When experience meets language statistics:

Individual variability in processing English compound words

Kaitlin Falkauskas

McMaster University, Canada

Victor Kuperman

McMaster University, Canada

Corresponding author:

Victor Kuperman, Department of Linguistics and Languages, McMaster University, Togo
Salmon Hall 626, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4M2, Phone: 905-525-9140, x. 20384, Email: vickup@mcmaster.ca
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) or belonged to larger morphological families, 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
Introduction

There is a consensus that the statistical patterns of language use are linked to both the mental representation of linguistic structure – from phonological segments to discourse units, as well as language production and comprehension (see, among many other reviews, Jaeger & Tily, 2011; Jurafsky, 2003; Jurafsky, Bell, Gregory & Raymond, 2001; Seidenberg & MacDonald, 1999). In particular, many recent theories of language processing argue that these statistical patterns are a causal factor in determining the processing effort in written language comprehension (MacDonald, 1999; 2013; Levy, 2008; Pickering & Garrod, 2007; 2013). It is a well-established notion that distributional preferences vary between individuals, and one major source of this variability is the amount of experience with these phenomena that individuals accumulate through their exposure to language (cf. MacDonald & Christiansen, 2002; Stanovich & West, 1989). Thus, an important and an under-researched empirical question for models that advocate the distribution-comprehension link is how individual differences in language experience translate into individual variability in comprehension behavior. The present study contributes to the investigation of this question 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 the 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, (Perfetti, 1985; 2007; Perfetti & Hart, 2002; Plaut, McClelland, Seidenberg, & Patterson, 1996). This includes representation of the word’s semantic and syntactic environments. The Lexical Quality Hypothesis further argues that automatically-activated mappings occur between these three components: orthography, phonology, and semantics. 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” (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 can develop in multiple ways. One of these ways is through individuals' repeated exposure to a specific word. As a word is seen repeatedly, its representation can become more entrenched, and links between the representational components more stable. Second, lexical quality can increase with individuals' increased experience with printed materials in general, which requires discriminating that word from other words. With respect to orthography, which is the focus of the present paper, lexical quality of a specific word can range between individuals 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). Also, different words will vary within an individual as a function to how frequently that person has encountered these words, similar words and all words in the lexicon.

The Lexical Quality Hypothesis and models of statistical learning makes a number of predictions on how variability in reading experience affects visual word comprehension. First, reduced exposure to print is expected to render reading more effortful overall (Perfetti, 2007). Exposure to print is thus likely to appear as a main effect in every aspect of reading behaviour
that is constrained by processing difficulty, including the speed and accuracy of word recognition. The expectation is that individuals with more experience will perform better on these tasks, compared to less experienced individuals. This prediction has been supported by studies that explored the relationship between variability in experience and skills, and reading performance in non-clinical adult populations. These studies have tested less proficient readers (i.e. those with a weaker performance on skill tests), and less experienced readers (i.e. those with fewer years of schooling, or lower scores on exposure-to-print tests). For comparability with the present set of findings, our review emphasizes studies that use eye-tracking as their experimental paradigm. 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 or less experienced readers 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).\(^1\)

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, along with weaker

---

\(^1\) While couched in terms of natural reading, which is a complex skill and requires extensive language background and specific training, statistical learning has an effect even within short experimental sessions (Fine, Jaeger, Farmer, & Qian, 2013; 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).
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 (Adelman, Brown, & Quesada, 2006). The Lexical Quality Hypothesis would therefore predict 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). 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, 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 all readers show more difficulty in processing words to which they have had less exposure, however, more 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). Individuals with greater proficiency or experience, therefore, seem to be less affected by differences in the above listed distributional properties.

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.

**Spelling alternations**

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 its own probability of realization, yet the meanings of all variants are near-identical, that is, they have the same denotation. In a minority of compounds, a difference in spelling does translate into a difference in denotation (*dishwasher* is a device, *dish washer* is a person employed to clean dishes) or a preference for one of denotations (*football* vs *foot ball* as a game or an object). Such compounds were excluded from consideration in both Kuperman and Bertram (2013) and in the present paper². 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 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). Since all of the variants have the same propositional content (for datives) or denotation (for compounds), any behavioural differences in the comprehension or production of these

² Experiment 1 in Marelli et al. (2013) also reports spelling-driven differences in connotations of compound meanings: there is a weak tendency for concatenated compounds to be interpreted in terms of their constituents, and for spaced ones as integral semantic units (e.g. *home work* is more readily associated with job and home, and *homework* with teacher and lesson). We acknowledge a possible subtle difference in connotations of spaced versus concatenated compounds (much like there might a difference in register between dative alternatives, e.g. Pinker, 1989, see below). In what follows, we define meaning-preservation narrowly, as an identity of denotations between spelling alternatives.
variants can be attributed to their differences in probabilities or the differences in their form.

When considering the spelling alternation between most 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.

Prior research on compound spelling

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 the 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 suggested 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, the studies mentioned above 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 unexamined the 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.

**Goals of the study**

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 probabilistic effect would be modulated by the frequency of occurrence of the compound in question, the presence of compounds that are similar in structure (i.e. share morphological constituents), and 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. Moreover, spelling preferences among the compound’s family members (complex words sharing a morpheme with the target compound, e.g. *girlfriend*, *girl scout*, *boyfriend*) have a strong analogical influence on the orthographic choice in a given compound (Kuperman & Bertram, 2013). Increased familiarity with whether and how the compound's family members alternate in spelling would therefore contribute to individuals' sensitivity to the distributional bias of that given word. Thus, it is possible that compounds from stronger (larger, or more frequent) families would show a stronger effect of the spelling bias. Finally, more experienced readers will have had greater exposure to any and all words, including the critical compounds and their morphological families. This exposure gives experienced readers more opportunity to include subtle orthographic information such as the compounds’ spelling biases into their lexical representations. It also provides better grounds for discriminating the alternative spellings from each other and from all other words in the person’s mental lexicon. It stands to reason that extensive experience with print will give an individual preferential access to and practice with lower-frequency compounds, in all their spelling variants. Thus, a three-way interaction is expected, where the spelling bias has more influence on relatively experienced readers, especially for lower-frequency words. Similarly, a three-way interaction between the spelling bias, experience and compound family size is possible, as is a
four-way interaction including the spelling bias, experience, compound frequency and compound family size. These possibilities are explored below. In sum, we hypothesized that the effect of the probabilistic bias towards the presented spelling of compound words would be stronger in more frequent compounds, compounds from stronger families, and in individuals with more reading experience.

Method

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 compounds 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 as either spaced or concatenated, 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 representing quartiles of the spacing bias range. Between twenty eight and thirty two compounds were then selected from each of these bins, for a total of 120 compounds, so that the distribution of the bias in selected stimuli would be nearly uniform. 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 the 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, as well as 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.

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 presented with 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 reading 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 spelling bias, compound length, and compound frequency. We also considered the number and the cumulative frequency of all compounds that share the left or the right constituent with the target compound (left and right family size and family frequency). The frequency of the target compound was subtracted from the family frequency estimate and family size was reduced by one to avoid circularity. All frequency-related measures 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. Of measures related to compound constituents, only family frequencies showed consistent results in our analyses: in what follows, we only report those predictors.

Marelli et al. (in press) have recently shown using the Latent Semantic Analysis technique that semantic similarity of constituents varies between spaced versus concatenated formats of English compounds and thus might confound the effect of spelling bias. We do not have format-specific estimates of semantic similarity at our disposal to test whether they affect present results. We note that this potential confound is unlikely in our data, as estimates obtained for concatenated compounds in our sample using the term-term Latent Semantic Analysis with 300 factors (http://lsa.colorado.edu/) show a very weak correlation ($r = -0.09, p = 0.39$) with the bias towards concatenation.

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.

**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 the 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.

Skill tests

Tests of orthographic segmentation, vocabulary size (Nation & Beglar, 2007) and reading
experience (Acheson, Wells, & MacDonald, 2008) were conducted in order to assess individual
variability in reading proficiency and experience. Only the Author Recognition Test of reading
experience showed reliable results across the eye-movement record: in what follows we only
report those.

ART/MRT. The Author Recognition Test (ART) was used to assess amount of reading
experience of participants (Acheson et al., 2008). This test provides participants with a list of
names that contain 50% names of published authors and 50% distracters in the form of non-
author names. Participants were instructed to indicate the names that they were certain were
authors. The score for these tests was the number of correctly identified authors or magazines
minus the number of incorrectly identified authors.

Statistical Considerations

All continuous predictors were z-transformed to allow the predictors to be compared on
the same scale: compound frequency and positional family frequencies were 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).
Linear mixed-effects multiple regression models were used for this study with participant and word as random effects with the Gaussian (for continuous predictors) or binominal (for binary predictors) underlying distributions (Baayen, 2008; Baayen, Davidson & Bates, 2008; Jaeger, 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). We used a single label (milk+shake) to represent random effects associated with both spaced and concatenated variants of the same compound (milkshake, milk shake). Nesting the spelling variance under this label did not lead to an improved model in any of the analyses. After fitting a model, we removed outliers if they were outside of the range of -2.5 to 2.5 units of standard deviation away from the residual error of the model. The model was then refitted to the trimmed data set.

Spelling bias is co-determined by several lexical properties, including compound length, frequency, family size and others (Kuperman & Bertram, 2013). Our decision to present each compound both in its more and less biased format effectively eliminated collinearity between bias and its co-determinants (all correlation coefficients r < 0.01, p > 0.1). This is because every value of frequency, family size or other co-determinants would be associated with two different values of bias: the bias towards concatenation and its inverse. 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 p-values obtained using the Satterthwaite's approximation for degrees of freedom as implemented in lmerTest() package v. 2.0 (Kuznetsova, Brockhoff, & Christensen, 2013).

**TABLE 1 APPROXIMATELY HERE**

**Results and Discussion**

The original data set contained 3465 data points. Trials were removed if the compound word was not fixated on (38 data points, 1.1%), and if the compound was skipped in the first pass, and fixated on in subsequent passes (17 data points, 0.5%). Trials with first fixation durations 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. Trimming was done based on participants' total reading times on each compound word. Trials in which the total reading time of the compound word was more than three standard deviations away from the participant's mean 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), bias by compound frequency, and bias by left or right constituent family frequencies, 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 also reported. 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

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 versus concatenated compounds. Spaced compounds elicited shorter single fixation durations \( b = 0.062, \ SE = 0.015, p < 0.001 \) and second fixation durations \( b = 0.112, \ SE = 0.031, p < 0.001 \), however they also were more likely to elicit a second fixation \( b = -0.720, \ SE = 0.086, p < 0.001 \). As a result, spaced compounds came with longer gaze durations \( b = -0.060, \ SE = 0.023, p = 0.0151 \) and total reading times \( b = -0.041, \ SE = 0.016, p = 0.012 \). 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 which were presented in their more frequently occurring spelling format (spaced or concatenated) to be processed 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 scores on the ART, a measure of overall reading experience, with compound frequency, a measure of experience with a particular compound word, and with compound family frequency, a position-specific measure of experience with the morphological neighborhood of the compound.

**Measures of experience.**

*ART/MRT:* Scores on the ART have been argued to be the most direct measure of exposure to print (Acheson et al., 2008; Stanovich & West, 1989). 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 \(b = -0.062, \ SE = 0.020, p = 0.005\), second fixations \(b = -0.108, SE = 0.039, p = 0.008\), gaze durations \(b = -0.179, SE = 0.025, p < 0.001\), and total reading times \(b = -0.143, SE = 0.035, p < 0.001\) on the target words. They were also less likely to refixate on the target word \(b = -0.611, SE = 0.109, p < 0.001\). 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.

**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 [for regression estimates of this nonlinear negative effect, see Table 3a], gaze durations \(b = -0.032, SE = 0.009, p < 0.001\), and total reading times
In addition, compound frequency interacted with spelling bias such that higher-frequency compounds showed stronger bias effects (see below), as predicted.

**Constituent family frequencies.** Spelling bias interacted with positional (left and right) constituent family measures, as reported below. The time-course of involvement of left and right positional family frequencies was mostly driven by the order in which constituents became available for foveal inspection: the left constituent family frequency in early eye-movement measures (first-of-many fixation duration and refixation probability), and the right one in later and cumulative measures (single fixation duration and total reading time). Specifically, higher left constituent family frequency led to lower refixation probability \( b = -0.090, SE = 0.040, z = -2.241, p = 0.025 \), especially in spaced compounds. Similarly, higher right constituent family frequency led to shorter total reading time \( b = -0.029, SE = 0.014, t = -1.998, p = 0.046 \), and the effect was much stronger in concatenated than spaced compounds. (Interactions of family measures with the spelling bias are discussed below).

**Interactions between lexical properties and skill tests.**

Importantly, the effect of bias was modulated by compound frequency, positional constituent family frequencies, and by individuals' scores on the ART. Where found, each type of the interaction had the same qualitative shape. The effect of bias was stronger if the experience with a compound, its family, or all printed words was more extensive, as attested by a higher compound or family frequency or a higher ART score. We also observed a theoretically predicted three-way interaction between ART, joint frequency, and bias in second fixation duration, described below. Figures 1-3 show a partial effect of spelling bias on (1) probability of
refixation, (2) gaze duration and (3) total reading time, as estimated by linear mixed-effects models and adjusted for values corresponding to 10th, 30th, 50th, 70th, and 90th percentiles of (1) family frequency, (2) joint frequency or (3) ART score distributions, respectively. In what follows, we present the interactions, while their time-course is interpreted in the General Discussion.

A significant interaction was observed between spelling bias and ART scores for second fixation duration (further qualified by joint frequency, see below) and for total reading time \([b = -0.017, \ SE = 0.007, \ p = 0.01]\). 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).

A bias by compound frequency interaction was also seen in second fixation duration (further qualified by ART, see below), gaze duration \([b = -0.016, \ SE = 0.006, \ p = 0.005]\) and total reading time \([b = -0.019, \ SE = 0.006, \ p = 0.001]\) (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 increased, the negative slope of the bias effect becomes steeper (see Figure 2).
The effect of bias was additionally modulated by constituent families. Participants were less likely to refixate on the compound word if it appeared in a more supported spelling format, especially if this compound came from a family with higher-frequency members [b = -0.090, SE = 0.040, z = -2.241, p = 0.025], Figure 3. There was a weak reversal of the bias effect in very low-frequency families, suggesting a slight increase in refixation probability\(^3\).

Similarly, single fixation durations were shorter for compounds presented in the preferred spelling format, especially if the compounds’ right constituent family frequency was higher [b = -0.24, SE = 0.769, t = -3.124, p = 0.002].

**FIGURE 3 APPROXIMATELY HERE**

Finally, a three-way interaction between bias, ART scores, and joint compound frequency was observed in second fixation duration [b = 0.038, SE = 0.017, t = 2.207, p = 0.027]. Second fixation durations were shorter if compounds were presented in the more preferred format. The effect of bias was stronger in all higher-frequency words, weaker but noticeable in lower-frequency words read by experienced readers, and negligibly small in lower-frequency words read by less experienced readers (figure not shown). No interactions were observed between bias towards spelling format and the presentation format of the compound, nor did we see additional reliable three-way or four-way interactions between constituent frequencies, joint frequency, ART, and bias (all |t|s < 1).

---

\(^3\) To closer inspect this pattern, we fitted a generalized additive model to refixation probability in which the interaction was modeled as a tensor product of bias and left family size, and thus the interaction surface was allowed to take any form, and not only the planar form allowed by mixed-effects models. This inspection revealed that the apparent reversal is even weaker in the generalized additive model and thus is an artifact of a linear approximation of the data, imposed by linear regression models.
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. Gaze durations \[b = -0.029, \text{SE} = 0.007, p < 0.001\] and total reading times \[b = -0.044, \text{SE} = 0.011, p < 0.001\] on the target compound were shorter, and refixations were less likely to occur \[b = -0.080, \text{SE} = 0.039, p = 0.034\] 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 \[b = 0.018, \text{SE} = 0.007, p = 0.008\]. The overall advantage in compound processing for relatively experienced readers was increasingly weaker as the experiment progressed, suggesting a training effect in less experienced readers. Trial number did not enter into other two-way or three-way interactions with any critical predictor (all \(|t|’s < 1.5\).

Word length. Longer target words elicited longer first-of-many fixation durations \[b = -0.012, \text{SE} = 0.006, p = 0.059\], and gaze durations \[b = 0.062, \text{SE} = 0.010, p < 0.001\], and were more likely to be refixated \[b = 0.232, \text{SE} = 0.039, p < 0.001\].

Whole compound plausibility. Compound words that were more semantically plausible in the context of the beginning of the sentence elicited shorter gaze durations \[b = -0.027, \text{SE} = 0.009, p = 0.004\] and total reading times \[b = -0.066, \text{SE} = 0.010, p < 0.001\].

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 a sufficient number of opportunities to learn statistical patterns. The present study explored this hypothesis by considering a specifically orthographic phenomenon: the spelling alternation of English compound words. Knowledge of the different probabilities associated with the alternating variants can only be acquired through reading. Given the degree of uncertainty in the spelling choice for many compounds (cf. Kuperman & Bertram, 2013), knowledge of the different probabilities 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 (with possible subtle differences in connotations, cf. Marelli et al., in press), 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, for words that occur most frequently when collapsed across both spelling formats, and for words that have many well-established structural analogues in the language, i.e. larger morphological families. Generally, 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 or compounds from weaker families. More specifically, the facilitatory effect of the bias towards the presented spelling was at its strongest in readers with higher scores on the ART, for total reading time. This effect was also stronger for compounds with the highest joint frequency of occurrence in gaze duration and total reading time, and for compounds with stronger constituent families in single
fixation duration and refixation probability. Moreover, the effect of bias was the strongest in the most experienced readers and for the most frequent compounds in second fixation duration, see Table 1. To summarize, statistical preferences for how a compound is to be spelled are most influential in situations which granted the most chance to absorb these preferences, through increased exposure to spelling variants of this particular compound, analogous compounds in the morphological family, and printed text overall. A previous study (Kuperman & Van Dyke, 2013) demonstrated that the effect of reading experience is greater for low-frequency words, in that individuals with less reading experience are less likely to be exposed to low-frequency words, compared to more experienced individuals. If our argument is valid, we would thus expect more experienced readers to be more sensitive to the spelling bias, even in low-frequency words compared to individuals with less reading experience, since the individuals with more experience would have had more exposure to the compound words, and, consequently, variation in their spelling. This prediction is borne out in the three-way interaction between bias, compound frequency and the index of exposure to print (ART).

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, or not at all affected. In most previous studies of 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,
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 through 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 windowsill (bias towards concatenation: 0.60) are more likely to support the concatenated presentation, than the spaced presentation (bias: 0.40), and thus individuals may be facilitated when they read the word as concatenated, and potentially harmed when they read it as spaced. Without sufficient exposure to the specific compound, its closest morphological neighbours, and print in general, readers are less able to achieve entrenched orthographic representations to the extent necessary to discriminate the spelling variants: from one another, from similar compounds in the same family, and from all other words in the mental lexicon.

Interestingly, our data make a compelling case that the mental representation of a word must encompass all orthographic representations of that word (for a related point, see Marelli et al., in press). Relative frequencies of alternative spellings impact compound word recognition in a graded way, which is proportional to the amount of support for a presented spelling which exists 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, see 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).

In the remainder of this section, we discuss specific details of our findings: the time-course of probabilistic effects on compound processing, as well as interactions between different (compound-specific and general) metrics of reading experience. We conclude by discussing limitations of this study and by making methodological points regarding studies of individual differences in reading.

*The time-course of probabilistic effects on compound processing*

The order of involvement of morphological families followed the usual left-to-right processing of compound words (Pollatsek, Hyönä, & Bertram, 2000). Only left constituent families affected the first-pass eye-movement measures (first of many fixation duration and refixation probability), while the right constituent family influenced the cumulative measures of single fixation duration and total reading time, after both constituents were inspected foveally.

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 reading 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.

The independent or interactive effect of spelling bias permeated the entire eye-movement record, except for the earliest durational measure that we considered: first of many fixation duration. Why is the earliest available measure unaffected by bias, even though the low-level orthographic discrimination can be argued to take place at the earliest stages of word processing (Rayner, 1998)? This is likely due to the demands of visual acuity. About 90% of presented compounds elicited more than one fixation, due to the presence of a space in half of the target words and to the length of the compound words (8-14 characters in the concatenated variant). Taken together, these facts suggest that the entire orthographic form of alternating compounds (i.e. both the left and the right constituent) needed to be visually inspected in order to discriminate the presented form from its alternative spelling (*firehall* and *fire hall*) and from compounds in the same family (e.g. *fire+man, fire+fly*) as well as unrelated words (e.g., *fir+tree, first+born*). In the minority of cases when a single fixation was sufficient to process the compound, the duration of this fixation was also influenced by spelling bias.

A similar dependency on the availability of sufficient visual information characterized the time-course of interactions of spelling bias with compound frequency, family frequency, and ART scores (i.e., a measure of individual exposure to print). Measures like the probability of refixation on a compound (a measure indexing a decision made during the first fixation on the word) and second fixation duration (a measure registered in the first reading pass) reveal both the effect of spelling bias and its variability over individuals of different experience levels and compounds offering different learning possibilities. To sum up, the effect of a subtle, orthographically specific lexical alternation affected the speed of compound processing strongly
and robustly as soon as sufficient visual information was obtained – typically, in two or more fixations on the target word, spread over one or multiple reading passes.

The three-way bias by frequency by exposure-to-print interaction was estimated as statistically reliable in second fixation duration but was not replicated in cumulative measures that blend together words that are read in a single fixation or multiple fixations (gaze duration) and in one reading pass or multiple passes (total reading time). One reason for the frailty of this interaction may be volatility in second fixation duration as a response variable: it is only estimated for those trials which required multiple fixations, and heavily depends on the duration of the first-of-many and, where applicable, subsequent fixations.

Another reason may be that the set of measures of experience currently available to researchers is suboptimal. In current practice, frequency counts for individual words or their morphological families are aggregated from large corpora and do no vary across participants with differing amounts of exposure to print, even though this variability is substantial (Kuperman & Van Dyke, 2013). Moreover, while reliable as a metric of an individual’s reading experience (Acheson et al., 2008; Stanovich & West, 1989) the score in the current version of the Author Recognition Test is partly contingent on one’s familiarity with a specific set of literary writers in a restricted range of time periods and genres (Moore & Gordon, in press). Finally, experimenters have a much tighter control over selecting stimuli for their distributional properties than over selecting an encompassing range of individuals with requisite variability in reading experience. Thus, until individualized estimates of word frequency and better calibrated tests become available, replicability of the interactions between word and participant properties within an experiment and across experiments will partly depend on (a) how well the fixed, corpus-based estimates approximate the differing amounts of exposure to specific words and word families in
selected individuals; (b) how precise the Author Recognition test is in estimating reading habits of those individuals, and (c) how variable those individuals are in their reading experience.

Limitations and methodological considerations

A limitation of this study is that our 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 not large in numbers (N = 29) and 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. Given our findings, however, future studies will benefit from testing individuals on the other end of the spectrum. Individuals with very extensive exposure to print, such as copy editors or English professors, for example, may show even larger effects. In this exquisitely proficient population, with extensive reading experience, we may see a larger benefit for more supported compounds, and more of a cost for less supported compounds.

Additionally, our battery of individual differences measures was fairly small and incorporated indices of exposure to print (the Author and Magazine Recognition Tests), skill of orthographic segmentation, and vocabulary size. Of these measures, only scores on the ART showed consistent effects on the eye-movement record (in line with the findings of Moore & Gordon, in press, and with the central role of language experience, cf. MacDonald & Christiansen, 2002), as well as entered into interactions with spelling bias. It is possible, however, that ART scores, as a specific construct of print exposure, are just a mediator of another skill that is practiced and improves as a result of extensive experience with printed materials. To establish the difference between the mere exposure to printed words as practice in discriminative word learning, or a specific skill that is linked to this practice, future research will
need to employ a more extensive battery of individual differences measures, especially those
tapping into the skill of orthographic processing.

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 the 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 these criteria, 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.

Acknowledgments
This paper is based on the MSc thesis of the first author (completed in September, 2014). This
research was partially supported by SSHRC CGS Master’s Scholarship 766-2013-0507 to the
first author; and the SSHRC Insight Development grant 430-2012-0488, the NSERC Discovery
grant 402395-2012, the NIH R01 HD 073288 (PI Julie A. Van Dyke), and the Early Researcher
Award from the Ontario Ministry of Research and Innovation to the second author. Thanks are
due to the audience of the Psychonomic Society's 54th Annual Meeting for providing valuable feedback about earlier stages of this project, and to Petar Milin and two anonymous reviewers for their insightful comments on earlier drafts.

References


New York: Continuum International Publishing Group Ltd..


### Tables

<table>
<thead>
<tr>
<th></th>
<th>Range</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>range of transformed values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Compound Frequency</strong></td>
<td>61, 9698</td>
<td>444.000</td>
<td>140.000</td>
<td>1217.466</td>
<td>-1.109, 3.936</td>
</tr>
<tr>
<td><strong>Left Constituent Family Frequency</strong></td>
<td>61, 33588</td>
<td>3164.735</td>
<td>868.000</td>
<td>5415.692</td>
<td>4.127, 10.422</td>
</tr>
<tr>
<td><strong>Right Constituent Family Frequency</strong></td>
<td>66, 44699</td>
<td>3595.693</td>
<td>875.000</td>
<td>6415.357</td>
<td>4.205, 10.708</td>
</tr>
<tr>
<td><strong>Length</strong></td>
<td>8, 14</td>
<td>9.700</td>
<td>10.000</td>
<td>1.239</td>
<td>-1.374, 3.469</td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>0.003, 0.997</td>
<td>0.500</td>
<td>0.500</td>
<td>0.299</td>
<td>-1.664, 1.658</td>
</tr>
<tr>
<td><strong>ART</strong></td>
<td>0, 40</td>
<td>11.080</td>
<td>9.000</td>
<td>8.175</td>
<td>-1.355, 3.538</td>
</tr>
<tr>
<td><strong>Whole compound plausibility</strong></td>
<td>2.263, 6.857</td>
<td>5.665</td>
<td>5.786</td>
<td>0.790</td>
<td>-4.307, 1.509</td>
</tr>
<tr>
<td><strong>Trial number</strong></td>
<td>11, 198</td>
<td>104.260</td>
<td>103</td>
<td>54.133</td>
<td>-1.723, 1.732</td>
</tr>
</tbody>
</table>

*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 measures

<table>
<thead>
<tr>
<th>Spelling format (with spaced as the reference level)</th>
<th>Bias</th>
<th>ART</th>
<th>Joint Frequency</th>
<th>Left Family Frequency</th>
<th>Right Family Frequency</th>
<th>ART x Bias</th>
<th>ART x Joint Frequency</th>
<th>Bias x Joint Frequency</th>
<th>Left Family Frequency x Bias</th>
<th>Right Family Frequency x Bias</th>
<th>ART x Bias x Joint Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single Fixation Duration</strong> N = 1420 Ntrimmed = 1386</td>
<td>b = -0.062 SE = 0.015 t = 4.129 p &lt; 0.001</td>
<td>b = -0.062 SE = 0.020 t = -3.053 p = 0.005</td>
<td>Better readers are faster.</td>
<td>b = -0.071 SE = 0.038 t = -1.864 p = 0.065 (restricted cubic splines term 2) More frequent words are read faster.</td>
<td>b = 0.001 SE = 0.009 t = 0.136 p = 0.892</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>b = -0.024 SE = 0.008 t = -3.124 p = 0.002 Greater effect of bias in higher family frequency compounds.</td>
<td>NS</td>
</tr>
<tr>
<td><strong>First of Many Fixation Duration</strong> N = 1913 Ntrimmed = 1857</td>
<td>b = -0.001 SE = 0.017 t = 0.087 p = 0.931</td>
<td>b = -0.001 SE = 0.005 t = -0.198 p = 0.843</td>
<td>b = -0.026 SE = 0.024 t = -1.095 p = 0.282</td>
<td>NS</td>
<td>NS (but see interaction with spelling format in text)</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Refixation Probability</strong> N = 3333 Ntrimmed = 3333</td>
<td>b = -0.720 SE = 0.086 p &lt; 0.001</td>
<td>b = -0.611 SE = 0.019</td>
<td>Refixation is more likely for spaced compounds.</td>
<td>b = -0.081 SE = 0.045 p = 0.070</td>
<td>b = 0.026 SE = 0.044 P = 0.559</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>b = -0.090 SE = 0.040 p = 0.025 Greater effect of bias in higher family frequency compounds.</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Second Fixation Duration</strong> N = 1913 Ntrimmed = 1862</td>
<td>b = 0.112 SE = 0.031 t = 3.649 p &lt; 0.001</td>
<td>b = 0.019 SE = 0.017 t = 1.101 p = 0.271</td>
<td>b = -0.108 SE = 0.039 t = -2.815 p = 0.008 Better readers are faster.</td>
<td>b = 0.019 SE = 0.017 t = 1.107 p = 0.270</td>
<td>NS</td>
<td>b = 0.0002 SE = 0.021 t = -0.009 p = 0.993</td>
<td>b = 0.044 SE = 0.020 t = 2.165 p = 0.030 Greater effect of ART in less frequent words.</td>
<td>b = 0.007 SE = 0.014 t = 0.493 p = 0.622</td>
<td>NS</td>
<td>NS</td>
<td>b = 0.038 SE = 0.017 t = 2.207 p = 0.027 Greater effect of bias for better readers, especially in less frequent words.</td>
</tr>
</tbody>
</table>
Table 2: Summary of the critical effects for each eye-movement measure. The regression coefficient estimate, the standard error, the t-value and the p-value 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).
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.
Figure 3: Partial effects of the probabilistic bias towards the presented format on the probability of refixating on the compound in the first pass, broken down by percentiles of left constituent family frequency. Values of family frequency are shown on the right edge and stand for the 10th, 30th, 50th, 70th and 90th percentiles of its distribution.
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 a **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 earthquake that happened two towns away. 0
The farmer harvested the honeycomb 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 albatross 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**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.340</td>
<td>0.027</td>
<td>194.576</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>ART</td>
<td>-0.062</td>
<td>0.020</td>
<td>-3.053</td>
<td>0.005</td>
</tr>
<tr>
<td>Bias</td>
<td>-0.001</td>
<td>0.007</td>
<td>0.144</td>
<td>0.885</td>
</tr>
<tr>
<td>Spelling Format</td>
<td>0.062</td>
<td>0.015</td>
<td>4.129</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>(concatenated)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joint frequency</td>
<td>0.036</td>
<td>0.027</td>
<td>1.305</td>
<td>0.195</td>
</tr>
<tr>
<td>rcs term 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joint frequency</td>
<td>-0.071</td>
<td>0.038</td>
<td>-1.864</td>
<td>0.065</td>
</tr>
<tr>
<td>rcs term 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right Family</td>
<td>0.001</td>
<td>0.009</td>
<td>0.136</td>
<td>0.892</td>
</tr>
<tr>
<td>Frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias x Right</td>
<td>-0.024</td>
<td>0.008</td>
<td>-3.124</td>
<td>0.002</td>
</tr>
<tr>
<td>Family Frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3a: Fixed effects of the multiple regression model fitted to single fixation duration. The R² of the model is 0.231 and the standard deviation of the residual is 0.264. All numeric predictors were scaled. “Rcs” stands for restricted cubic splines, fitted with 3 knots. The reference level for Spelling Format is spaced.

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound</td>
<td>0.042</td>
</tr>
<tr>
<td>Participant</td>
<td>0.103</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.295</td>
<td>0.024</td>
<td>222.043</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Trial Number</td>
<td>0.004</td>
<td>0.006</td>
<td>0.636</td>
<td>0.525</td>
</tr>
<tr>
<td>ART</td>
<td>-0.026</td>
<td>0.024</td>
<td>-1.095</td>
<td>0.282</td>
</tr>
<tr>
<td>Spelling Format</td>
<td>-0.001</td>
<td>0.017</td>
<td>0.087</td>
<td>0.931</td>
</tr>
<tr>
<td>(concatenated)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>-0.001</td>
<td>0.005</td>
<td>-0.198</td>
<td>0.843</td>
</tr>
<tr>
<td>Word Length</td>
<td>-0.012</td>
<td>0.006</td>
<td>-1.909</td>
<td>0.059</td>
</tr>
<tr>
<td>Whole Compound</td>
<td>-0.011</td>
<td>0.006</td>
<td>-1.736</td>
<td>0.085</td>
</tr>
<tr>
<td>Plausibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4a: Fixed effects of the multiple regression model fitted to first of many fixation duration. The $R^2$ of the model is 0.265 and the standard deviation of the residual is 0.229. All numeric predictors were scaled. The reference level for Spelling Format is spaced.

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Standard Deviation</th>
<th>Correlations between by-participant slopes and intercepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td>Participant</td>
<td>0.121</td>
<td></td>
</tr>
<tr>
<td>Presentation format by participant</td>
<td>0.064</td>
<td>0.05</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard Error</th>
<th>z value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.550</td>
<td>0.410</td>
<td>-3.776</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>ART</td>
<td>-0.611</td>
<td>0.109</td>
<td>-5.583</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Bias</td>
<td>-0.051</td>
<td>0.039</td>
<td>-1.297</td>
<td>0.195</td>
</tr>
<tr>
<td>Spelling Format</td>
<td>-0.720</td>
<td>0.086</td>
<td>-8.408</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>(concatenated)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial Number</td>
<td>-0.080</td>
<td>0.039</td>
<td>-2.064</td>
<td>0.039</td>
</tr>
<tr>
<td>Word Length</td>
<td>0.232</td>
<td>0.039</td>
<td>5.986</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Joint Frequency</td>
<td>-0.081</td>
<td>0.045</td>
<td>-1.811</td>
<td>0.070</td>
</tr>
<tr>
<td>Left Family Frequency</td>
<td>0.026</td>
<td>0.044</td>
<td>0.584</td>
<td>0.559</td>
</tr>
<tr>
<td>Bias x Left Family Frequency</td>
<td>-0.090</td>
<td>0.040</td>
<td>-2.241</td>
<td>0.025</td>
</tr>
</tbody>
</table>

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.
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

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Standard Error</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.587</td>
<td>0.039</td>
<td>118.389</td>
</tr>
<tr>
<td>ART</td>
<td>-0.108</td>
<td>0.039</td>
<td>-2.815</td>
</tr>
<tr>
<td>Bias</td>
<td>0.019</td>
<td>0.017</td>
<td>1.101</td>
</tr>
<tr>
<td>Spelling Format (concatenated)</td>
<td>0.112</td>
<td>0.031</td>
<td>3.649</td>
</tr>
<tr>
<td>Joint Frequency</td>
<td>0.019</td>
<td>0.017</td>
<td>1.107</td>
</tr>
<tr>
<td>ART x bias</td>
<td>0.0002</td>
<td>0.021</td>
<td>-0.09</td>
</tr>
<tr>
<td>ART x frequency</td>
<td>0.044</td>
<td>0.020</td>
<td>2.165</td>
</tr>
<tr>
<td>Bias x frequency</td>
<td>0.007</td>
<td>0.014</td>
<td>0.493</td>
</tr>
<tr>
<td>ART x Bias x Frequency</td>
<td>0.038</td>
<td>0.017</td>
<td>2.207</td>
</tr>
</tbody>
</table>

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 0.636. All numeric predictors were scaled. The reference level for Spelling Format is spaced.

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

Gaze Duration

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Standard Error</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.805</td>
<td>0.027</td>
<td>218.760</td>
</tr>
<tr>
<td>ART</td>
<td>-0.179</td>
<td>0.025</td>
<td>-7.144</td>
</tr>
<tr>
<td>Bias</td>
<td>-0.012</td>
<td>0.008</td>
<td>-1.580</td>
</tr>
<tr>
<td>Spelling Format (concatenated)</td>
<td>-0.060</td>
<td>0.023</td>
<td>-2.552</td>
</tr>
<tr>
<td>Trial Number</td>
<td>-0.029</td>
<td>0.007</td>
<td>-4.149</td>
</tr>
</tbody>
</table>
Table 7a: Fixed effects of the multiple regression model fitted to gaze duration. The R² of the model is 0.322 and the standard deviation of the residual is 0.395. All numeric predictors were scaled.

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Length</td>
<td>0.062</td>
<td>0.010</td>
<td>6.235</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Joint Frequency</td>
<td>-0.032</td>
<td>0.009</td>
<td>-3.541</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Whole Compound Plausibility</td>
<td>-0.027</td>
<td>0.009</td>
<td>-2.956</td>
<td>0.004</td>
</tr>
<tr>
<td>Bias x Joint Frequency</td>
<td>-0.016</td>
<td>0.006</td>
<td>-2.778</td>
<td>0.005</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Standard Deviation</th>
<th>Correlations between by-participant slopes and intercepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound</td>
<td>0.064</td>
<td></td>
</tr>
<tr>
<td>Participant</td>
<td>0.127</td>
<td></td>
</tr>
<tr>
<td>Presentation format by participant</td>
<td>0.091</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Total Reading Time

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>6.015</td>
<td>0.037</td>
<td>160.624</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>ART</td>
<td>-0.143</td>
<td>0.035</td>
<td>-4.114</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Bias</td>
<td>-0.011</td>
<td>0.008</td>
<td>-1.380</td>
<td>0.168</td>
</tr>
<tr>
<td>Spelling Format (reference level = spaced)</td>
<td>-0.041</td>
<td>0.016</td>
<td>-2.502</td>
<td>0.012</td>
</tr>
<tr>
<td>Word Length</td>
<td>0.072</td>
<td>0.011</td>
<td>6.450</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Joint Frequency</td>
<td>-0.048</td>
<td>0.011</td>
<td>-4.230</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Trial Number</td>
<td>-0.044</td>
<td>0.011</td>
<td>-4.036</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Whole Compound Plausibility</td>
<td>-0.066</td>
<td>0.010</td>
<td>-6.310</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Right Family Frequency</td>
<td>0.028</td>
<td>0.014</td>
<td>2.056</td>
<td>0.041</td>
</tr>
<tr>
<td>Spelling Format x Right Family Frequency</td>
<td>-0.029</td>
<td>0.014</td>
<td>-1.998</td>
<td>0.046</td>
</tr>
<tr>
<td>Bias x ART</td>
<td>-0.017</td>
<td>0.007</td>
<td>-2.550</td>
<td>0.011</td>
</tr>
</tbody>
</table>
Bias x Joint Frequency | -0.019 | 0.006 | -3.249 | 0.001

Table 8a: Fixed effects of the multiple regression model fitted to total reading time. The $R^2$ of the model is 0.357 and the standard deviation of the residual is 0.387. All numeric predictors were scaled.

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Standard Deviation</th>
<th>Correlations between by-participant slopes and intercepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound</td>
<td>0.082</td>
<td></td>
</tr>
<tr>
<td>Participant</td>
<td>0.189</td>
<td></td>
</tr>
<tr>
<td>Trial Number by participant</td>
<td>0.045</td>
<td>0.31</td>
</tr>
</tbody>
</table>

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