Spelling errors impede recognition of correctly spelled word forms

Running Head: Spelling errors impede word recognition

Sadaf Rahmanian[a,*]

Victor Kuperman[a]

[a] Department of Linguistics and Languages
McMaster University
1280 Main Street West
Hamilton, Ontario, Canada, L8S 4M2

*Corresponding author: email rahmas28@mcmaster.ca

Abstract

Spelling errors are typically thought of as an *effect* of a word's weak orthographic representation in an individual mind. What if existence of spelling errors is a partial *cause* of effortful orthographic learning and word recognition? We selected words that had homophonic substandard spelling variants of varying frequency (e.g., *innocent* and *inocent* occur in 69% and 31% of occurrences of the word). Conventional spellings were presented for recognition either in context (Experiment 1, eye-tracking sentence reading) or in isolation (Experiment 2, lexical decision). Words elicited longer fixation durations and lexical decision latencies if there was more uncertainty (higher entropy) regarding which spelling is a preferred one. The inhibitory effect of frequency was not modulated by spelling or other reading skill. This finding is in line with theories of learning that predict spelling errors to weaken associations between conventional spellings and the word's meaning.

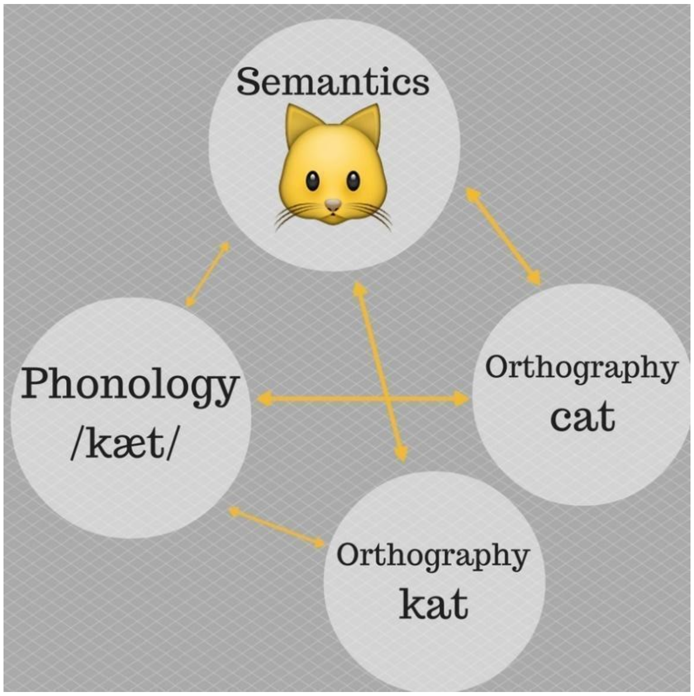
Keywords: spelling, information theory, reading, individual differences

This paper focuses on spelling errors, a phenomenon of written language that is typically thought of as an *effect* of inadequate word knowledge, a footprint of a word's weak orthographic representation in an individual mind (Perfetti, 1985; 2007). We consider the possibility that existence of spelling errors is a partial *cause* of effortful orthographic learning and word recognition. In what follows, we present a theoretical background for our hypothesis and report two studies examining the influence of incorrectly spelled English words (e.g., *comit*) on recognition of their correctly spelled counterparts (*commit*).

A lower-quality mental representation of a word implies that either some formal or semantic constituents of the word are not sufficiently specified in an individual's mind, or mappings between them are too weak (Nelson Taylor & Perfetti, 2016; Perfetti, 1985; 2007; Perfetti & Hart, 2001; 2002; Stanovich & West, 1989). One challenge for developing crisp lexical representations and word learning is a lack of stability in a word's spelling, sound, or semantics (Perfetti & Hart, 2001). We examine how the lack of orthographic stability caused by naturally occurring homophonic spelling errors influences learning and recognition of correctly spelled English words in isolation and in context.

Spelling errors that do not alter the pronunciation of correctly spelled words (e.g., *beginning* vs *begining*, or *receive* vs *recieve*) present a special variant of homophony. Unlike typical homophones (*ate* and *eight*) they share not just phonology but also meaning. Moreover, unlike pseudo-homophones (*spais* vs *space*), they occur in naturally produced written language with variable frequencies. Thus, homophonic spelling errors instantiate a kind of a pure orthographic alternation that can also be found in English compounds (e.g., *girlfriend*, *girl-friend* and *girl friend*, see Falkauskas & Kuperman, 2015; Kuperman & Bertram, 2013) or regional spelling variants (*color* vs *colour*). Figure 1 adapts the well-known triangle model of word representation (see Harm & Seidenberg, 2001, Perfetti, 2007, and Plaut et al., 1996) for meaning-preserving homophonic orthographic alternations.

Figure 1: *The Triangle Model of Reading with orthographic variants*



Why would the presence of an incorrectly spelled word affect learning of that word when spelled correctly? As speech production research demonstrates, alternative forms generate their own mental representations which are stored simultaneously in the mental lexicon and compete with one another (see review by Ernestus, 2014). This suggests that spelling errors "steal exposure" from correctly spelled words. Every encounter with one orthographic alternative (e.g., incorrectly spelled *asymetric*) in a context where another alternative (e.g., correctly spelled *asymmetric*) might occur comes at the expense of encountering the latter alternative (Perfetti & Hart, 2002). Since exposure to an orthographic form is a central vehicle of orthographic learning (Andrews & Hersch, 2010; Hersch & Andrews, 2012), spelling errors diminish opportunities to learn and entrench a fully-specified orthographic representation and strengthen the mental representation of an error instead. Another reason derives from theories of learning and indicates that the impact of errors transcends that of reduced exposure. Under the Naïve Discriminative Learning theory (Baayen, Milin, Durdevic, Hendrix, & Marelli, 2011;

Ramscar, Dye, & McCauley, 2013; Rescorla & Wagner, 1972), the strength of a connection between a cue (a word's spelling) and the outcome of learning (the word's meaning) is attenuated when another cue can point to the same meaning. The more often the form *asymetrical* is found to mean "lopsided, uneven", the weaker the association becomes between this meaning and the correct form *asymmetrical*, and the more uncertainty the reader experiences when mapping word forms and meanings.

An underlying assumption for our hypothesis is that readers are sensitive to relative frequencies with which correctly and incorrectly spelled words occur in language (see Protopapas, Fakou, Drakopoulou, Skaloumbakas, & Mouzaki, 2013). Recent studies of another orthographic alternation – the presence or absence of a space in English compounds – provides evidence that this sensitivity might exist. Falkauskas and Kuperman (2015) have found that compounds have been read faster and with fewer fixations and regressions in an eye-tracking study if they were presented in the form that was more frequent in natural language. Similarly, lexical decision latencies to concatenated compounds (*ballpoint*) with a stronger bias towards concatenation (Kuperman & Bertram, 2013).

The Lexical Quality Hypothesis makes further predictions that are applicable to the present study. Distributional patterns of orthographic alternatives are expected to have a stronger effect in situations that afford more opportunity to learn those patterns, e.g., in more frequent words and in readers with a greater exposure to print. Less proficient readers may not be aware that an orthographic alternation exists in the first place, due to their limited exposure to orthographic variants of an alternating word; this is especially true of lower-frequency words with alternating orthography (see motivation in Falkauskas & Kuperman, 2015).

Based on the prior literature, our current prediction is that readers would experience the most difficulty learning and recognizing words when uncertainty of choosing between alternative spellings is the strongest, e.g., when there are many alternative spellings and/or when their relative frequencies of occurrence are more even. A word that is spelled uniformly (conventionally or not) is not expected to lead to an excessive learning or recognition effort. Not only processing efficiency to be higher in words whose spelling is less uncertain, but we also expect to further increase in words that occur more frequently (in all spelling variants) and in more experienced readers.

We operationalize uncertainty between spelling variants using an information-theoretic measure of entropy (for worked examples see Milin, Kuperman, Kostić, & Baayen, 2009). Entropy *H* over a probability distribution is defined as:

$$H = -\sum_{i=1}^{n} p_i log\, p_i,$$

where *p* is the relative frequency (probability) of each spelling variant *i* in the summed frequency of all *n* variants of the word. Entropy is a non-negative measure of an average amount of effort associated with selecting one of the available variants given their probability distribution. Entropy is low when there is one of the variants is dominant, and high when there are many variants or when available variants are of similar probabilities. For instance, the word *innocent* occurs 141,960 times (69%) in this form, and 62,665 (31%) as *inocent*, yielding a relatively high value of entropy H = 0.89. Words that mostly occur in one form (e.g., *necessary* 98% vs neccessary 2%) yield relatively low values of entropy (H = 0.14).

In information-theoretic terms, our predictions can be reformulated as follows. In words that are misspelled in English, we expect a higher entropy (i.e. a greater uncertainty) of spelling variants to correlate with a greater effort of recognizing a correctly spelled word. We further predict the effect of entropy to be stronger in words of a higher rather than lower total frequency. A final prediction is that more experienced readers will be more sensitive to entropy than less experienced readers because they will have had more exposure to both correct and incorrect variants in print.

While our hypothesis advocates a causal link between one's exposure to spelling errors and one's effort of recognizing correctly spelled words, we cannot rule out either an opposite or a reciprocal nature of causality[1]. For instance, the very existence of misspellings may be a result of a poor, imprecise phoneme-to-grapheme mapping in a conventionally spelled word (e.g., receive). Thus, an apparent effect of spelling entropy may mask a penalty that comes with this lack of precision. Finally, both the exposure to spelling errors and the quality of the word's phoneme-to-grapheme mapping might contribute to word recognition behavior. A clear distinction – to be pursued in future research – could be achieved in an artificial word learning experiment where participants are exposed to homophonic spelling variations with different rates of exposure and are administered a post-training spelling assessment.

In sum, the present paper examines the influence of "spelling entropy" (a measure of competition between a word's spelling variants) on the word recognition effort. Experiment 1 studies word recognition in sentence context using eye-tracking, and Experiment 2 focuses on word recognition in isolation and uses the lexical decision task. We employed the two experimental paradigms to test whether context and different task demands on the depth of semantic and orthographic processing would affect the existence and strength of interactions between entropy and frequency. Finally, given known discrepancies between tasks (Kuperman, Drieghe, Keuleers, & Brysbaert, 2013), finding effects of spelling entropy on both would speak to the reliability of our results.

Experiment 1: Eye-Tracking

In this experiment, participants saw correctly spelled words embedded in semantically neutral context while their eye-movements were recorded. The central manipulation was the amount of entropy in the probability distribution of the correct and incorrect variants of target words. In both Experiments 1 and 2 below, readers were only exposed to correctly

[1] We thank an anonymous reviewer for raising and elaborating on this point.

spelled words (except for typos, see below). Thus, any effect related to incorrect variants of those words would be implicit and can only stem from the readers' awareness of distributional patterns of orthographic variants in the language.

Method

Stimuli:

The stimuli consisted of 70 correctly spelled target words. The frequencies of the correct forms of the target words and of their single most common misspellings were identified from a 7-billion token USENET corpus of unedited electronic communication (Shaoul & Westbury, 2013). For simplicity, we opted for selecting words that had only one relatively common misspelling, while other potential misspellings of that word were negligibly infrequent. There was no correlation between the frequencies of the correct and incorrect forms. This ensured that these words did not just occur misspelled more often in the corpus because they were overall more frequently used.

The correct variants of each of the 70 words were then embedded into syntactically simple, context-neutral sentences. All sentences occupied exactly one line on the screen. The target word was not the first or last word in the sentence. An example of a sentence with the correctly spelled target word *asymmetric* is given below: the entire list of stimuli is available in Supplementary materials S1.

He was bothered by the asymmetric painting in the office.

The sentences were displayed on a 17-inch monitor with a resolution of 1600 x1200 pixels, and a refresh rate of 60 Hz. Sentences were presented one at a time in a Courier New, a monospace font, size 22, in black on a white background, and occupied exactly one line on the screen. Each character subtended 0.36 degree of visual angle.

Participants:

A total of 35 undergraduate McMaster students participated in this study (24 females, mean age of 20.9). The participants were all native speakers of English, and reported either normal or corrected-to-normal vision. None of the participants indicated that they had any reading or learning disabilities. Participants were compensated with course credit.

Procedure:

Participants were asked to read the sentences silently while their eye movements were recorded. After they finished reading the sentence, they were to press any key on the keyboard. A quarter of the trials were followed by a yes-or-no comprehension question to ensure that participants were in fact reading the sentences. Participants could record their responses by the ‘a’ key for true and ‘’’ key for false. These keys were marked with a green and a yellow sticker respectively to make them easier to spot.

Participants sat 60 cm away from the screen. A chin rest was used to reduce movement during the experiment. Participants would undergo a three-point horizontal calibration and validation prior to beginning any data collection. Data collection would only begin once a good validation (maximum error below 1 degree of visual angle) was achieved. Although participants read naturally, the eye movement data was only collected from one eye using an eye-tracker Eyelink 1000 (SR Research, Kanata, ON, Canada).

Each trial began with a single point shift correction to increase the accuracy of eye-movements. The sentences would appear individually. Before the experiment began, there were four practice trials to familiarize the participants with the process. All participants read the same set of sentences but the sentence order was randomized.

After the eye-tracking portion of the study was complete, the participants were given a a spelling recognition and spelling dictation test (Andrews & Hersch, 2010; see also Andrews & Lo, 2013), two tests of exposure to print – the Author Recognition Test (ART) and Magazine Recognition Test (MRT) (Stanovich & West, 1989) – and the Test of Word

Reading Efficiency (TOWRE) (Torgeson et al., 1999). Details on administration and scoring of the tests are available in Supplementary materials S2. An unrelated eye-tracking study was run during the same block as this study. In total, the participants spent approximately 40-50 minutes completing the experiment.

Variables

The critical independent variable was entropy of the probability distribution formed by the frequency of the correct spelling of the word and its single most common misspelled counterpart (*until vs untill*), see above for definition of entropy. In all words, frequency of the word's conventional spelling was much higher than that of its erroneous spelling, making lower values of entropy more prevalent than its higher values in our sample: see Table 1 for descriptive statistics.

We also considered as a predictor frequency of occurrence of the correctly spelled word (i.e. the word form that readers were exposed to). Both main effects of entropy and frequency and their interaction were critical tests of our predictions above. All frequency counts were based on the 7-billion token USENET corpus (Shaoul & Westbury, 2013). We considered word length in characters as a major determinant of reading behavior. We also included the length and log-transformed frequency of the words preceding and following the target: none of these affected eye-movements to the target words and are not reported further. A final set of independent variables that we took into account were individual scores in the offline skill tests described above, ART, MRT and the Test of Reading Efficiency. Since only the ART test showed any effects on the eye-movement record, we do not report other tests.

The dependent variables examined were: first fixation duration (the duration of the first fixation on the word), single fixation duration (the duration of the first and only fixation on the target word), gaze duration (the sum of all fixations before leaving the word for the first time), and total reading time (the sum of all fixations on the word), as well as the likelihood of a regression (look-back or left-bound saccade to the target word). The eye-movement record enables a fine-grained analysis of the time-course of word processing, with first fixation duration as the initial measure of word decoding and lexical access, regression rate the measure of sentence reanalysis, 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 (Boston, Hale, Kliegl, Patel, & Vasishth, 2008). Table 1 reports descriptive statistics for dependent and independent variables.

Table 1: Descriptive statistics of independent and dependent variables.

Variable	Range	Mean	Median	SD	Range of transformed values
Word length	3:12	8.294	8	1.955	
Word frequency (correct spelling)	1117: 1946830	180687	80138	313448	3.048:6.289
Entropy	0.061:0.954	0.437	0.403	0.258	
Author recognition test (ART)	1:41	10.74	7.50	8.951	
First fixation duration	81:919	219	205	77	4.394:6.823
Single fixation duration	81:919	220	205	80	4.394:6.823
Gaze duration	81:996	269	232	125	4.394:6.904
Regression rate	0:1	0.18	0	0.385	
Total reading time	81:1097	334	289	175	4.394:7.000

Statistical considerations

We made use of generalized additive mixed effects models (GAM models; Hastie & Tibshirani, 1990; Wood, 2006) as implemented in the mgcv package 1.8-7 (Wood, 2006; 2011) of the R statistical computing software (R Core Team, 2015). Unlike a linear regression model, in a GAM model the functional relation between a predictor and the response variable need not be linear (Baayen, Kuperman, & Bertram, 2010). Instead, the GAM enables a flexible smoothing of nonlinear relations in any number of dimensions. In a GAM, multiple predictors may be combined in a single smooth, yielding either a nonlinear functional relationship (between one independent variable and a dependent variable), a wiggly surface (when two independent variables are combined) or a wiggly hypersurface (when three or more independent variables are combined). Our critical interactions (entropy x frequency and entropy x skill) used tensor products to approximate complex surfaces formed by two independent variables and a dependent one. We opted for GAM models to avoid imposing a specific (linear) nature of the relationship between critical predictors and outcomes. These nonlinear interactions are not reported in an ANOVA-style (as a combination of main effects and an interaction term) but rather as parameters characterizing the shape of the smoothed surface and its inferential statistics.

To visualize a wiggly three-dimensional surface we opted for presenting its two-dimensional cross-sections as effects of one predictor on a dependent variable at fixed levels (25% and 75% percentiles) of another predictor. These visualizations are made on the basis of models reported below and use the function *plot_smooth* in *itsadug* package (van Rij, Wieling, Baayen, & van Rijn, 2015). For detailed description and worked examples of the use of generalized mixed-effects additive models in psycholinguistics see Balling and Baayen (2012), Matuschek, Kliegl, and Holschneider (2015).

We log-transformed continuous dependent variables and word frequency. Models with continuous dependent variables (e.g., gaze duration and total reading time) used the underlying Gaussian distribution, while binary variables (regression rate) used the binomial distribution. To attenuate the influence of outliers, we removed datapoints with absolute scaled residuals exceeding 2.5 standard deviations, and refitted each model after trimming: these are the models reported below. Inclusion of random effects (including by-item and by-participant splines for critical experimental effects) was permitted when it improved the performance of the model, as indicated by the *compareML* function of the *itsadug* library (van Rij et al., 2015). In the interest of space, we only reported models where critical predictors and interactions reached statistical significance: other models are available upon request. Random effects were only maintained in the final model when significantly contributing to the model fit, as indicated by the model comparison test.

Results and Discussion

One of the participants was removed from data analysis as they scored less than 80% on the comprehension questions: the average accuracy was above 92%. Each of the participants read a total of 70 sentences. Two sentences were removed from the data analysis because of a mistake in their target words: the remainder of analysis only reports data from 68 target sentences. Only the eye-movements to the target words were considered. After this trimming, the data set consisted of a total of 2312 observations.

All trials in which the target word was not fixated during the first reading pass were removed (196 observations, 8.5%). In addition, trials in which first fixation durations that were shorter than 80 ms (43 observations, 1.9%) or in the top 1% of longest fixation durations (21 observations, 1.3%) were removed. This resulted in a total of 2052 observations for analysis.

A generalized additive model was fitted to total reading time on the target words and contained a critical interaction of entropy by word frequency and additional predictors: see Tables 2 and 3 below. The interaction was statistically significant ($p = 0.019$): we visualize the resulting complex surface in Figure 3 left panel by plotting effects of entropy on the eye-movement measure for fixed levels of word frequency. Thus, the solid line represents a partial effect of entropy on total reading time estimated for words in the 25^{th} percentile of frequency, and the dashed line for words in the 75th percentile of frequency. Figure 3 left panel reveals that entropy had no apparent effect on the word processing effort in lower-frequency words, with estimated total reading time fluctuating around 300 ms across the entire range of entropy. Conversely, words of a higher frequency were more sensitive to the effects of entropy, especially in the low entropy range (0-0.4). Estimated total reading time was at its lowest (260 ms) in higher-frequency words with zero entropy, i.e. words that had little uncertainty regarding the correct and incorrect word spelling. However, with an increase in entropy came an increase in total reading times and it reached about 310 ms (a contrast of 50 ms) in the mid-range of entropy. Furthermore, an advantage to higher-frequency words over lower-frequency ones was confined to a lower entropy range, since total reading times in both frequency bands were indistinguishable in a higher entropy range. There were no statistically significant effects of entropy or an interaction with entropy any other eye-movement measure. That the effect is restricted to total reading time suggests that the influence of spelling errors was late and was likely to emerge in a behavioral measure typically associated with re-analysis of read material and its semantic integration within a sentence (Boston et al., 2008; Rayner 1998).

Table 2. Parametric coefficients of the generalized additive mixed model fitted to log total reading times in Experiment 1. N after trimming = 2035 (N before trimming = 2052). Deviance explained = 23.4%.

	Estimated	Standard Error	t-value	p-value
Intercept	5.514	0.091	60.459	< 0.001
Length	0.029	0.009	3.101	0.002
ART	-0.007	0.003	-2.278	0.023

Table 3. Estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F and p values for the splines and random effects in the generalized additive mixed model fitted to log total reading times in Experiment 1.

	Edf	Reference df	F	P

Tensor product of frequency x entropy	6.071	6.572	2.577	0.019
Random by-participant intercepts	40.944	63.000	2.030	< 0.001
Random by-item intercepts	28.305	32.000	7.895	< 0.001

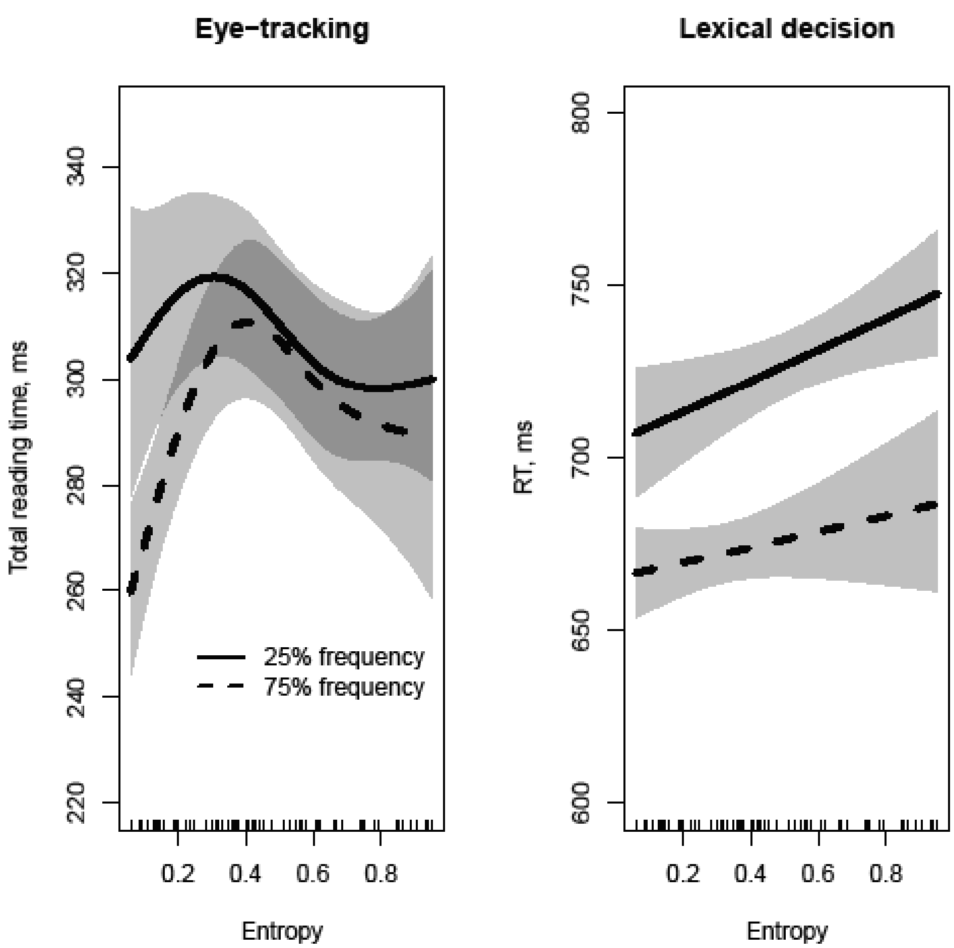


Figure 3. Partial effects of spelling entropy on total reading times (left panel) and lexical decision latencies (right panel), estimated for the 25% (solid line) and 75% (dashed line) percentiles of log word frequency.

This finding demonstrates that the statistical distribution of correct and incorrect spellings in one's mental lexicon is able to affect the effort of recognizing a word in context even when that word is correctly spelled. This implicit effect is particularly pronounced in high-frequency words, i.e. words that afford more opportunities to encounter all their spelling variants. Both observations are in line with our predictions above.

The ART/MRT, spelling recognition and spelling dictation tests were completed by all of the participants. The ART score correlated negatively and reliably with total reading times: readers with greater exposure to print showed shorter fixation times ($b = -0.007$, $SE = 0.003$, $t = -2.278$, $p = 0.023$, see Table 2). No other measure of individual difference reliably predicted any of the eye movements. Moreover, none of the individual difference measures entered into a critical interaction with entropy of the orthographic alternation, or a three-way interaction with entropy and word frequency. This finding runs counter to our predictions of a great reading experience being associated with greater sensitivity to statistical patterns of orthographic alternations. The null effect of individual variability cannot be explained away by a relatively uniform reading experience in our cohort. In fact, compared to the normative data on university students by Moore and Gordon (2015) the ART scores in our cohort were both lower ($M = 10.74$ vs. 14.72) and, importantly, more variable ($sd = 8.95$ vs 7.32). We discuss this finding in the General Discussion.

We also tested the two-way interactions of entropy and word length, as well as the three-way interaction of entropy, word length and skill or entropy, word length and word frequency: none reached significance in any of the models.

Experiment 2: Lexical Decision Task

Experiment 1 uncovered the influence of spelling errors on recognition of correctly spelled words in context. Even though our target words were not predictable in experimental sentences, context might still have provided top-down syntactic cues that could make the readers' reliance on the orthographic form less critical than in a situation when a word is presented in isolation. This consideration motivated Experiment 2, which examined recognition effort of target words out of context. We expected patterns in lexical decision latencies to confirm the detrimental effect of spelling variability on the word processing effort (see Introduction).

Methods

Stimuli:

The target words used as stimuli in Experiment 2 were identical to the stimuli in Experiment 1. We considered the same independent variables as described above. The dependent variable of interest was the response time to the word in lexical decision task.

Procedure:

We made use of the lexical decision data in the English Lexicon Project mega-study (Balota et al., 2007). The English Lexicon Project (ELP) is a collection of normative data on lexical decision and word naming for 40,481 words and 40,481 non-words. Data from 816 participants across 6 USA universities was collected for the lexical decision task specifically. Each participant completed approximately 3400 responses: for details of procedure see Balota et al. (2007). No data on individual skills is made available with the ELP, and so we did not pursue an examination of individual variability in this study[2]. The age ranges and the educational levels (mostly undergraduate students) were comparable between cohorts tested in Experiments 1 and 2.

---

[2] A similar lexical decision mega-study (the British Lexicon Project; Keuleers, Lacey, Rastle, & Brysbaert, 2012) only contained 43 of our 68 target words, and so we did not analyze this additional dataset.

Results and Discussion

Before analysis, all of the data for non-words, inaccurate responses and outliers (as defined by Balota et al., 2007) were removed from the data set. Only the lexical decision data for the 68 target words from Experiment 1 was considered. There were 2134 observations after the trimming. Due to the design of the ELP data collection, most of participants contributed only one data point to this pool of observations, precluding an accurate estimation of inter-subject variability. We calculated average response times for every word, and analyzed contributions of predictors using a GAM model without random effects.

We fitted a GAM model to lexical decision response times with a critical interaction of entropy by log word frequency and additional controls. Model outcomes are visualized in Figure 2 right panel. The critical interaction was linear in shape and statistically reliable (Tensor product of frequency by entropy: Edf = 3, Reference df = 3, $F = 13.66$, $p < 0.001$): parametric coefficients are reported in Table 4. Similar to the results of the eye-tracking data, the solid line represents a partial effect of entropy on response times estimated for words in the 25th percentile of frequency. The dashed line represents a partial effect of entropy on response times estimated for words in the 75th percentile of frequency.

Table 4. Parametric coefficients of the generalized additive mixed model fitted to mean lexical decision latencies in Experiment 2. N after trimming = 67 (N before trimming = 68). Deviance explained = 52.6%.

	Estimated	Standard Error	t-value	p-value
Intercept	6.387	0.048	133.581	< 0.001
Length	0.024	0.006	4.238	< 0.001

The results were largely comparable to those found in Experiment 1. As expected, longer and lower-frequency words were processed slower. Critically, higher spelling entropy (i.e. a higher level of uncertainty about what the correct spelling of the word is) translated into longer response times across the entire range of entropy, and in both higher- and lower-frequency words. This evidence corroborates our hypothesis that the existence of spelling errors hampers our perception of lexical units, even when they are spelled correctly. It also demonstrates that the increased effort due to spelling uncertainty emerges both with and without the cues provided by the sentence context.

Unlike Experiment 1 and our hypothesis, lower-frequency words showed a slightly stronger positive effect of entropy on response times than higher frequency words (the difference in estimated RTs between the extreme values of entropy was on the order of 35 ms vs 20 ms, respectively). We discuss this discrepancy below.

General Discussion

The present paper explores the influence of spelling errors and their frequency of occurrence on the ease of word recognition. We argue that existence of alternative spellings creates an important challenge to word learning and recognition, especially when those alternatives are homophonic (e.g., *achieved* vs *acheived*), see Figure 1. As follows from theories of statistical learning and the Lexical Quality Hypothesis Perfetti (1985; 2007), the lack of stability in a word's orthographic representation may be harmful in two ways. First, an orthographic alternative like a spelling error steals exposure from the conventional spelling, and offers fewer opportunities to entrench that spelling. Second, one's experience with seeing a given meaning associated with multiple orthographic forms (e.g., the meaning *dweller* with *tenant* and *tennant*) weakens the strength of either orthographic form as a cue to that meaning. As a result, exposure to spelling errors – whether their own or made by others – is predicted to lead to unlearning the conventional spelling and learning the erroneous one, leading to a lower quality of the word's orthographic representation as well as a greater effort of spelling the word in its correct form.

We tested the theorized influence of spelling errors by selecting words that varied in their spelling entropy, a measure quantifying how frequently words occurred in their conventional spelling relative to a non-conventional spelling. We hypothesized that higher values of entropy would come with more effortful word recognition, especially in situations which offered more opportunities to encode all and any word spellings, i.e. words with higher frequency of occurrence across spelling formats, and individuals with greater exposure to print.

The results of our two experiments demonstrated that higher spelling entropy came with an increased effort of recognition of words both in sentence context (Experiment 1) and in isolation (Experiment 2). That is, words in which conventional and erroneous spellings had more similar frequencies of occurrence in natural language elicited longer eye-fixation times and longer latencies of the lexical decision response than words in which the conventional spelling was more dominant (see Kuperman & Bertram, 2013, for a similar finding). This finding is remarkable, given that either experiment only exposed participants to correctly spelled word forms, and no explicit reference to alternative spellings (i.e. spelling errors) took place either in the instructions or the stimuli[3]. Taken together, these observations suggest that readers possess gradient statistical knowledge of distributional patterns in language, even if these patterns are subtle: in the majority of words in our sample, incorrect spellings account for no more than 10% of the total frequency of the word's occurrence. Importantly, readers appear to rely on mappings between a meaning and multiple orthographic alternatives of a word regardless of whether words appear in context or in isolation, and even in the absence of a direct task demand (i.e., in lexical decision).

Reliable effects of the measure based on a mathematical function of both correct and incorrect form frequencies imply that the strength of a word's orthographic representation in an individual is affected by his or her exposure to *all* alternative spellings (Kuperman & Bertram, 2103). This is consistent with the notion advocated by the Naïve Discriminative Learning theory (Baayen et al., 2011) that frequencies of alternatives dynamically shape

[3] Lexical decision makes use of non-words, however, none of these were homophonic with existing words of the English language (Balota et al., 2007).

how well a word is learned. Existence of spelling errors undermines the quality of a conventional orthographic representation: the more common the error, the more a reader "unlearns" the correct spelling and the more effort is required to recognize that spelling. Our data demonstrate that this is true for relatively proficient readers (undergraduate students recruited for Experiments 1 and 2) who are likely to have strong awareness of what spelling is codified as conventional.

Our additional hypotheses were that spelling entropy would be more influential in higher-frequency words and more experienced readers. These hypotheses found mixed support in the data. While higher-frequency words showed a stronger effect of entropy on total reading times than lower-frequency words in Experiment 1, the opposite was true in Experiment 2. We speculate that the discrepancy is due to differences in the task demand. The goal of the lexical decision task is to differentiate between existing and non-existing strings of characters. Strictly speaking, lexical decision can be done purely on orthographic grounds (even if semantic effects are common in this task). This makes precision of the orthographic representation critical for the task. It stands to reason that higher-frequency words come with more familiar orthographic patterns, and so they can elicit a response faster and without as much recourse to stored exemplars of alternative spellings, as compared to less familiar lower-frequency words. Thus, the form of the interaction in Experiment 2 is similar to the well-documented interaction of word frequency and spelling-sound regularity, where low-frequency words benefit more from regular spelling and are penalized more by irregular one than higher-frequency words (see Hino & Lupker, 2000, Keuleers, Lacey, Rastle, & Brysbaert, 2010). However, sentence reading for comprehension (Experiment 1) is obviously contingent on retrieval of semantics of words in the sentence. Context may partly compensate the relative lack of orthographic precision in a target word, but the strength of the mapping between lexical orthography and semantics has more influence on the recognition effort. As we argued above, the mapping is likely to be more uncertain in higher-frequency words, which allow for more exposure and better encoding of all spelling alternatives: hence a stronger inhibitory effect of spelling entropy in higher- rather than lower-frequency words. Future research will need to pin down relative contributions of examine how the context and task demand to the influence of spelling errors.

Also, contrary to our predictions and findings of Falkauskas and Kuperman (2015) we did not observe an interaction between entropy and any individual difference measure in Experiment 1 (no participant-level data was available for Experiment 2). Even though our undergraduate participants varied in their reading proficiency (see above), it is possible that a greater variability is required to reveal the predicted interactions. So far, however, the absence of the skill x entropy interaction suggests that exposure to inconsistency in spelling affects all readers equally, even those whose own spelling is impeccable. Taken together, our observations support the use of (human or automatic) resources that standardize spelling, i.e. spellcheckers, language editors and proof-readers, and spelling and literacy instructors. The outcome of their work is not only a more consistent orthographic inventory of a language, but also more efficient word learning and recognition.

A future direction of this study is to engage populations varying in their age, educational background and reading proficiency, and their L1. It is also worthwhile to conduct this study with individuals that were taught spelling under different pedagogical methods. Of particular interest are consequences of the inventive spelling method, which encourages early production of written forms based on the word's sound and without regard for the conventional spelling. Pending this exploration, we refrain from formulating recommendations for educational practices for children or adults.

In sum, we demonstrated that the trajectory of learning a word's orthography is shaped by exposure to all spelling variants in which this word occurs. These findings enrich our present understanding of how word learning and word representation in the mental lexicon are connected. Moreover, they pave the way to the further examination of factors impeding and facilitating proficient word spelling and identification.

Acknowledgments

We wish to thank three anonymous reviewers and Raymond Bertram for their insightful comments on earlier drafts of this paper. This work was supported by the following research grants: The Canadian NSERC Discovery grant RGPIN/402395-2012 (Kuperman, PI), the Ontario Early Researcher award (Kuperman, PI), the Canada Research Chair (Tier 2; Kuperman, PI), the SSHRC Partnership Training Grant 895-2016-1008 (Libben, PI), and the NSERC USRA scholarship to the first author. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Canadian government. This work has been presented at the 10th International Mental Lexicon meeting (2016, Ottawa, Canada).

References

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General, 139(2)*, 299.

Andrews, S., & Lo, S. (2013). Is morphological priming stronger for transparent than opaque words? It depends on individual differences in spelling and vocabulary. *Journal of Memory and Language, 68(3)*, 279-296.

Baayen, R. H., Milin, P., Durdevic, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118 (3)*, 438–481.

Baayen, R. H., Kuperman, V., & Bertram, R. (2010). Frequency effects in compound processing. *Compounding, Amsterdam/Philadelphia: Benjamins*, 257-270.

Balling, L. W., & Baayen, R. H. (2012). Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition*, *125*(1), 80-106.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior research methods*, *39*(3), 445-459.

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

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

Falkauskas, K. and Kuperman, V. (2015). When experience meets language statistics: Individual variability in processing English compounds. Journal of Experimental Psychology. Learning, Memory and Cognition. 41(6),1607-1627.

Harm, M. W., & Seidenberg, M. S. (2001). Are there orthographic impairments in phonological dyslexia?. *Cognitive Neuropsychology*, *18*(1), 71-92.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models* (Vol. 43). CRC Press.

Hersch, J., Andrews, S. (2012). Lexical Quality and Reading Skill: Bottom-Up and Top-Down Contributions to Sentence Processing. *Scientific Studies of Reading, 16(3),* 40-262.

Hino, Y., & Lupker, S. J. (2000). Effects of word frequency and spelling–to–sound regularity in naming with and without preceding lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 26(1)*, 166.

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, *1*:1-174.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44*(1), 287-304.

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

Kuperman, V., Drieghe, D., Keuleers, E., and Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *Quarterly Journal of Experimental Psychology*, *66*, 563-580.

Matuschek, H., Kliegl, R., & Holschneider, M. (2015). Smoothing spline ANOVA decomposition of arbitrary splines: An application to eye movements in reading. *PloS One*, *10*(3), e0119165.

Milin, P., Kuperman, V., Kostic, A., & Baayen, R. H. (2009). Paradigms bit by bit: An information theoretic approach to the processing of paradigmatic structure in inflection and derivation. *Analogy in grammar: Form and acquisition*, 214-252.

Nelson Taylor, J. & Perfetti, C. A. (2016). Eye movements reveal readers' lexical quality and reading experience. *Reading and Writing*, *29(6)*, 1069 – 1103.

Perfetti, C. A. (1985). *Reading ability*. Oxford University Press.

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

Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill.

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

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological review*, *103*(1), 56.

Protopapas, A., Fakou, A., Drakopoulou, S., Skaloumbakas, C., & Mouzaki, A. (2013).

What do spelling errors tell us? Classification and analysis of errors made by Greek schoolchildren with and without dyslexia. *Reading and Writing*, *26*(5), 615-646.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of mouses in adult speech. *Language*, *89*(4), 760-793.

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

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, *2*, 64-99.

Shaoul, C., & Westbury, C. (2013). A reduced redundancy USENET corpus (2005-2011). *University of Alberta*, *39*(4), 850-863.

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

Torgeson, J. K., Wagner, R. K., & Rashotte, C. A. (1999). Test of word reading efficiency (TOWRE). *Austin, TX: ProEd*.

Van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2015). itsadug: Interpreting time series, autocorrelated data using gamms. *R package, version*, *1*(1).

Wood, S. N. (2006). On confidence intervals for generalized additive models based on penalized regression splines. *Australian & New Zealand Journal of Statistics, 48*(4), 445-464.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73(1)*, 3-36.