	0.010	5.67
Joint Frequency	-0.032	0.009	-3.34
Whole Compound Plausibility	-0.026	0.010	-2.70
Bias x Joint Frequency	-0.015	0.006	-2.48

Table 7a: Fixed effects of the multiple regression model fitted to gaze duration. The R^2 of the model is 0.296 and the standard deviation of the residual is 0.412. All numeric predictors were scaled.

Random Effect	Standard Deviation	Correlations between by-participant slopes and intercepts
Compound	0.066	
Participant	0.118	
Presentation format by participant	0.090	0.542

Table 7b: Random effects of the multiple regression model fitted to gaze duration, including random intercepts for compound word and participant by presentation format. The reference level for Spelling Format is spaced.

Total Fixation Time

	Estimate	Standard Error	t value
Intercept	6.008	0.037	161.33
ART	-0.139	0.035	-4.01
Bias	-0.008	0.008	-1.03
Spelling Format (reference level = spaced)	-0.035	0.017	-2.11
Word Length	0.072	0.011	6.47
Joint Frequency	-0.042	0.010	-4.03
Trial Number	-0.046	0.011	-4.11
Whole Compound Plausibility	-0.062	0.010	-5.98
Bias x ART	-0.019	0.007	-2.79
Bias x Joint Frequency	-0.022	0.006	-3.83

Table 8a: Fixed effects of the multiple regression model fitted to total fixation time. The R^2 of the model is 0.342 and the standard deviation of the residual is 0.396. All numeric predictors were scaled.

Random Effect	Standard Deviation	Correlations between by-participant slopes and intercepts
Compound	0.082	
Participant	0.187	
Trial Number by participant	0.048	0.338

Table 8b: Random effects of the multiple regression model fitted to total fixation time, including random intercepts for subject and trial number by participant.

3.0 Discussion

3.1 Summary of Main Findings

The goal of the present study was to examine the role of spelling bias, a probabilistic measure of support that alternative spellings of compound words receive in natural language use, plays in the processing of compound words across individuals with different amounts of reading experience, and across words of differing frequencies of occurrence. Compounds with higher biases towards the format in which they were presented (i.e. with a higher relative frequency of occurrence in that format), were read faster than compounds with less biased spellings. This effect was only seen for the most experienced readers, and for the most frequent words, suggesting that only words with high quality representations, gained either by experience with a specific word, or much exposure to print in general can be specific enough to encode this information.

3.2 Contributions to the Literature

The present study has contributed to the literature in a variety of ways, which will be discussed by topic area.

3.2.1 Compound spacing

This study has replicated the early processing advantage and later cost of adding a space to concatenated compounds (Bertram, Kuperman, Baayen & Hyönä, 2011; Cherng, 2008; Inhoff, et al., 2000; Juhasz et al., 2005). These previous studies only considered compounds with very high biases, typically towards concatenation. The present study used compounds across the range of biases, thus extending the early advantage and later cost of spaced compound words to compound words with any spelling bias.

3.2.2 Isolating the role of probability

This research has examined a meaning-preserving alternation on the lexical level, and has demonstrated that variants of different probabilities affect the way in which they are read, for those individuals and for those words with strong enough representations to encode the distributional information about word spellings. These effects are proportional to the probabilities of the variants.

This research has demonstrated the effect of spelling bias for spaced compound words, which had previously not been explored. Previous studies indicated that there was an effect of bias on lexical decision latencies for concatenated compound words (Kuperman & Bertram, 2013). The present study has shown that there is a processing benefit for spaced as well as concatenated compound words. Additionally this study presented the novel finding that the effect of bias is modulated by word frequency and reading experience.

3.2.3 Individual differences in visual word processing

This study has demonstrated a novel interaction whereby individuals with more reading experience are preferentially affected by the spelling bias of compound words, compared to individuals with less reading experience. Previous studies have found interactions of the opposite pattern - individuals with less reading skill have been more affected by lexical properties compared to individuals with higher scores on skill tests (Adelman, Sabatos-DeVito, Marquis,

& Estes, 2014; Ashby, Rayner & Clifton, 2005; Butler & Hains, 1979; Chateau & Jared, 2000; Hawelka et al., 2010; Hersch & Andrews, 2012; Jared, Levy & Rayner, 1999; Kuperman & Van Dyke, 2011a, 2011b, 2013; Sears, Siakaluk, Chow, & Buchanan, 2008; Yap, Balota, Sibley, & Ratcliff, 2012; but see Whitford & Titone, 2014). These results have implications for the study of other print-specific, or rare patterns. If individuals have little exposure to print, then we may not see effects for these individuals of previously studied, rare patterns. This will be discussed in the Future Directions section.

3.2.4 Lexical quality hypothesis

The present study has added to the body of evidence suggesting that lexical representations can vary across individuals with varying levels of reading experience, and within individuals for words of varying frequencies. One of the consequences of exposure to a given words, and, in turn, high quality representations, appears to be that subtle differences in orthographic information can factor into these representations, even in the absence of differences in phonology or semantics, and that these differences in representation can affect reading behaviour. Orthographic representations seem then to be separable from phonological and semantic information. In addition, the results demonstrate that multiple orthographic variants can be stored in the mental lexicon for a given lexical unit.

3.2.5 Statistics of language

Finally, this research has provided evidence that the spelling bias of compound words affects linguistic processing. Adding to the research by Kuperman and Bertram (2013), spelling bias has been demonstrated to facilitate processing. Since this effect has been shown to be confined to only high quality representations, through increased exposure to print or through high frequency words, it can be suggested that not all language statistical properties can be used across all individuals. Some degree of experience with a lexical item appears to be required in order for individuals to make use of this distributional property.

3.3 Future Directions

A limitation of the present study was that the participant pool was limited to the undergraduate population. It is expected that individuals differing greatly from these individuals in reading experience would respond differently to the spelling bias of compound words. Individuals with much less experience would not be expected to be sensitive to differences in spelling bias. Individuals with very extensive exposure to print may show a larger facilitation and greater costs when the spelling bias supports or does not support the presented spelling format. It is also possible that individuals with mid-to-high and high reading experience would perform similarly, if there was an upper-limit to this effect. This line of research may thus be useful in examining the consequences of very high quality representations, such as how much of a processing advantage could be gained from increasingly high quality representations.

Also yet to be considered is how individuals' own biases to distributional patterns affect their processing. Using corpus frequencies does not allow for a consideration of individuals' experience with any given word, rather an estimate based on a collection of written or spoken language. In order to gain a more fine-grained understanding of the role of distributional properties on the individual level, it would be ideal to determine participants' own experience with a given word, or to base bias estimates on a subject-level variable, such as reading experience, or scores on a reading-related skill test. As well as following this line of testing for spelling bias of compound words, this can also be done for the biases of verbs to take a given argument, such as in the processing of reduced relative clauses, or for the biases of verbs to appear in a double object or prepositional dative.

As mentioned previously, since individuals with less reading experience have been shown to be less sensitive to a print-specific pattern, it is possible that an analogous effect would be seen for other patterns. Some candidate patterns are described in the corpus study by Roland, Dick and Elman (2007). For example, it is possible that differences in processing of subject and object cleft sentences may be due to differences in frequency of these structures, although both of these structures are relatively rare in both spoken and written corpora. Since we have demonstrated a print-specific pattern that requires a certain amount of experience in order for participants to be sensitive to it, then it may be prudent to consider individual differences in reading experience and/or skill when examining rare, or print-specific linguistic patterns. It is possible that the effects previously found by researchers would be attenuated in individuals with less experience.

Finally, future research may consider the electrophysiological consequences of the strengths of lexical representations, and how this relates to the inclusion of information about spelling bias in these representations. Harris, Perfetti and Rickles (2014) conducted a spelling decision task to elicit the Error-Related Negativity (ERN), which is typically more negative when an error is made during a task. They found that the ERN was more positive for misses and false alarms than for hits and correct rejections. They also found that individual differences in spelling scores was correlated with, and significantly predicted, ERN magnitude. A similar experiment may be conducted with compound words with varying spelling biases, to determine whether the ERN would be seen in response to going against the spelling bias of these compound words. For example, would an 'incorrect' response to a concatenated word that was highly biased towards being concatenated be perceived as an error. Additionally, if found, would this ERN vary with exposure to print, or some other reading-related skill?

3.4 Conclusion

The present study has demonstrated that individuals are sensitive to the spelling bias of compound words, but only when their representations of those words are of sufficiently high quality. Gaining high enough quality representations seems to be possible though either having large amounts of exposure to printed text in general, or through increased experience with the given word, through that word having a high frequency. These results have implications for theories of language storage and processing, and suggest that individual differences in reading experience or proficiency should be considered in any study that examines the role of language statistical properties on language production or comprehension.

References

Adelman, J. S., Sabatos-DeVito, M. G., Marquis, S. J., & Estes, Z. (2014). Individual differences in reading aloud: A mega-study, item effects, and some models. *Cognitive psychology*, *68*, 113-160.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67-82.

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1065-1086.

Balass, M., Nelson, J. R., & Perfetti, C. A. (2010). Word learning: An ERP investigation of word experience effects on recognition and word processing. *Contemporary educational psychology*, *35*(2), 126-140.

Bertram, R., Kuperman, V., Baayen, R. H., & Hyönä, J. (2011). The hyphen as a segmentation cue in triconstituent compound processing: It's getting better all the time. *Scandinavian journal of psychology, 52(6)*, 530-544.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation.
In G. Boume, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69-94). Amsterdam: Royal Netherlands Academy of Science.

Bresnan, J., & Ford, M. (2010). Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language, 86*, 168-213.

Buswell, G. T. (1922). *Fundamental reading habits, a study of their development*. Chicago: Chicago University Press.

Butler, B., & Hains, S. (1979). Individual differences in word recognition latency. *Memory & Cognition*, *7*(2), 68-76.

Bybee, 2007

Chace, K. H., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview: effects of reading skill. *Canadian Journal of Experimental Psychology*, *59*(3), 209-217.

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition*, *28*(1), 143-153.

Cherng, M. (2008). *The role of hyphenation in English compound word processing* (Unpublished bachelor's thesis). Wesleyan College, Middletown, CT.

Everatt, J., & Underwood, G. (1994). Individual differences in reading subprocesses: Relationships between reading ability, lexical access, and eye movement control. *Language and speech*, *37*(3), 283-297.

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*(1), 130-147.

Griffin, D. C., Walton, H. N., & Ives, V. (1974). Saccades as related to reading disorders. *Journal of Learning Disabilities, 7*(5), 310-316.

Hawelka, S., Gagl, B., & Wimmer, H. (2010). A dual-route perspective on eye movements of dyslexic readers. *Cognition, 115*(3), 367-379.

Hersch, J., & Andrews, S. (2012). Lexical Quality and Reading Skill: Bottom-Up and Top-Down Contributions to Sentence Processing. *Scientific Studies of Reading*, 16(3), 240-262.

Inhoff, A. W., Radach, R., & Heller, D. (2000). Complex compounds in German: Interword spaces facilitate segmentation but hinder assignment of meaning. *Journal of Memory and Language*, *42*(1), 23-50.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*,*40*(6), 431-439.

Jaeger, T. F., & Tily, H. (2011). On language 'utility': Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science, 2(*3), 323-335.

Jared, D., Levy, B. A., & Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: evidence from proofreading and eye movements. *Journal of Experimental Psychology: General, 128*(3), 219-264.

Joseph, H. S., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition*, *133*(1), 238-248.

Juhasz, B. J., Inhoff, A. W., & Rayner, K. (2005). The role of interword spaces in the processing of English compound words. *Language and cognitive processes, 20*(1-2), 291-316.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay & and S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 39-96). Massachusetts: MIT Press.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typological studies in language, 45*, 229-254.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1-2), 262-284.

Kuperman, V. & Bertram, R. (2013). Moving spaces: Spelling alternation in English noun-noun compounds. *Language and Cognitive Processes, 28*(7), 939-966.

Kuperman, V., & Bresnan, J. (2012). The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language*, *66*(4), 588-611.

Kuperman, V., Van Dyke, J.A. (2011a). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language, 65*(1), 45-73.

Kuperman, V. and Van Dyke, J.A. (2011b). Individual differences in visual comprehension of morphological complexity. In L. Carlson, C. Hoelscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 1643-1648). Austin, TX: Cognitive Science Society.

Kuperman, V. & Van Dyke, J.A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance, 39*(3), 802 – 823.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*(1), 35-54.

McConkie, G. W., Hogaboam, T. W., Wolverton, G. S., Zola, D., & Lucas, P. A. (1979). Toward the use of eye movements in the study of language processing∗. *Discourse processes*, *2*(3), 157-177.

Misyak, J. B., Christiansen, M. H., & Bruce Tomblin, J. (2010). Sequential Expectations: The Role of Prediction-Based Learning in Language. *Topics in Cognitive Science*, *2*(1), 138-153.

Perfetti, C. A. (1985). Reading ability. New York: Oxford University Press.

Perfetti, C. A., (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading*, *11*(4), 357-383.

Perfetti, C. A., & Hart, L. (2001). The lexical bases of comprehension. In D. Gorfien (Ed), *On the consequences of meaning selection* (pp. 67-86). Washington, DC: American Psychological Association.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy*, *11*, 67-86.

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1281.

Rayner, K. (1978). Eye movements in reading and information processing.*Psychological bulletin*, *85*(3), 618.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372.

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*(3), 191-201.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton Jr, C. (1989). Eye movements and on-line language comprehension processes.*Language and Cognitive Processes*, *4*(3-4), SI21-SI49.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: a comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(5), 1188.

Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic bulletin & review*, *17*(6), 834-839.

Saffran, J. R. (2003). Statistical language learning mechanisms and constraints. *Current directions in psychological science*, *12*(4), 110-114.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of memory and language*, *35*(4), 606-621.

Sears, C. R., Siakaluk, P. D., Chow, V. C., & Buchanan, L. (2008). Is there an effect of print exposure on the word frequency effect and the neighborhood size effect?. *Journal of Psycholinguistic Research*, *37*(4), 269-291.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive science*, *23*(4), 569-588.

Sepp, M. (2006). *Phonological constraints and free variation in compounding: A corpus study of*
English and Estonian noun compounds. Unpublished doctoral dissertation, City University of New York, NY.

Shie, J. S. (2002). English hyphenated compounds. *Journal of the Da-Jeh University, 11,* 89-98.

Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1162-1169.

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*, *1*(2), 147-165.

Underwood, G., Hubbard, A., & Wilkinson, H. (1990). Eye fixations predict reading comprehension: The relationships between reading skill, reading speed, and visual inspection. *Language and speech*, *33*(1), 69-81.

Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. *The Quarterly Journal of Experimental Psychology*, *67*(4), 703-727.

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*(2), 729-742.

Whitford, V., & Titone, D. (2014). The effects of reading comprehension and launch site on frequency–predictability interactions during paragraph reading. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1151-1165.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 53.