Running head: WORD FREQUENCY: A LINEAR EFFECT?

The word frequency effect during sentence reading:
A linear or nonlinear effect of log frequency?

Sarah J. White[1]

Denis Drieghe[2]

Simon P. Liversedge[2]

Adrian Staub[3]

[1] Department of Neuroscience, Psychology and Behaviour, University of Leicester, UK

[2] School of Psychology, University of Southampton, UK

[3] Department of Psychological and Brain Sciences, University of Massachusetts Amherst, USA

Corresponding Author: Sarah J. White

Department of Neuroscience, Psychology and Behaviour, University of Leicester, University Road, Leicester, LE1 7HA, UK

s.j.white@le.ac.uk

Word count excluding References, Tables and Figure Legends: 4674

Acknowledgements: We would like to thank Vanessa Srawley and David Walters for their help with data collection. The authors would also like to thank Eyal Reingold and an anonymous reviewer for their helpful comments on a previous version of the manuscript.

Abstract

The effect of word frequency on eye movement behaviour during reading has been reported in many experimental studies. However, the vast majority of these studies compared only two levels of word frequency (high and low). Here we assess whether the effect of log word frequency on eye movement measures is linear, in an experiment in which a critical target word in each sentence was at one of three approximately equally-spaced log frequency levels. Separate analyses treated log frequency as a categorical or a continuous predictor. Both analyses showed only a linear effect of log frequency on the likelihood of skipping a word, and on first fixation duration. Ex-Gaussian analyses of first fixation duration showed similar effects on distributional parameters in comparing high and medium frequency words, and medium and low frequency words. Analyses of gaze duration and the probability of a refixation suggested a nonlinear pattern, with a larger effect at the lower end of the log frequency scale. However, the nonlinear effects were small, and Bayes Factor analyses favoured the simpler linear models for all measures. The possible roles of lexical and post-lexical factors in producing nonlinear effects of log word frequency during sentence reading are discussed.

Keywords: word frequency, eye movements, ex-Gaussian

Rayner and Duffy (1986) and Inhoff and Rayner (1986) first demonstrated that reading times for words within sentences are longer for low compared to high frequency words (controlling for word length). Since then, many studies have shown effects of word frequency on eye movement behavior during reading (see White, 2008). However, although a few experimental (Murray & Forster, 2008; Radach, Huestegge, & Reilly, 2008) and some corpus (Cop, Keuleers, Drieghe, & Duyck, 2015; Kliegl, Nuthmann & Engbert, 2006; Kuperman, Drieghe, Keuleers, & Brysbaert, 2013; Kuperman & Van Dyke, 2013) eye movement studies have reported word frequency effects across a range of values, most experimental studies have simply compared two levels of frequency.

Graded word frequency effects have been shown across a range of single word tasks (Monsell, Doyle, & Haggard, 1989) and the effect of word frequency on lexical decision times has been shown to be approximately logarithmic (Howes & Solomon, 1951; Keuleers, Lacey, Rastle, & Brysbaert, 2012; but see Adelman & Brown, 2008; Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Murray & Forster, 2004; Yap & Balota, 2009). When absolute frequency is used as a predictor variable (as measured in number of occurrences per million words of text), its effect is clearly nonlinear: The difference in response time between words with very low and moderate absolute frequency is much larger than is the difference in response time between words with moderate and high absolute frequency. Indeed, although words with extremely low or high frequencies may show a different pattern (see Keuleers, Diependaele, & Brysbaert, 2010; Murray & Forster, 2008), in general response time does appear to depend in an approximately linear fashion on the log of a word's absolute frequency value.

In line with this work, models of eye movement control during reading use log word frequency to simulate the effects of word frequency on reading times and fixation probabilities (Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Kliegl, & Richter, 2005; Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003). However there is some indication that effects of log word frequency on eye movement behavior may not be linear. Both Wotschack and Kliegl (2013) and Kuperman and Van Dyke (2013) found, in analyses of eye movement corpora, significant nonlinear effects of log word frequency on reading time measures, with larger effects of log frequency occurring at the lower end of the range than at the higher end. In other words, the log transform may not be extreme enough to fully account for the pronounced effect of encountering a low frequency word (see also Keuleers et al., 2010). However, it is important to note that corpus analyses include a broad range of words, from various syntactic categories, and that words of varying frequencies that occur in the texts used in these corpus studies are not matched on other lexical characteristics and occur in uncontrolled contexts.

The present study investigates whether there are nonlinear effects of log frequency in a controlled experiment where words of varying frequencies occur in matched sentence frames, and where other variables are controlled. In particular, we assessed the effects of word frequency on eye movement behavior for nouns varying in length from four to six letters. Word frequencies were calculated on the Zipf scale, based on the SUBTLEX-UK corpus (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). The Zipf scale is logarithmic, ranging from very low (Zipf value 1) to very high (Zipf value 7) frequency words; it is calculated as log10 (frequency per million words) + 3, or equivalently log10 (frequency per billion words). This scale has a number of desirable properties, discussed by Van Heuven et al. (2014). The words that we use in the present study are almost entirely in the Zipf range from 3 to 6; we did not explore reading of words with very low Zipf frequencies of 1 or 2 (an issue we revisit in the Discussion). We explored the effect of log word frequency in two ways. First, we conducted analyses comparing three experimental conditions (high vs. medium, medium vs. low). Second, we treated Zipf frequency as a continuous predictor, assessing whether there was a significant quadratic effect of this variable in regression models of the various eye movement measures.

The effect of word frequency was also explored here by fitting the ex-Gaussian distribution (Ratcliff, 1979) to first fixation durations for the high, medium and low frequency words. The ex-Gaussian distribution is the convolution of a normal and an exponential distribution and can be characterized by three parameters: the mean of the normal component (μ), the standard deviation of the normal component (σ) and the mean and standard deviation of the exponential component (τ). Staub, White, Drieghe, Hollway, and Rayner (2010) demonstrated that a manipulation of word frequency produces both a shift in the location of the distribution for low frequency words (μ parameter) and increases the degree of skew (τ parameter) (see also Reingold, Reichle, Glaholt, & Sheridan, 2012). The present study tests if these differences hold for high compared to medium, and medium compared to low frequency words.

Finally, although the vast majority of studies that have examined word frequency effects during reading have employed monocular eye movement recordings, some have recorded binocularly (Jainta, Blythe, & Liversedge, 2014; Jainta, Nikolova, & Liversedge, 2016; Juhasz, Liversedge, White, & Rayner, 2006; Kirkby, Blythe, Drieghe, Benson, &

Liversedge, 2013; Kliegl et al., 2006; Nikolova, Jainta, Blythe, Jones, & Liversedge, 2015; Nikolova, Jainta, & Liversedge, 2016). However, none of these studies restricted analyses of word frequency effects to cases for which critical words were skipped or fixated by both eyes (for similar restrictions see Kliegl et al., 2006). The two eyes are not always perfectly aligned during reading (see Kirkby et al., 2013 for a review) and therefore the two eyes occasionally fixate two different words. For the measure of word skipping it is particularly important to verify that effects of word frequency hold for cases for which both eyes skipped the critical word. Therefore, in the present study word skips and reading times were calculated only for cases for which the critical word was skipped or first fixated with both eyes.[1]

Method

Participants. Thirty-three members of the University of Leicester community completed the study. Participants were replaced if they did not complete the experiment (due to failing visual acuity or calibration criteria). All participants were naïve regarding the purpose of the experiment, and were native English speakers with no history of reading disorders.

Apparatus. Sentences were presented on a ViewSonic P227fb monitor with a refresh rate of about 7ms (150Hz) at a viewing distance of 80cm. Movements of both eyes were recorded using an EyeLink 1000 eye tracker (SR Research Ltd.). Pupil location was sampled at a rate of 1000Hz. Sentences were presented on a single line in Courier New bold font, the text was black on a light grey background. Approximately three characters subtended one degree of visual angle.

Materials and design. There were 60 high, 60 medium and 60 low frequency critical words. All were nouns between four and six letters long ($M = 5.3$, $SD = 0.65$). The high frequency words had significantly higher Zipf values ($M = 5.24$, $SD = 0.30$) than the medium frequency words ($M = 4.32$, $SD = 0.32$) ($t(118) = 16.05$, $p < 0.001$), and the medium frequency words had significantly higher Zipf values than the low frequency words ($M = 3.36$, $SD = 0.32$) ($t(118) = 16.30$, $p < 0.001$)[2]. The three classes of words were almost perfectly evenly spaced on the Zipf scale, with a difference of .92 between the high frequency and medium frequency means, and a difference of .96 between the medium and low frequency means. The within-item difference scores (high vs. medium, medium vs. low) were not significantly different from each other ($t < 1$). Each set of critical words was embedded in the same neutral sentence frame up to and including the word after the critical word. Twelve different participants completed a cloze task: they were provided with the experimental sentences up to the critical word and were asked to guess words that could fit as the next word in the sentence. None of the guesses were correct, demonstrating that the critical words were not predictable from the initial sentence context. Example sentences are shown in Table 1 and the full stimulus set is available in the supplementary materials.

Insert Table 1 here

Participants read all 180 experimental sentences, distributed across three blocks following a Latin square design. Each block began with four practice trials. The order of the experimental items was randomized for each participant within each block. The order of the blocks was counterbalanced across participants. 34% of sentences were followed by a comprehension question.

Procedure. Visual acuity was tested separately for each eye using a Bailey Lovie chart (Bailey & Lovie, 1976). Participants were instructed to read the sentences for comprehension and to respond "yes" or "no" to questions by pressing buttons on a game controller. A chinrest and forehead rest minimized head movements. The eye tracker was calibrated monocularly for each eye using a three point horizontal calibration (other eye was occluded), with monocular calibration checks at least every third trial. Recalibrations were undertaken when necessary (ensuring maximum spatial error $<0.3^{\circ}$).

Analyses. Fixations less than 80 and more than 1200ms were excluded. First-pass reading behavior is reflected in the likelihood of skipping a word, and fixations on that word, before moving to the right or left of it. 5.2% of trials were excluded due to blinks before, during, or after any first-pass (left or right eye) fixation on the critical word. For word skipping, the critical word was categorized as skipped only if it was skipped on first-pass with both eyes. For the other first-pass measures, trials were only included if the initial first-pass fixation on the critical word was with both eyes (~72% of cases). Three measures are reported: First fixation durations, gaze durations (the sum of first-pass fixations on the word), and refixations (the proportion of trials for which the critical word was fixated more than once on first-pass). For brevity, single fixation durations are not reported, but the pattern was the same as for first fixation durations which was not surprising given that re-fixations were quite rare (see Table 2). These measures were analyzed using linear mixed effects models (Baayen, Davidson, & Bates, 2008) using R (R Core Team, 2016) and the lme4 package (Bates, Maechler & Bolker, 2011). Participants and items were treated as crossed random effects. A maximal random effects structure was employed (Barr, Levy, Scheepers, & Tily, 2013), with random subject and item intercepts, and random subject and item slopes as justified by the design. Values of *t*/*z* values greater than 1.96 are considered significant. Word skipping and refixations (binary data) were analyzed using logistic models. Two sets of analyses were undertaken. First, two contrasts were programmed, one comparing high and medium frequency words, and another comparing medium and low frequency words. These analyses were undertaken on both the raw data and log-transformed data. As the pattern of results was the same, only results for the raw data are reported. A second set of analyses further explored if the effects of word frequency were linear by treating word frequency as a continuous predictor. These models included the centered Zipf value, and the square of this value to test for a nonlinear (i.e., quadratic) effect. This second set of analyses was undertaken only on the raw data and with maximum likelihood estimation (rather than restricted maximum likelihood (REML)). In addition to directly assessing the statistical significance of the quadratic term, we compared the mixed effects models that did and did not contain this term by computing an approximate Bayes Factor based on the Bayesian Information Criterion (BIC) for each model (i.e., the exponent of the half the difference in BIC between the models; Kass & Raftery, 1995; Wagenmakers, 2007).

Insert Table 2 here

The distribution of first fixation durations was also examined by fitting the ex-Gaussian distribution. The same data restrictions were applied as above. The fitting procedure requires a large number of observations per participant per condition. Therefore, five participants with less than 30 observations in more than one condition were removed (analyses based on 28 participants). The QMPE software (Cousineau, Brown, & Heathcote, 2004; Heathcote, Brown, & Mewhort, 2002) was used to determine best-fitting parameters, with all procedures the same as those employed by Staub et al. (2010).

Results

All participants scored more than 90% correct on the comprehension questions (M = 97%). Mean eye movement measures are reported in Table 2 (see also Footnote 3). The Linear mixed effects model for the effects of three levels of word frequency are reported in Table 3 and the effects of centered Zipf word frequency are reported in Table 4. Figure 1 shows the item means for each measure as a function of Zipf word frequency.

Insert Tables 3 and 4 and Figure 1 here

Word skipping. The contrasts comparing word skipping for the medium vs. high / low frequency conditions were not significant. However, an additional contrast did show that there were more skips for high compared to low frequency words (b = -0.59, se = 0.23, t = -2.53). Importantly, the model testing the effect of centered Zipf word frequency showed an effect of the linear term, but not the quadratic term. Furthermore, the results demonstrate that effects of word frequency on word skipping hold for cases for which both eyes skip the word.

First fixation duration. First fixation durations were significantly shorter for high compared to medium, and medium compared to low frequency words. Similar to the results for word skipping, the model with centered Zipf word frequency showed an effect of the linear term but no effect of the quadratic term, consistent with a linear effect of Zipf word frequency on first fixation durations.

The effect of word frequency on the distribution of first fixation durations was explored by fitting the ex-Gaussian distribution. Table 5 shows the mean best fitting ex-Gaussian parameters for each of the conditions. The μ parameter was significantly smaller for high compared to medium, $t(27) = 2.74$, $p < 0.05$, and for medium compared to low frequency words, $t(27) = 3.44$, $p < 0.01$. The τ parameter was significantly smaller for high compared to medium words, $t(27) = 2.33$, $p < 0.05$. The τ parameter was numerically smaller for medium compared to low frequency words, but this difference was not significant, $t(27) = 1.00$, $p = 0.325$. There was a significant effect of the σ parameter for medium compared to low, $t(27) = 2.56$, $p < 0.05$, but not high compared to medium frequency words, $t(27) = 1.23$, $p = 0.229$. In order to test if the effect of word frequency was different for the μ and τ parameters for high compared to medium and medium compared to low frequency words, the difference scores for each of these comparisons was calculated for each participant. A one sample t-test was used to test if the difference between these difference scores was significantly different from zero. Importantly, the test showed no difference for either the μ, $t(27) = 1.13$, $p = 0.269$, or τ ($t < 1$) parameters. Thus, while the high vs. medium difference in the τ parameter was significant and the medium vs. low difference was not, a direct test did not suggest that these differences were different from each other. These results are therefore consistent with a linear effect of Zipf word frequency on both the location and skew of the distribution of first fixation durations.

Insert Table 5 here

The results for the ex-Gaussian parameters are consistent with the vincentile plot (Ratcliff, 1979; Vincent, 1912) shown in Figure 2. The points on this plot represent the mean of the subject means within successive bins of the distribution of fixation durations; the shortest 10% of observations are in the first bin, the next shortest 10% of observations are in the second bin etc. A shift in the distribution is reflected in the separation of the curves, and a difference in the weight of the right tail is reflected in greater separation of the curves at the higher vincentiles. The similar degree of separation between the high vs. medium and medium vs. low frequency conditions across all of the vincentiles is, again, consistent with a linear effect of Zipf word frequency on both the location and the skew of the fixation duration distribution.

Insert Figure 2 here

Gaze duration and refixations. Gaze durations were significantly shorter for high compared to medium and medium compared to low frequency words[4]. The proportion of trials with a refixation was significantly higher for low compared to medium frequency words, but there was no difference between high and medium frequency words. For gaze durations the model with centered Zipf word frequency produced a significant linear effect, but also a significant effect of the quadratic term, and for refixations the quadratic term approached significance ($p = 0.07$).

In contrast to word skipping and first-fixation durations, these results suggest non-linear effects of Zipf word frequency on gaze duration and the probability of a refixation. The results are consistent with the pattern shown by the Loess curves for these measures in Figure 1. Note the flattening of the curves for higher Zipf values, consistent with a larger effect of Zipf word frequency for the low vs. medium frequency words compared to the medium vs. high frequency words. It is important to note, however, that although there is a statistically reliable nonlinear effect of log word frequency for gaze duration, this effect is quite small compared to the linear effect; the parameter estimates from the mixed effects model are 6.44 ms and -29.20 ms, respectively. It is also notable that the nonlinearity for refixation probability may arise partly due to a floor effect, given that refixations are already very unlikely for words that have Zipf frequency between 4 and 5.

Bayes Factor analysis. The Bayes Factor analysis provides an additional test of whether there are nonlinear effects of log word frequency. This analysis favored the model without the quadratic term in all cases, even in gaze duration (skipping BF in favor of the smaller model: 70.11; first fixation duration: 54.60; gaze duration: 4.71; refixation: 13.46). In other words, though the effect of the quadratic term on gaze duration reaches the conventional standard for statistical significance, the data may actually be regarded as more likely under the simpler model, when the more complex model is penalized for its additional complexity. The Bayes Factor analysis also strengthens the (null) conclusion that there is only a linear effect of log word frequency on skipping and first fixation duration. (See Abbott & Staub, 2015, for discussion of Bayes Factors in the context of eye movement data.)

Discussion

The present study demonstrates the expected graded effects of word frequency on first-pass eye movement behaviour. The main question this study was designed to address is whether these effects were linear, when log word frequency was used as a predictor, or whether, as suggested by some corpus studies (e.g., Wotschak & Kliegl, 2013; Kuperman & Van Dyke, 2013) effects are stronger at the lower end of the frequency scale even after a log transformation of word frequency has already been applied. In the present experimental study, in which word length and context were matched across frequency classes, the effects of log word frequency on word skipping and first fixation duration appeared to be linear. In contrast, there was a significant non-linear trend in gaze duration, such that the influence of log word frequency was especially large at the lower end of the frequency scale. The pattern for gaze duration is similar to that previously shown for corpus data (Kuperman & Van Dyke, 2013). The proportion of refixations shows a similar nonlinear pattern. Nevertheless, the nonlinear effects for gaze duration and the proportion of refixations were small. Moreover, the Bayes Factor analyses favoured the simpler linear models for all measures.

The pattern for word skipping and first fixation duration is consistent with the approximately logarithmic effects of word frequency shown for lexical decision times (Howes & Solomon, 1951). Furthermore, the present study showed linear effects of word frequency on the location and skew of the distribution of first fixation durations. Note that Staub and Benatar (2013) suggested that differences in the exponential (τ) parameter may reflect the frequency of processing disruption. If the effect of word frequency on the skew of the distribution is determined by the frequency of processing disruption, then the present study indicates that this occurs across a range of word frequencies, not just for low frequency words.

For gaze durations, it appears that while there may be a nonlinear effect of log word frequency, the effect is small. Note, however, that the present study did not include very low frequency words with Zipf frequencies of 1 or 2, corresponding to frequencies per million words of about .01 and .1, respectively. Van Heuven et al. (2014) provide *antifungal* and *harelip* as examples of words with Zipf frequencies of about 1, and *airstream* and *outsized* as examples of words with values of about 2. The lowest frequency words used in this study, with Zipf frequencies of about 3, occur about once per million words. It is entirely possible that more pronounced nonlinear effects would be present in the eye movement record if such very low frequency words were included. However, such words are by definition encountered very rarely indeed, and nonlinearities in this range would have little implication for understanding how word frequency modulates eye movement behavior in "normal" reading.

If first fixation duration is indeed a linear function of log word frequency, but gaze duration and refixation probability vary non-linearly with log word frequency, this dissociation requires some explanation. One possibility is that first fixation duration and the probability of a refixation are differentially influenced by lexical processing. The E-Z Reader model suggests that programming of a refixation is initiated at the start of a fixation on a word. The likelihood of initiating such a programme may be greater for longer words (Reichle et al., 2003) or may increase with saccadic error (distance between fixation position and the word centre) (Reichle, Warren, & McConnell, 2009). Crucially, refixations are less likely to occur on high frequency words because lexical processing completes quickly, cancelling the labile programme to refixate the word. One possibility is that lexical processing for both high and medium frequency words is likely to complete during the labile refixation programme, such that the refixations are cancelled, whereas for low frequency words lexical processing may be much less likely to complete until after the non-labile refixation programme. Model simulations seem necessary to further explore the effects of log word frequency on refixations, along with associated variables such as word length (see Pollatsek, Juhasz, Reichle, Machacek, & Rayner, 2008).

Another possibility is that the specific effects of encountering a low frequency word on refixation probability, and therefore gaze duration, may reflect differences in post-lexical processing. E-Z Reader includes a parameter (I) that reflects post-lexical integration difficulty (Reichle et al., 2009). Even if lexical and post-lexical processing stages are independent (see Staub, 2011), it may be more difficult to integrate the meaning of less familiar words into the sentence context. That is, low frequency words may on average be associated with higher “I” values than medium and high frequency words, which could perhaps result in much longer gaze durations for low frequency words and an overall nonlinear effect of word frequency. Note again, though, that formal computational simulations are necessary to examine these suggestions.

To summarise, the present study shows graded effects of word frequency on eye movement behavior. The linear pattern shown for first fixation durations and word skipping is consistent with the approximately logarithmic effects of word frequency shown in lexical decision (Howes & Solomon, 1951). For gaze duration, however, it appears that there is a small nonlinear effect of log word frequency, which might be explained by the mechanisms underlying programming of refixations or effects of integration difficulty.

References

Abbott, M. J., & Staub, A. (2015). The effect of plausibility on eye movements in reading: Testing E-Z Reader's null predictions. *Journal of Memory and Language, 85,* 76-87. http://dx.doi.org/10.1016/j.jml.2015.07.002

Adelman, J.S. and Brown, G.D.A. (2008) Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review, 115,* 214–227. http://dx.doi.org/10.1037/0033-295X.115.1.214

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390-412. http://dx.doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Bailey, I.L., & Lovie, J.E. (1976). New design principles for visual acuity letter charts. *American Journal of Optometry and Physiological Optics, 53,* 740-745.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133,* 283–316. http://dx.doi.org/10.1037/0096-3445.133.2.283

Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., & Bolker, B. (2011). LME4: linear mixed-effects models using S4 classes. http://cran.R-project.org/package=lme4

Cousineau, D., Brown, S., & Heathcote, A. (2004). Fitting distributions using maximum likelihood: Methods and packages. *Behavior Research Methods, Instruments, & Computers, 36,* 742–756. http://dx.doi.org/10.3758/BF03206555

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency Effects in Monolingual and Bilingual Natural Reading. *Psychonomic Bulletin & Review, 22,* 1216 - 1234. http://dx.doi.org/10.3758/s13423-015-0819-2

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research, 42,* 621-636. http://dx.doi.org/10.1016/S0042-6989(01)00301-7

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777–813. http://dx.doi.org/10.1037/0033-295X.112.4.777

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review, 9,* 394–401. http://dx.doi.org/10.3758/BF03196299

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word probability. *Journal of Experimental Psychology, 41,* 401–410. http://dx.doi.org/10.1037/h0056020

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40,* 431–439. http://dx.doi.org/10.3758/BF03208203

Jainta, S., Blythe, H.I., & Liversedge, S.P., (2014). Binocular advantages in reading. *Current Biology, 24,* 526-530. http://dx.doi.org/10.1016/j.cub.2014.01.014

Jainta, S., Nikolova, M., & Liversedge, S.P. (2016). Does text contrast mediate binocular advantages in reading? *Journal of Experimental Psychology: Human Perception and Performance* (invited resubmission).

Juhasz, B.J., Liversedge, S.P., White, S.J., & Rayner, K. (2006). Binocular coordination of the eyes during reading: Word frequency and case alternation affect fixation duration but not fixation disparity. *Quarterly Journal of Experimental Psychology, 59,* 1614-1625. http://dx.doi.org/10.1080/17470210500497722

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795. http://dx.doi.org/10.1080/01621459.1995.10476572

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,* 1-15. http://dx.doi.org/10.3389/fpsyg.2010.00174

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012) The British Lexicon Project: lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavioral Research Methods, 44,* 287–304. http://dx.doi.org/10.3758/s13428-011-0118-4

Kirkby, J.A., Blythe, H.I., Drieghe, D., Benson, V., & Liversedge, S.P. (2013). Investigating eye movement acquisition and analysis technologies as a causal factor in differential prevalence of crossed and uncrossed fixation disparity during reading and dot scanning. *Behavioral Research Methods, 45,* 664-678. http://dx.doi.org/10.3758/s13428-012-0301-2

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General, 135,* 12-35. http://dx.doi.org/10.1037/0096-3445.135.1.12

Kuperman, V., Drieghe, D., Keuleers, E., & 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. http://dx.doi.org/10.1080/17470218.2012.658820

Kuperman, V., & Van Dyke, J.A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 802-823. http://dx.doi.org/10.1037/a0030859

Monsell, S., Doyle, M.C., & Haggard, P.N. (1989). Effects of word frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General, 118,* 43-71. http://dx.doi.org/10.1037/0096-3445.118.1.43

Murray, W.S. and Forster, K.I. (2004) Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review, 111,* 721–756. http://dx.doi.org/10.1037/0033-295X.111.3.721 .

Murray, W.S., & Forster, K.I. (2008). The rank hypothesis and lexical decision: A reply to Adelman and Brown (2008). *Psychological Review, 115,* 240-252. http://dx.doi.org/10.1037/0033-295X.115.1.240

Nikolova, M., Jainta, S., Blythe, H.I., Jones, M.O. and Liversedge, S.P. (2015). Vergence responses to vertical binocular disparity during lexical identification. *Vision Research, 106,* 27-35. http://dx.doi.org/10.1016/j.visres.2014.10.034

Nikolova, M., Jainta, S., & Liversedge, S.P. (2016). Using a dichoptic moving window presentation technique to investigate binocular advantages during reading. *Journal of Experimental Psychology: Human Perception and Performance* (invited resubmission).

Nuthmann, A., Engbert, R., & Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Research, 45,* 2201-2217. http://dx.doi.org/10.1016/j.visres.2005.02.014

Nuthmann, A., & Kliegl, R. (2009). An examination of binocular reading fixations based on sentence corpus data. *Journal of Vision, 9(5):31,* 1-28. http://dx.doi.org/10.1167/9.5.31

Paterson, K.B., McGowan, V.A., & Jordan, T.R. (2013). Aging and the control of binocular fixations during reading. *Psychology and Aging, 28,* 789-795. http://dx.doi.org/10.1037/a0033328 .

Pollatsek, A., Juhasz, B.J., Reichle, E.D., Machacek, D., & Rayner, K. (2008). Immediate and delayed effects of word frequency and word length on eye movements in reading: A reversed delayed effect of word length. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 726-750. http://dx.doi.org/10.1037/0096-1523.34.3.726

R Core Team (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research, 72,* 675-688. http://dx.doi.org/10.1007/s00426-008-0173-3

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446–461. http://dx.doi.org/10.1037/0033-2909.86.3.446

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity and lexical ambiguity. *Memory & Cognition, 14,* 191–201. http://dx.doi.org/10.3758/BF03197692

Reichle, E.D., & Drieghe, D. (2015). Using E-Z Reader to examine the consequences of fixation-location measurement error. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 262-270. http://dx.doi.org/10.1037/a0037090

Reichle, E.D., Liversedge, S.P., Pollatsek, A., & Rayner, K. (2008). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences, 13,* 115-119. http://dx.doi.org/10.1016/j.tics.2008.12.002

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105,* 125-157. http://dx.doi.org/10.1037/0033-295X.105.1.125

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26,* 445–476. http://dx.doi.org/10.1017/S0140525X03000104

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E–Z Reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16,* 1–20. http://dx.doi.org/10.3758/PBR.16.1.1

Reingold, E.M., Reichle, E.D., Glaholt, M.G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology, 65,* 177-206. http://dx.doi.org/10.1016/j.cogpsych.2012.03.001

Staub, A. (2011). Word recognition and syntactic attachment in reading: Evidence for a staged architecture. *Journal of Experimental Psychology: General, 140,* 407-433. http://dx.doi.org/10.1037/a0023517

Staub, A., & Benatar, A. (2013). Individual differences in fixation duration distributions in reading. *Psychonomic Bulletin & Review, 20,* 1304–1311. http://dx.doi.org/10.3758/s13423-013-0444-x

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1280–1293. http://dx.doi.org/10.1037/a0016896

Ŝvede, A., Treija, E., Jaschinski, W., & Krūmiņa, G. (2015). Monocular versus binocular calibrations in evaluating fixation disparity with a video-based eye-tracker. *Perception, 44,* 110-1128. http://dx.doi.org/10.1177/0301006615596886

Van Heuven, W.J.B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67,* 1176-1190. http://dx.doi.org/10.1080/17470218.2013.850521

Vincent, S. B. (1912). *The function of the viborissae in the behavior of the white rat* (Animal Behavior Monographs Vol. 1, No. 5). New York, NY: H. Holt.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problem of p values. *Psychonomic Bulletin & Review, 14,* 779–804. http://dx.doi.org/10.3758/BF03194105

White, S. J. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 205–223. http://dx.doi.org/10.1037/0096-1523.34.1.205

Wotschack, C., & Kliegl, R. (2013). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *The Quarterly Journal of Experimental Psychology, 66,* 548–562. http://dx.doi.org/10.1080/17470218.2011.625094

Yap, M.J. and Balota, D.A. (2009) Visual word recognition of multisyllabic words. *Journal of Memory and Language, 60,* 502–529. http://dx.doi.org/10.1016/j.jml.2009.02.001

Footnotes

[1] Measurement error will often result in some erroneous classification of word skips or fixations (Reichle & Drieghe, 2015). Defining skips and fixations binocularly (excluding cases for which the eyes fixate different words) is likely to minimise the proportion of erroneous classifications. However, these restrictions are also likely to result in an underestimation of the proportion of intended skips or fixations. Also note that fixations at either end of word regions are more likely to be associated with fixations on different words, even though only one word may be attended at any one time (Reichle, Liversedge, Pollatsek, & Rayner, 2009). Such fixations at word endings are also more likely to be mislocated (see Nuthmann, Engbert, & Kliegl, 2005).

[2] The orthographic familiarity of letter sequences within the critical words was controlled based on both position specific and nonposition specific bigram and trigram token frequency counts. These were calculated using the CELEX English word form corpus (Baayen, Piepenbrock, & Gulikers, 1995) and summed for each word (see White, 2008). There were no differences in these counts between the high and medium, or medium and low, word frequency conditions (ts < 1.42, ps > 0.16).

[3] Across all of the data, fixations were horizontally disparate by on average 1.39 characters (SD = 1.06) (excluding both fixations preceded/followed by blinks and with disparities > 2.5SDs greater than each participant's mean). Note that fixation positions were based on DataViewer output. Eye positions were therefore based on the average eye position during

the fixation, with relatively early demarcation of the start of the fixation. The reported average disparities may be larger compared to alternative data processing methods (see Kirkby et al., 2013) or recordings with binocular calibrations (Ŝvede, Treija, Jaschinski, & Krūmiņa, 2015). 43.6% of fixations were aligned (within one character width), 45.8% of fixations were crossed by more than one character width and 10.6% of fixations were uncrossed by more than one character width. The pattern of alignment is in line with previous studies that employed similar viewing conditions (Nuthmann & Kliegl, 2009; Kliegl et al., 2006; Paterson, McGowan, & Jordan, 2013; see Kirkby et al., 2013).

[4] Total time (sum of all fixations on the critical word) produced a similar pattern to gaze duration. Total times were significantly shorter for high (M = 282, SD = 161) compared to medium (M = 316, SD = 196) and medium compared to low (M = 367, SD = 227) frequency words. There was a trend for a nonlinear effect of word frequency, with particularly long total times in the low frequency condition, however the model with centered Zipf frequency produced no significant effect of the quadratic term (b = 7.56, se = 5.04, t = 1.50).

Table 1. Example Experimental Items.

Example	Word frequency	Sentence (critical word shown in italics)
1.	High	He knew that the small *room* would be really useful for storage.
	Medium	He knew that the small *fund* would be sufficient to cover the costs.
	Low	He knew that the small *crib* would be ideal for his baby nephew.
2.	High	She tried to open the little *game* that she had bought for her niece.
	Medium	She tried to open the little *trap* that had been set to catch the mouse.
	Low	She tried to open the little *clam* that she found on the beach.
3.	High	I was given a special *number* that I had to keep secret.
	Medium	I was given a special *ticket* that gave me unlimited travel for a week.
	Low	I was given a special *lotion* that I could use for my eczema.

Table 2. Mean Eye Movement Measures for the High, Medium, and Low Frequency Critical Words. Standard Deviations Shown in Parentheses.

Measure	High	Medium	Low
Skip (proportion) [a]	0.09 (0.28)	0.06 (0.24)	0.04 (0.21)
First fixation duration (ms)[b]	218 (62)	235 (76)	255 (85)
Gaze duration (ms) [b]	231 (78)	252 (99)	285 (117)
Refixation (proportion) [b]	0.07 (0.25)	0.08 (0.28)	0.15 (0.35)

Notes:

[a] Skip represents the proportion of trials for which the critical word was skipped by both eyes on first pass.

[b] First fixation duration, gaze duration and refixation measures include only trials for which the critical word was first fixated with both eyes on first pass.

Table 3. Linear Mixed Effect Model Statistics: Effects of Three Levels of Word Frequency.

		Word skip	First fixation duration	Gaze duration	Refixation
Intercept	*b*	-3.35	234.74	254.66	-2.54
	se	0.23	5.87	7.45	0.16
	t/z	-14.46 *	39.98 *	34.17 *	-15.86 *
High vs. Medium	*b*	-0.37	17.26	20.88	0.12
	se	0.21	3.88	5.59	0.22
	t/z	-1.80	4.45 *	3.74 *	0.56
Medium vs. Low	*b*	-0.21	20.03	34.41	0.80
	se	0.23	3.95	5.60	0.18
	t/z	-0.93	5.07 *	6.14 *	4.42 *

Note: * Denotes statistical significance ($t > 1.96$).

Table 4. Linear Mixed Effect Model Statistics: Effects of Centered Zipf Word Frequency.

		Word skip	First fixation duration	Gaze duration	Refixation
Intercept	*b*	-3.37	233.90	250.28	-2.62
	se	0.24	5.93	7.59	0.17
	t/z	-13.95 *	39.48 *	33.00 *	-15.30 *
Centered Zipf (linear)	*b*	0.26	-19.49	-29.20	-0.44
	se	0.12	2.76	3.81	0.09
	t/z	2.14 *	-7.06 *	-7.67 *	-5.18 *
Centered Zipf (quadratic)	*b*	0.03	1.17	6.44	0.17
	se	0.11	2.05	2.79	0.09
	t/z	0.33	0.57	2.31 *	1.85

Note: * Denotes statistical significance ($t > 1.96$).

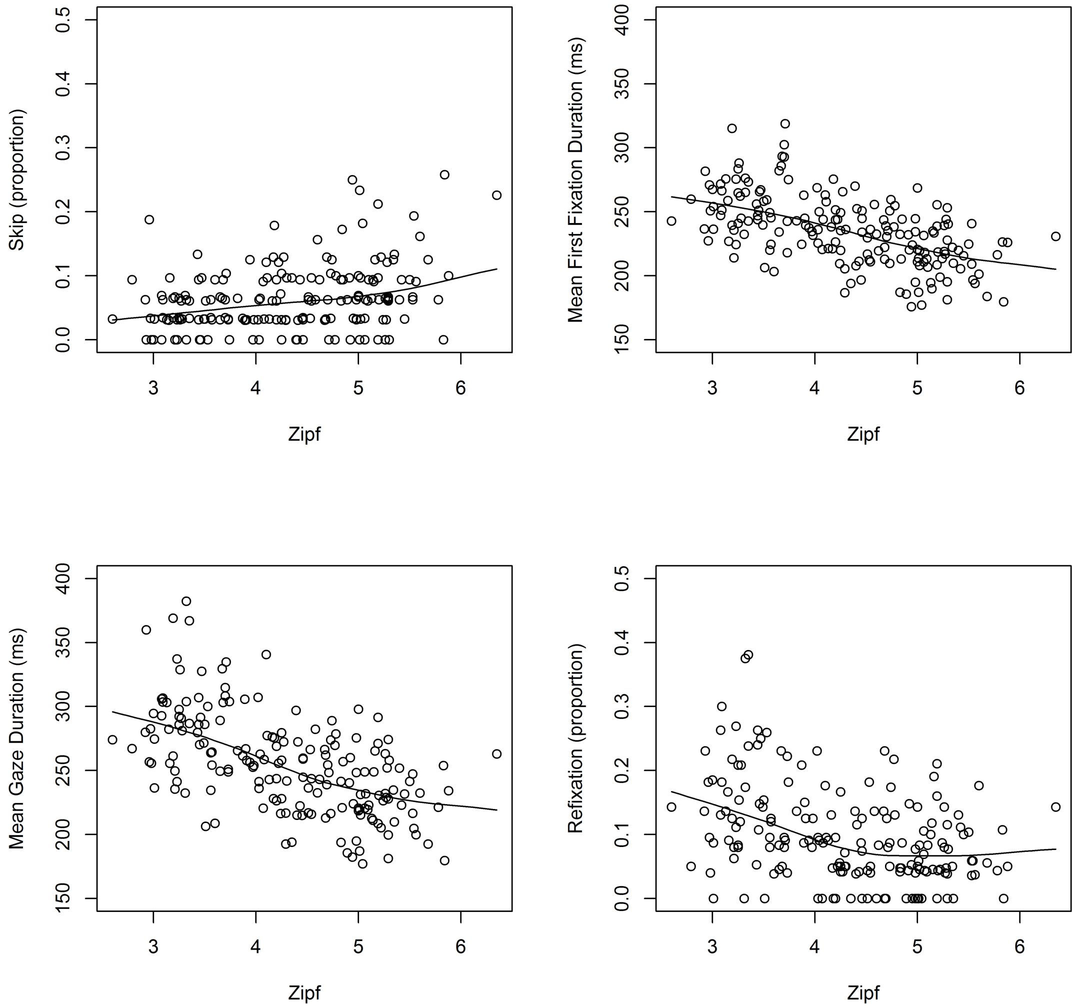
Table 5. Mean Best Fitting Ex-Gaussian Parameters for First Fixation Durations on the High, Medium and Low Frequency Critical Words.

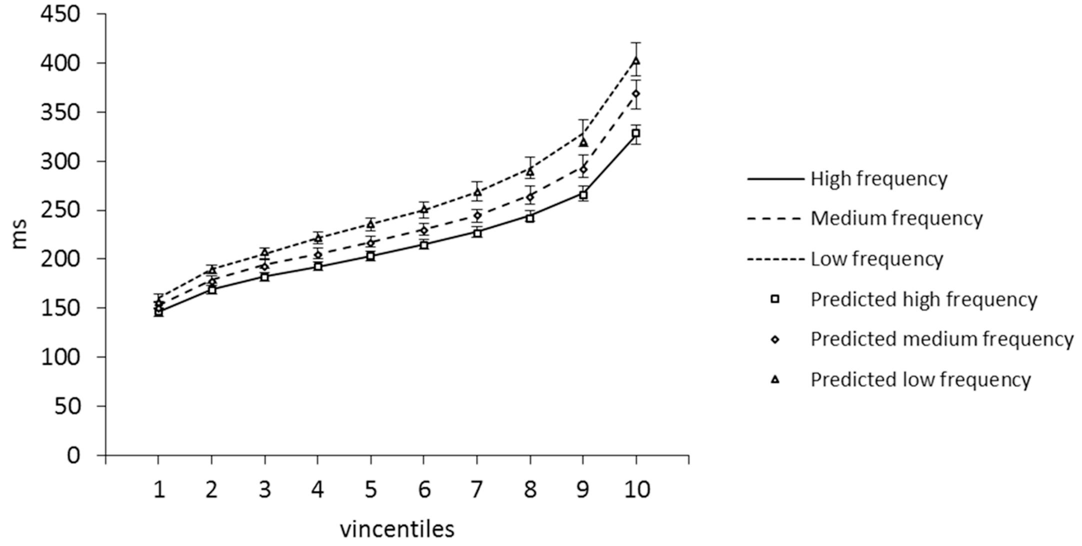
Measure	Frequency	μ	σ	τ
First fixation duration	High	171	26	46
	Medium	179	30	55
	Low	195	38	59

Figure captions

Figure 1. Scatterplot of item means and Zipf word frequency, with Loess curves,
measure.

Figure 2. Vicentile plot for first fixation durations on the critical high, medium a
frequency critical words. Error bars show the standard error of the mean. Predict
vincentiles are based on the mean of the best-fitting ex-Gaussian parameters and
random samples.


Skip (proportion)
0.0
0.1
0.2
0.3
0.4
0.5
3
4
5
6
Zipf
Mean First Fixation Duration (ms)
150
200
250
300
350
400
3
4
5
6
Zipf
Mean Gaze Duration (ms)
150
200
250
300
350
400
3
4
5
6
Zipf
Refixation (proportion)
0.0
0.1
0.2
0.3
0.4
0.5
3
4
5
6
Zipf


450
400
350
300
250
200
150
100
50
0
ms
1 2 3 4 5 6 7 8 9 10
vincentiles
High frequency
Medium frequency
Low frequency
Predicted high frequency
Predicted medium frequency
Predicted low frequency