Frequency and predictability effects for line-final words

Adam J. Parker[1] & Timothy J. Slattery[2]

Affiliations: [1] Department of Experimental Psychology, Division of Psychology and Language Sciences, University College London, London, WC1H 0AP; [2] Department of Psychology, Faculty of Science & Technology, Bournemouth University, Fern Barrow, Poole, BH12 5BB.

Correspondence: Correspondence regarding this article should be addressed to Adam J. Parker, Department of Experimental Psychology, Division of Psychology and Language Sciences, University College London, 26 Bedford Way, London, WC1H 0AP. Email: adam_parker@ucl.ac.uk.

Data availability statement: The materials and the data sets generated and analysed are available in the Open Science Framework (OSF) repository, https://doi.org/10.17605/OSF.IO/E4R2H. This repository also includes an R Markdown script to reproduce all analyses and generate the manuscript.

ORCID iD: Adam J. Parker: 0000-0002-1367-2282; Timothy J. Slattery 0000-0002-2652-289X

Abstract

Computational models of eye movement control during reading have revolutionized the study of visual, perceptual, and linguistic processes underlying reading. However, these models can only simulate and test predictions about the reading of single lines of text. Here we report two studies that examined how place-holders for lexical processing (frequency and predictability) influence the processing of line-final words. The first study was a linear mixed-effects analysis of the Provo Corpus, which included data from 84 readers reading 55 multi-line texts. The second study was a pre-registered eye movement experiment, where 32 participants read 128 items where frequency, predictability, and position (intra-line vs line-final) were orthogonally manipulated. Both studies were consistent in showing that reading times were shorter on line-final words. While there was mixed evidence for frequency and predictability effects in the Provo Corpus, our experimental data confirmed additive effects of frequency and predictability for line-final words which did not differ from those for intra-line words. We conclude that while models that make additive assumptions about the role of frequency and predictability may be better suited to modelling the current findings, additional assumptions are required if models are to be capable of modelling shorter reading times on line-final words.

Keywords: eye movements, reading, line-final words, return-sweeps, lexical processing.

Public Significance Statement

Our research adds to the growing body of work on return-sweeps during reading. Return-sweeps are the eye movements made at the end of a line and bring a reader's gaze to the start of a new line. Historically these eye movements have been understudied because eye movement studies typically present participants with single sentences. This work examined how input variables in computational models predict reading times for line-final words (words from which return-sweeps are commonly made). We report additive effects of frequency and predictability for line-final words. These findings are consistent with claims from the E-Z Reader model about the additive nature of these linguistic variables. This research complements earlier findings reported in the *Journal of Experimental Psychology: Human Perception & Performance* which suggest that, with minor additional assumptions, the E-Z Reader model may also be able to model reading times across line boundaries (Parker & Slattery, 2019).

Through the study of readers' eye movements, we have learned a great deal about the cognitive processes underlying sentence processing (Liversedge & Findlay, 2001; Rayner, 1998, 2009). For example, the implementation of gaze-contingent paradigms has indicated that readers extract meaningful information not only from the fixated word but also from the upcoming parafoveal word (e.g., McConkie & Rayner, 1975; Rayner, 1975). Benchmark findings such as these have been incorporated into computational models of eye movement control during reading, which make clear and testable predictions about how the eyes move through the text (see Engbert & Kliegl, 2011; Rayner, 2009b, Reichle, 2011, 2021 for reviews). While there is no doubt that these models have revolutionized the field, they are limited by the fact they have been fitted to data where participants have read single lines of text. As such, these models can only test predictions about single-line sentence reading[1]. Single-line reading is, of course, far removed from real reading. We read complex, multi-line sentences and paragraphs, and this presents a challenge to current accounts of the cognitive processes underlying reading. Therefore, in an attempt to better understand how readers process multiline texts, we conducted two eye movement studies which examined frequency and predictability effects for line-final words to inform the next generation of eye movement models that look to simulate eye movements across line boundaries. This work is critical as even with the proliferation of research on return-sweep saccades (the eye movement from the end of one line to the start of the next) and their effect on lexical processing across line boundaries, there still exists no model that allows for multi-line reading.

During reading, we make a series of rapid, ballistic eye movements (saccades) to bring visual information into high acuity foveal vision. The pauses between saccades, known as fixations, are when visual encoding of the text occurs. A plethora of eye movement research has fueled the argument that eye movements are under direct lexical control (Dambacher et al., 2013) and stages of lexical processing (e.g., lexical access) are what drive the eyes through the text (e.g., Liversedge & Findlay, 2001; Rayner et al., 1996). For instance, lexical variables such as word length, frequency of occurrence, and predictability from sentence context influence not only fixation durations but also the likelihood that a word is fixated (see Rayner, 1998, 2009, for reviews).

Reading times are shorter on highly frequent words (Angele et al., 2014; Inhoff & Rayner, 1986, Just & Carpenter, 1980; Kliegl et al., 2004; Miellet et al., 2007; Rayner et al., 2004; Rayner & Duffy, 1986; Slattery et al., 2007, 2012; Whitford & Titone, 2014). Reading times are also shorter on words that are highly predictable from the preceding sentence context (AlJassmi et al., 2022; Balota et al., 1985; Erlich & Rayner, 1981; Gollan et al., 2011; Rayner et al., 2011; Rayner & Well, 1996; Slattery & Yates, 2018). Moreover,

[1]Note that these models are also unlikely to be able to adequately model reading at the very start and the very end of a sentence given the sudden appearance of the sentence at the start of the trial will likely contaminate the first fixation of the trial and button press preparations will likely contaminate the final fixation. While the very first and last fixations during paragraph reading will be contaminated by the same artefacts, paragraphs will have sentences that do not receive trial initial or trial final fixations. Thus understanding paragraph reading will benefit our understanding of single-sentence reading also.

the probability of fixating a word is influenced by its frequency and predictability, with highly predictable words having greater skipping rates during first-pass reading (Brysbaert et al., 2005). More frequent words are also skipped more frequently, particularly when fixations land close to the start of the word (see Rayner et al., 2004, for a discussion).

Evidence of frequency and predictability effects on word skipping and first-pass fixation times indicate that these variables have an early influence on readers' eye movements. Findings from divergence point analyses (see Reingold & Sheridan, 2018, for a review) indicate that frequency and predictability effects emerge at 145 ms (Reingold et al., 2012) and 140 (Sheridan & Reingold, 2012) after the onset of a fixation respectively. Given that frequency and predictability both exert early influences on eye movement measures, the extent to which they interact has been debated (see Staub, 2015, for a review). For example, predictability effects have been hypothesized to be limited to low-frequency words as high-frequency words are already processed very rapidly. The experimental literature, however, is clear in that the effects of frequency and predictability on fixation duration are additive (Altarriba et al., 1996; Ashby et al., 2005; Kennedy et al., 2013; Miellet et al., 2007; Rayner et al., 2004; Rayner et al., 2001; Slattery et al., 2012; cf. Sereno et al., 2018[2]). While the joint effects of frequency and predictability on skipping are a little more complicated given mixed results (Gollan et al., 2011; Hand et al., 2010; Rayner et al., 2004), there exists no decisive evidence in favour of an interaction. Thus, it is safe to conclude that the two variables have an early effect on the decision of where and when to move the eyes, but these decisions are influenced by independent mechanisms.

Given the robustness of frequency and predictability effects on readers' eye movements, they are central to computational models of eye movement control during reading. One such model is the E-Z Reader (e.g., Reichle et al., 1998). At its core, E-Z Reader assumes that lexical processing and word identification drive the eyes through the text. E-Z Reader posits that attention is allocated to words in their printed canonical order such that words are identified in a strictly serial manner. As such, words are serially identified one after the other. E-Z Reader assumes two stages of lexical processing (*L1* and *L2*). *L1* represents an initial stage of

[2]Sereno et al. (2018) investigated the effects of target word frequency (low- vs high-frequency), predictability (low-, medium- vs high-predictability), and preview (valid vs invalid), where preview was varied between experiments. Importantly, target words in the high-predictability condition were of very high cloze probability (0.96 for low-frequency words and 0.97 for high-frequency words), which is a much higher value than those reported in previous studies (e.g., high-predictability words in Rayner et al., 2004, had a cloze probability of 0.78). Data from the valid preview experiment indicated a frequency by predictability interaction in first-fixation duration and single-fixation duration, where word frequency effects were absent in the high-predictability condition but present in the medium- and low-predictability conditions. Sereno et al.'s data, therefore, suggests that the frequency by predictability interaction may be observed under very high predictability conditions. However, this study is not without limitations. The vast majority of eye movement studies on prediction during sentence reading compare reading times on the same target words in different sentence contexts or different target words in the same context. By comparison, Sereno et al. compared reading times on different words presented in different sentence contexts. This arguably less controlled experimental design makes it difficult to compare the results of Sereno et al. with other studies. Therefore, while this study suggests that a frequency by predictability interaction can be observed under highly predictable conditions, there needs to be verification of this in an experimental study that compares more carefully controlled stimuli. Furthermore, the primary comparison of interest here is whether the frequency and predictability effects seen at intra-line locations are similar to those in line-final locations. If the effects differed appreciably across the locations then models would need to account for this.

lexical processing, called the familiarity check, which triggers the programming of a saccade. *L2* represents lexical access and triggers a shift of attention from the currently fixated word (*n*) to the upcoming word (*n+1*). Both *L1* and *L2* are influenced by frequency and predictability, with the two variables having an additive effect. Furthermore, E-Z Reader assumes that frequency and predictability influence the probability of fixating a word in an additive manner.

E-Z Reader can also explain skipping behaviour. The completion of the familiarity check on the fixated word (*n*) initiates a saccade program to *n+1*. *L2* then continues on *n* until it is identified. This is followed by a shift of attention to *n+1* and lexical processing for *n+1* begins. Because of the decoupling of the eyes and attention that is necessary for lexical processing, lexical processing of *n+1* can begin in the parafovea before it is directly fixated. This parafoveal processing is sometimes sufficient to complete the familiarity check for *n+1* before the saccade program to *n+1* is ready. As a result, the saccade to word *n+1* will be cancelled and a new saccade program to word *n+2* begins. Due to the time-consuming cancellation and reprogramming of saccades, E-Z Reader predicts inflated fixation times on word *n* prior to skipping word *n+1* (i.e., a skipping cost). Thus, E-Z Reader assumes skipping costs to be a consequence of word skipping (see Reichle & Drieghe, 2012, for further discussion).

To date, the role of frequency and predictability within computational models of reading has only ever been considered for the reading of single lines of text. Therefore, if we are ever to model the reading of multiline texts, it will be essential to better understand how these two variables operate across line boundaries. Before focusing on frequency and predictability across line boundaries, we briefly summarize relevant literature on return-sweeps.

To navigate between lines readers make return-sweeps, which are saccades that direct a reader's gaze from the end of one line to the start of the next. Return-sweeps are typically launched from five to seven characters from the end of the line (Hofmeister et al., 1999; Parker, Slattery, et al., 2019; Rayner, 2009; Slattery & Vasilev, 2019). The distance traversed by a return-sweep is largely determined by the layout of the text, with longer lines requiring longer return-sweeps. There is substantial variability in where fixations land following a return-sweep with landing positions being shifted towards the right for longer lines (Hofmeister et al., 1999; Parker, Nikolova, et al., 2019; Parker & Slattery, 2021) and for text displayed in larger fonts (when distance is measured in visual angle; Vasilev et al., 2021).

Like any saccade, return-sweeps are prone to systematic and random error (McConkie et al., 1998). Return sweeps have been reported to undershoot their target on 40-60% of occasions and require an immediate corrective saccade towards the left margin (Slattery & Vasilev, 2019). The rate of undershoot error is again determined by characteristics of the text, such as line length (e.g., Parker & Slattery, 2021) and line spacing

(Christofalos et al., 2023), as well as by reader-level characteristics (i.e., reading skill; Parker, Slattery, et al., 2019; Parker & Slattery, 2021) and task demands (Adedeji et al., 2021). Due to the two trajectories of return-sweeps, the fixations following a return-sweep can be grouped into two fixation populations: accurate line-initial fixations (where the line-initial fixation is followed by a rightwards saccade) and under-sweep fixations (where the line-initial fixation is followed by a leftwards saccade a regression or refixation, before a rightwards pass)[3].

In a reanalysis of the Provo Corpus (Luke & Christianson, 2018) and an eye movement experiment, Parker and Slattery (2019) tested several predictions about the nature of frequency and predictability effects that were derived from a modified E-Z Reader framework. Parker and Slattery assumed that if no lexical processing for the first word on a new line can occur until there is a fixation on the new line that places the first word within the fovea or parafovea then, from E-Z Reader's standpoint, a return-sweep may be viewed as any other inter-word saccade with the exception that the shift of attention to the first word of the next line would not result in the start of parafoveal pre-processing of this word, due to it being located in the periphery. As such, lexical processing (*L1*) of line-initial words must wait for these words to be both attended and located in the fovea or parafovea. With only a single additional assumption, Parker and Slattery stated that this modified framework would predict that: (1) the duration of the line-initial fixation following an accurate return-sweep should be longer compared to words fixated during left-to-right reading pass; (2) fixation times on line-initial words would be reduced if preceded by an undersweep-fixation due to the possible availability of preview benefit provided by these fixations; and (3) the effects of word frequency and predictability would remain the same as for other words. A pattern of results that were consistent with predictions (1) and (2) was observed in both data sets. Furthermore, the eye movement experiment showed (3) clear evidence of additive frequency and predictability effects for line-initial words and their analysis of the Provo Corpus indicated that the effects of frequency and predictability did not differ between intra-line and line-initial words. The fact that the data aligned with these predictions illustrates the potential capability of a modified E-Z Reader framework to predict the influence of frequency and predictability on reading times for line-initial words. Hence, when investigating the effects of frequency and predictability for line-final words in the current work, we again derived predictions from the E-Z Reader model.

While research endeavours have commenced to understand frequency and predictability effects at the start of a line, there is no previous study (to our knowledge) that has looked to understand how these variables

[3]Note that some studies define under-sweeps as line-initial fixations followed by an inter-word leftwards eye movement (e.g., Parker et al., 2020) while others use more relaxed criteria where under-sweeps are defined as line-initial fixations followed by either inter- and intra-word leftwards eye movement (e.g., Parker & Slattery, 2021). Studies that use the inter-word definition are typically concerned with word-level analyses while studies using both inter- and intra-word leftwards eye movements to define under-sweeps are typically focused on character-level information.

jointly impact the processing of line-final words. Parker, Slattery, et al. (2019) reported that skilled adult readers fixated 75% of line-final words and that readers' return-sweeps are not always initiated from the line-final word. Instead, only 67% of return-sweeps come from line-final words. Fixations prior to a return-sweep have been termed line-final fixations. These fixations are typically shorter than intra-line reading fixations (e.g., Abrams & Zuber, 1972; Parker, Nikolova, et al., 2019), as are reading times on line-final words (Tiffin-Richards & Schroeder, 2018). Two general accounts have been put forward to explain this phenomenon: the return-sweep planning account and the parafoveal processing account.

The return-sweep planning account of shorter line-final fixation durations stems from findings where there is a general speed-up as readers move across a line of text (Kuperman et al., 2010)[4]. A tentative suggestion from this evidence is that the line-final fixation serves the purpose of preparing the oculomotor system to shift a reader's gaze a large distance to the start of a new line. Consistent with this, Hofmeister (1997) reported that following a 50% degradation of the text there was a 20 ms increase in duration for all reading fixations other than line-final fixations, suggesting that line-final fixations are relatively uninvolved in linguistic processing. If line-final fixations, which are often made from line-final words, are uninvolved in lexical processing, then we might expect that the typical frequency and predictability effects observed in single-line reading may be absent for line-final words (particularly in cases where return-sweeps are made from these). This would result in an interaction in statistical models comparing lexical predictors across intra-line and line-final words; necessitating additional assumptions within computational models of eye movement control. Of course, the conclusion that line-final fixations are uninvolved in lexical processing may seem somewhat premature given the argument that eye movements are under direct lexical control (c.f. Liversedge & Findlay, 2001). An alternative account is one that instead focuses on parafoveal processing. Reader argues that fixations prior to word skipping are longer and that readers incur skipping costs. Thus, the absence of an opportunity to engage in parafoveal processing may eliminate the opportunity to engage in skipping and result in shorter line-final fixations. Estimates of skipping costs range greatly, with some estimates being sizable (e.g. 84 ms; Pynte et al., 2004) and others negligible (2 ms; Reichle & Drieghe, 2013). If the true effect of skipping costs exists within these bounds, then reduced skipping costs may be able to capture the differences in fixation duration that we see for line-final fixations. At current, there is no strong evidence base from which we can tease these explanations apart.

Here, we introduce the novel suggestion that E-Z Reader's assumptions about post-lexical integration may help explain the reduced line-final fixations. Integration can fail if word *n* is not successfully integrated with the sentence before the identification of word *n+1* occurs. This type of failure has important implications

[4]Note that although Kuperman et al. (2010) observed speed-up effects across a line of text, they removed line-initial and line-final fixations from their analysis of paragraph data, so suggestions here are based on the general trend across a line.

for the processing of line-final words as, in these cases, the identification of word *n+1* (the first word of a new line) will be delayed until after the execution of the return-sweep saccade (Parker & Slattery, 2019). Therefore, integration failures should be less likely for line-final words than for intra-line words and the resulting time costs associated with reprogramming saccades back to the location of the integration failure should be reduced leading to shorter line-final fixations. Evidence of such reduced integration failures can be assessed by comparing refixation rates and regression rates from intra-line fixations and line-final fixations, which we examine in our exploratory analyses. Both accounts derived from the E-Z Reader framework would predict additive effects of frequency and predictability effects for line-final words and a null interaction when comparing these lexical effects between intra-line and line-final words. Of course, given the shorter time course of reading times on line-final words, the effects of frequency and predictability may be attenuated for line-final words and this could result in statistically significant differences when comparing lexical effects between intra-line and line-final words.

Models of eye movement control use word frequency and predictability as language input variables to simulate the reading of single lines of text. In the hope of extending these models to the reading of multiline texts, it is essential to first understand how these input variables influence the processing of line-final words. To be clear, our goal is not to assess whether E-Z Reader (or a competitor model) can accurately predict the observed data as there is currently no model of eye movement control that allows for multiline reading. Instead, our goal is to provide benchmark findings that will be of importance for future modeling efforts. In the current work, we report two eye movement studies of frequency and predictability effects for line-final words. Specifically, we compared the effects of frequency and predictability for intra-line and line-final words, that is regardless of whether they were the word from which a return-sweep was made or not. The first study is a corpus-style analysis of the Provo Corpus (Luke & Christianson, 2018). The second is a pre-registered eye movement experiment involving 32 participants who read 128 stimuli where frequency, predictability, and position of the target word were orthogonally manipulated within participants. Borrowing from E-Z Reader's additive assumption about frequency and predictability, we anticipated additive effects of frequency and predictability for intra-line reading. Furthermore, under the assumption that reduced skipping costs or reduced failures of integration are responsible for shorter line-final reading times, then we may also assume that E-Z Reader's assumptions about the additive effects of frequency and predictability would hold for line-final words. However, given the argument of reduced lexical processing for line-final words, it also remains conceivable the effects of frequency and predictability may differ between intra-line and line-final words although explanations derived from E-Z Reader would likely predict highly similar effects of frequency and predictability acoss these locations. Demonstrating consistent and comparable effects across the two

Table 1: Mean word length, zipf Frequency, and cloze predictability for all words in the Provo Corpus, line-final words, analysed line-final words, and analysed intra-line words. Pearson correlation estimates are reported for each dataset.

Data set	Variable	Mean (SD)	Range	Length	Frequency	Predictability
Full corpus	Length	4.76 (2.55)	1 - 19	-		
	Frequency	5.70 (1.43)	1.17 - 7.67	-0.801	-	
	Predictability	0.41 (0.23)	0.05 - 1.00	-0.263	0.295	-
Line-final words	Length	5.15 (2.95)	1 - 19	-		
	Frequency	5.46 (1.47)	2.28 - 7.67	-0.784	-	
	Predictability	0.45 (0.26)	0.07 - 1.00	-0.181	0.192	-
Analysed line-final	Length	6.28 (2.10)	4 - 12	-		
	Frequency	4.56 (1.04)	2.32 - 6.45	-0.503	-	
	Predictability	0.33 (0.20)	0.07 - 0.95	-0.025	0.148	-
Analysed intra-line	Length	6.37 (2.01)	4 - 12	-		
	Frequency	4.65 (1.01)	1.17 - 7.19	-0.568	-	
	Predictability	0.34 (0.20)	0.05 - 1.00	-0.112	0.090	-

approaches (corpus and experimental) would provide compelling evidence for either outcome in naturally occurring corpus of written language and in experimentally manipulated items. However, to preempt our results, this would not be the case. Instead, our corpus analysis would provide only robust evidence for shorter reading times on line-final words while our experimental work would provide strong evidence for both shorter reading times on line-final words and additive effects of frequency and predictability on line-final words.

Eye movement corpus analysis

We first examined frequency and predictability effects for line-final words via a linear mixed-effects analysis of the Provo Corpus (Luke & Christianson, 2018), which is a freely available corpus of eye-tracking data with accompanying predictability norms (https://osf.io/sjefs). The corpus contains both interest area (word-based) and fixation reports for 84 participants who read 55 multiline texts (mean length= 50 words; range: 39-62 words) while their gaze positions were sampled via an SR Research EyeLink 1000+ eye-tracker sampling at 1000 Hz. Each text had 3-4 lines (mean= 3.5 lines), with a mean length of 84.2 characters (range: 5-100 characters). Lines from which readers will have made return-sweeps (i.e., non-final line) were 96.7 characters in length (range: 91 - 100 characters). Word length, Zipf frequency *(log10(frequency per billion words)* obtained from the SUBTLEX-UK corpus (van Heuven et al., 2014), and cloze predictability for the raw, unfiltered corpus are shown in Table 1, accompanied by means for filtered data. In the *Online Supplemental Materials*, we visualise the distribution of lexical predictors for intra-line and line-final words entering our analyses.

Transparency and Openness

For our eye movement corpus analysis, we report all data exclusions, all manipulations, and all measures entered into our analysis. All data and analysis code are available at https://doi.org/10.17605/OSF.IO/E4R2H. Our analyses were not pre-registered.

Data analysis

We analysed two eye movement measures for line-final words, regardless of whether readers' return-sweeps were made from these words or not: single-fixation duration (the duration of the initial first-pass fixation on a word given that it received only one first-pass fixation) and gaze duration (the sum of all first-pass fixations on a word before moving to another). Our analysis was restricted to these two measures as our primary goal was to examine how frequency and predictability influence reading times prior to the decision to shift the eyes across a line boundary and execute a return-sweep during first-pass reading. While we could have additionally analysed first-fixation durations on target words to achieve our goals, these fixations often represent a mixture of single-fixations and first of multiple fixations. Fixations that are the first of multiple fixations are often shorter in duration and land further from the optimal viewing position than their single-fixation counterparts (i.e., inverted optimal viewing position effects, see Nuthmann et al., 2007; Vitu et al., 1990, 2001, for discussions). By analysing single-fixation cases we can assess the effects of frequency and predictability in the earliest of eye movement measures while reducing effects of the IOVP. Analysing gaze durations enabled us to examine cases where readers made multiple fixations on a line-final word before a return-sweep. Analysing single-fixation duration and gaze duration also gave us parity with Parker and Slattery's (2019) investigation of frequency and predictability effects for line-initial words. For each measure, we present two sets of analyses: (1) a comparison of intra-line and line-final words; and (2) an analysis of line-final words. Analysis 1 enabled us to first replicate frequency and predictability effects for intra-line words before comparing these effects with those for line-final words, Analysis 2 enabled us to directly examine frequency and predictability effects for line-final words.

Data cleaning

Luke and Christianson (2018) prepared the dataset so that fixations shorter than 80 ms and longer than 800 ms were removed from the eye movement records. We then imposed five additional data cleaning steps: (1) we removed the first and last word in each passage (8.7% of words); (2) following previous corpus analyses (e.g., Miellet et al., 2007; Parker & Slattery, 2019; Whitford & Titone, 2014), we removed function words

(42.4% of words); (3) we removed words that were less than 4- or greater than 12-letters in length (following Parker & Slattery, 2019; 18.3% of words); (4) we removed words if they were preceded or followed by a blink (12% of words). This left us with usable data for 4,539 line-final words and 81,654 intra-line words. Of the 86,193 words, single fixation data was present for 50,336 words and gaze duration data was present for 61,673 words. We then adopted (5) Hoaglin and Iglewicz's (1987) approach to identifying and removing outliers on a participant-level basis, separately for line-final and intra-line words[5]. This procedure defined outliers as data points that were 2.2 times the difference between the first quartile (Q1) and the third quartile (Q3), above or below the Q1 and Q3 values (e.g., lower boundary = Q1 – 2.2 × (Q3-Q1); upper boundary = Q3 + 2.2 × (Q3-Q1)). For our analysis of single fixation durations, there were 47,586 observations following cleaning, indicating that the Hoaglin and Iglewicz procedure led to the removal of 5.5% of observations. For our analysis of gaze durations, there were 57,717 observations following cleaning, indicating that the Hoaglin and Iglewicz procedure led to the removal of 6.5% of observations.

Linear mixed-effects analysis

For each eye movement measure, a series of linear mixed-effects models were fitted using the *lmer()* function from the lme4 package (version 1.1.35.3; Bates et al., 2015) within R (version 4.3.3; R Development Core Team, 2020). The model comparing reading times on intra-line and line-final words adopted an identical fixed effects structure for both single fixation duration and gaze duration: *dv~ frequency × predictability × length × position + (1 | participant) + (1 | word)*, where participant and word are random factors[6]. Word length was included as a control variable within the model and allowed to interact with all other predictors. This is because word length has a strong influence on reading times (Rayner, 2009) and, as indicated in Table 1, it is correlated with other lexical predictors. Word frequency, predictability, and length were scaled and centred before analysis using the *scale()* function within R, where the mean is subtracted from each score before dividing by the standard deviation, to reduced the impact of the intercorrelated nature of the data. Position, a categorical variable coding whether the word was intra-line or line-final, was coded so that intra-line words corresponded to the intercept to which line-final words were compared (i.e., treatment coding). Given that the intra-line word represented the intercept, main effects of each lexical variable were assessed for intra-line words. Any interaction with position indicated whether the main effect of lexical variables differed for line-final words relative to intra-line words. To specifically examine lexical effects for line-final

[5]The advantage of using this method was that it enabled us to take into account the whole distribution when defining outliers instead of relying on summary statistics. Furthermore, because we identified outliers separately for each participant for both intra-line and line-final words, subtle variation in each dependent measured was not unnecessarily screened out as noise.

[6]We originally included random intercepts for item number. However, this resulted in convergence warnings for several models or the intercept captured little variance and resulted in poor model fitting.

words, we fitted an additional model to line-final reading data: *dv~ frequency* $\times$ *predictability* $\times$ *length + (1 | participant) + (1 | word)*. For all dependent variables, we applied a log-transformation to remove the rightwards skew of the distribution. Inspection of the skewness values indicated that the log-transformation reduced the skew in the data as skewness fell from 1.067 to 0.047 for single-fixation duration and from 1.544 to 0.299 for gaze duration. For all models, we report regression coefficients (b), standard errors (*SE*), and t-values.

To estimate the best fitting random structure for each model, the *buildmer()* function from the buildmer package (version 2.11; Voeten, 2021) was used. First, a maximal structure was fitted to the data before applying a backwards elimination process based on the significance of the change in log-likelihood between models. The most basic and possible model retained all fixed effects and random intercepts for participants and words.

To evaluate the evidence for the critical null effects, we supplemented our analyses with Bayes Factor analysis. Bayes Factors quantify how much evidence the data (and priors) provide in favour of two competing models and allow us to infer how much a given hypothesis is consistent with the data (for reviews see Nicenboim et al., 2023, and Wagenmakers, 2017). Bayes Factors were computed by first fitting Bayesian linear-mixed effects models to reading time data using the *brm()* function from the brms package (version 2.21.0; Bürkner, 2007). The models included the same fixed effects as the *lmer()* models. Non-informative priors *normal(0,1)* were assumed for each fixed effect. Each model used 12,000 iterations with four chains, where the first 2,000 iterations were discarded due to warm-up. Then the *hypothesis()* function was implemented to calculate the Bayes Factors (*BF10*) for each fixed effect. The *hypothesis()* function computes Bayes Factors using the Savage-Dickey density ratio method (Dickey, 1971), where Bayes Factors for individual parameters within a model are taken as the posterior density of the model parameter of interest divided by the prior density at the critical point of inference (e.g., zero if assessing whether an estimate is greater than zero). Bayes Factors greater than one indicate that evidence in favour of a given hypothesis has increased.

The combination of frequentist and Bayesian analysis enabled us to take a two-stage approach to inference. We considered results to be statistically significant where $|t| > 1.96$. If $|t| < 1.96$ and $BF10 > 1/3$, we considered there to be insufficient evidence. If $|t| < 1.96$ and $BF10 < 1/3$, we concluded that there was evidence in favour of the null hypothesis.

Following our analyses, we calculated Variance Inflation Factors (VIFs) to assess the extent to which correlations between lexical variables impacted estimates of the fixed effects reported for each model. The VIFs, which are reported in the *Online Supplemental Materials*, indicate that multicollinearity was not a concern for the models fitted to data from the Provo Corpus.

Results

Approximately 70.1% of intra-line words were fixated during first-pass reading while 68.4% of line-final words were fixated during first-pass reading. For the filtered data, return-sweeps were made from 39.0% of line-final words. The effects of frequency, predictability, and length are visualized in Figure 1 for intra-line and line-final words.

Single fixation duration

To compare single fixation durations for intra-line and line-final words, we fitted a linear mixed-effect model to 47,586 data points *(lmer(dv~ frequency × predictability × length × position + (1 + position | participant) + (1 + predictability | word)))*. As indicated in Table 2, there were significant main effects of frequency, predictability, and length, indicating that intra-line single fixation durations were shorter for high frequency, high predictability, and shorter words. The simple main effect of position significantly impacted single fixation duration, indicating that line-final words received shorter single fixation durations than intra-line words. Importantly, the interaction between frequency and predictability did not impact intra-line single fixation durations and the Bayes Factor (*BF10*= 0.002) indicated evidence in favour of the null, indicating that the frequency and predictability had an additive effect on single fixation durations for intra-line words. There were no other statistically significant interactions, for which Bayes Factors favoured the null, indicating that lexical variables did not jointly impact single fixation durations for intra-line words, nor did the effects of frequency, predictability, length, or their interactions differ between intra-line and line-final words.

We then fitted a model to single fixation durations for line-final words. The final model *(lmer(dv~ frequency × predictability × length + (1 + frequency | participant) + (1 | word)))*, fitted to 2,547 data points, indicated significant main effects of frequency and length where single fixation durations on line-final words were shorter for words of a higher frequency and shorter lengths. The effect of predictability was not significant and the Bayes Factors indicated evidence in favour of the null despite the results of our analysis comparing single fixation durations on intra-line and line-final words indicating that the predictability effect for line-final did not differ from the predictability effect for intra-line words. There was no evidence to conclude that higher-level interactions between lexical variables impacted line-final single fixation durations. Importantly, there was evidence in favour of a null interaction between frequency and predictability for single fixation durations on line-final words.

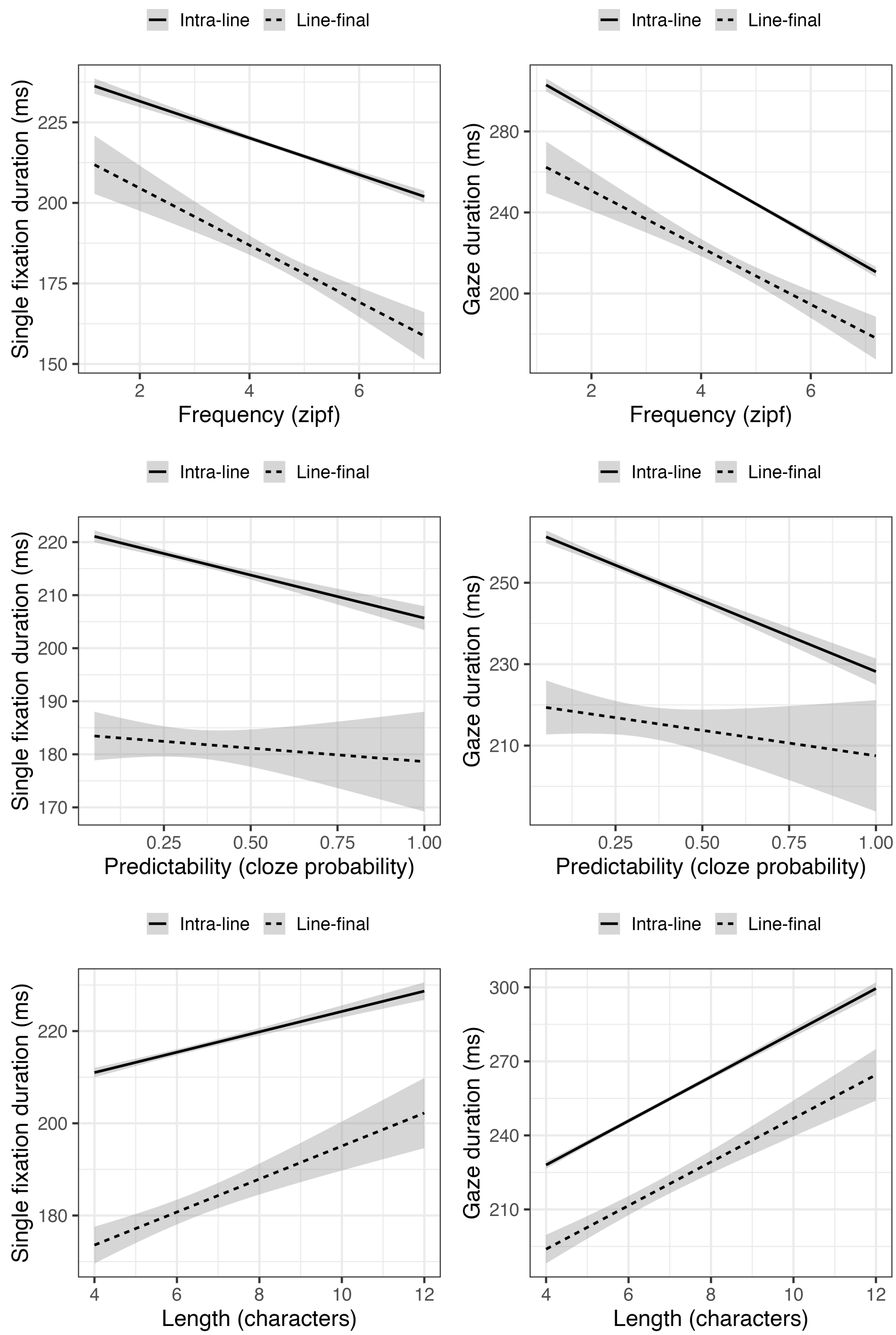

Figure 1: Plots showing the effect of frequency, predictability, and length for single fixation durations and gaze durations. Slopes for intra-line words are represented by solid black lines. Slopes for line-final words are presented by the dashed black lines. The gray bands represent the 95% confidence intervals.

Table 2: Linear mixed-effects model coefficients for the Provo Corpus.

		Single fixation duration				Gaze duration			
Model	Fixed effect	b	SE	t	BF10	b	SE	t	BF10
Comparison model	(Intercept)	2.313	0.005	425.29	-	2.357	0.007	348.41	-
	Frequency (F)	-0.009	0.002	-4.87	1.630e+04	-0.014	0.002	-7.05	9.497e+17
	Predictability (P)	-0.004	0.002	-2.59	0.047	-0.008	0.001	-6.02	5.357e+14
	Length (L)	0.006	0.002	3.43	0.719	0.021	0.002	10.58	-3.547e+15
	Position	-0.089	0.009	-9.87	1.227e+17	-0.094	0.010	-8.98	6.422e+40
	F × P	0.001	0.002	0.31	0.002	0.001	0.001	1.14	0.002
	F × L	-0.002	0.002	-1.02	0.003	-0.002	0.002	-1.27	0.004
	P × L	-0.001	0.002	-0.41	0.002	-0.002	0.001	-1.10	0.003
	F × Position	-0.011	0.006	-1.74	0.028	-0.006	0.006	-0.89	0.009
	P × Position	0.006	0.007	0.97	0.011	0.012	0.006	1.98	0.040
	L × Position	0.003	0.006	0.50	0.007	0.003	0.007	0.43	0.007
	F × P × L	0.002	0.002	1.21	0.003	0.000	0.001	0.42	0.001
	F × P × Position	-0.004	0.006	-0.63	0.008	-0.005	0.005	-1.03	0.009
	F × L × Position	-0.007	0.006	-1.12	0.012	-0.012	0.006	-2.12	0.054
	P × L × Position	-0.001	0.008	-0.14	0.008	0.002	0.007	0.28	0.008
	F × P × L × Position	0.000	0.008	0.01	0.008	0.010	0.007	1.47	0.019
Line-final words	(Intercept)	2.230	0.009	246.63	-	2.272	0.012	194.33	-
	Frequency (F)	-0.019	0.007	-2.79	0.264	-0.019	0.008	-2.35	0.112
	Predictability (P)	0.001	0.006	0.15	0.006	0.001	0.007	0.13	0.007
	Length (L)	0.017	0.006	2.57	0.153	0.032	0.008	4.21	7.350
	F × P	-0.008	0.006	-1.30	0.014	-0.008	0.007	-1.18	0.014
	F × L	-0.004	0.007	-0.62	0.009	-0.008	0.008	-0.93	0.013
	P × L	0.002	0.008	0.32	0.008	0.003	0.009	0.37	0.010
	F × P × L	0.004	0.008	0.48	0.009	0.008	0.009	0.80	0.013

Gaze duration

To compare gaze durations for intra-line and line-final words, we fitted a linear mixed-effect model to 57,717 data points *(lmer(dv~ frequency × predictability × length × position + (1 + position | participant) + (1 | word)))*. As indicated in Table 2, there were significant main effects of frequency, predictability, and length, indicating that intra-line gaze durations were shorter for high frequency, high predictability, and shorter words. The simple main effect of position significantly impacted gaze durations, indicating that line-final words received shorter gaze durations than intra-line words. Importantly, the interaction between frequency and predictability did not impact intra-line gaze durations and the Bayes Factor (*BF10*= 0.002) indicated evidence in favour of the null. The interaction between predictability and position significantly impacted gaze duration, indicating the effect of predictability differed for line-final relative to intra-line words. If reliable, this would indicate that predictability effects were negligible, if not reversed, for line-final words as indicated by the difference in the model estimates for the main effect of predictability and the position by frequency interaction being positive (indicating that as predictability increases so does gaze duration). Our analysis also revealed that the interaction between frequency and length differed for line-final relative to intra-line words. From Figure 2 of the *Online Supplemental Materials*, it appears that word length effects are stronger for low-frequency words than for high-frequency words with this difference being more pronounced for line-final relative to intra-line words. The remaining higher-level interactions were not statistically significant and had Bayes Factors that favoured the null.

We then fitted a model to gaze durations on line-final words. The final model *(lmer(dv~ frequency × predictability × length + (1 + frequency | participant) + (1 | word)))*, fitted to 3,030 data points, indicated significant main effects of frequency and length where gaze durations on line-final words were shorter for words of a higher frequency and shorter lengths. The effect of predictability was not significant and the Bayes Factors indicated evidence in favour of the null, despite the results of our analysis comparing gaze durations on intra-line and line-final words. There was no evidence to conclude that higher-level interaction between lexical variables impacted line-final gaze durations.

Discussion

Our analysis of the Provo Corpus set out to examine frequency and predictability effects for line-final words. Specifically, we fitted a series of linear mixed-effects models to two eye movement measures: Single-fixation duration and gaze duration. For each dependent variable, we started by fitting a comparative model with fixed effects for frequency, predictability, length, a categorical variable that coded whether a word was presented as intra-line or line-final, and all possible interactions between these variables. This comparative model enabled us to first examine joint effects of frequency, predictability, and length for intra-line words before comparing these lexical predictors between intra-line and line-final words. We then supplemented our comparative model with a reduced model fitted to data for line-final words. This enabled us to explicitly examine the effects of lexical predictors for line-final words.

For single-fixation duration, results from our comparative model indicate that we were able to replicate additive effects of frequency and predictability during intra-line reading while controlling for word length. The same model indicated that although single-fixation durations were shorter on line-final words, the effects of frequency, predictability, and length did not differ between intra-line and line-final words. However, our restricted model fitted to single-fixation durations for line-final words indicated effects of frequency and word length but not predictability. While the results for single-fixation durations are relatively straightforward, the results for gaze duration are a little more complex. Our comparative model fitted to gaze duration data for intra-line and line-final words indicated statistically reliable effects of frequency, predictability, and length during intra-line reading. As with single-fixation duration, our comparative model indicated that gaze durations were shorter on line-final words compared to intra-line words. However, there was evidence to suggest that frequency and predictability effects differed between intra-line and line-final words. The predictability by position interaction indicated that predictability effects were negligible, if not reversed, for line-final words and the three-way interaction between frequency, length, and position indicated that frequency effects were larger for line-final words; an interaction, which is largely driven by larger word length effects for low-frequency line-final words. Our reduced model fitted to gaze duration confirmed a lack of predictability effects for line-final words while there was clear evidence of frequency and length effects. There was no frequentist or Bayesian evidence of an interaction between any of the lexical predictors in readers' gaze durations in our reduced model fitted to gaze duration, which is surprising given the three-way interaction in our comparative analysis.

A consistent finding across both eye movement measures is that reading times were shorter on line-final words relative to intra-line words. This is consistent with an empirical body of work showing that reading times for line-final words are typically shorter than those for intra-line words (Tiffin-Richards & Schroeder, 2018). This is a similar observation to shorter line-final fixations relative to intra-line fixations (Abrams & Zuber, 1972; Adedeji et al., 2021; Parker, Nikolova et al., 2019; Parker, Slattery, et al., 2019; Rayner, 1977).

After statistically controlling for word length, we found clear evidence of frequency effects for both single-fixation durations and gaze durations in our reduced analysis of line-final words. Our comparative analysis of single-fixation duration indicated that frequency effects did not differ between intra-line and line-final words. However, our comparative analysis of gaze duration indicated that frequency effects for long words may have been more pronounced for line-final words relative to intra-line words. Regardless, the emergence of frequency effects for line-final words is problematic for accounts which posit that shorter line-final fixations are the result of reduced, or even an absence of, lexical processing. Instead, the evidence suggests that these fixations are under lexical control.

Regarding predictability effects, the evidence from the Provo Corpus was mixed. For single fixation duration, our comparative analysis indicated that the effects of predictability did not differ between intra-line and line-final words. Yet our reduced model fitted to single-fixation duration yielded a null result. This pattern of results is highly similar to that reported in Parker and Slattery's (2019) analysis of the Provo Corpus. Parker and Slattery focused on the processing of line-initial words and conducted analyses comparing both intra-line and line-initial words. As with the current study, predictability effects were observed for intra-line words and the interaction between predictability and position was null, indicating that predictability effects did not differ significantly between intra-line and line-initial words. Yet predictability effects were absent when analyzing reading times on line-initial words. Parker and Slattery argued that an absence of predictability effects for intra-line words could have resulted from a restricted range of cloze values entering the analysis. For the current analysis, the range of cloze values in Table 2 is highly similar for the analysed intra-line and line-final words and Figure 1 of the *Online Supplemental Materials* shows a highly similar distribution of cloze probabilities, which opposes such a possibility. For gaze durations, the effect of predictability differed between intra-line and line-final words and our reduced analysis confirmed that predictability effects were absent for line-final words. This perhaps more convincingly illustrates that predictability effects differ for line-final words relative to intra-line words than did the interaction in the comparative model. However, it is important to note that these interactions differed across eye movement measures, which makes the pattern of results difficult to interpret as both single-fixation duration and gaze duration index the early stages of lexical processing. von der Marlsburg and Angele (2017) made the case that at least two dependent measures showing consistent results should be considered as evidence for an effect. Based on this criteria then it makes interpreting whether predictability effects differ between intra-line and line-final words difficult and suggests that there is ambiguous evidence in the current study. These potentially spurious results that are inconsistent across eye movement measures could be explained by intercorrelated variables entering the analysis. Indeed, interactions between lexical predictors have been reported in corpus studies (e.g., between frequency and length; Kliegl et al., 2006) that are absent under experimental conditions. Inspection of the VIFs in the *Online Supplemental Materials*, however, would suggest that intercorrelations in these analyses may not have been as problematic as one would think given the correlations in Table 1.

Regarding the statistically significant interactions between lexical predictors and the categorical fixed effect coding for word position in gaze duration, it is also important to note that the Bayes Factors indicated evidence for the null. Our pre-registered inference criteria were to only use Bayes Factors to supplement our null frequentist results. That said, it is important to highlight that Bayes Factors favoured the null for the interactions between lexical predictors and the categorical variable coding for position in gaze duration. Currently, there is (to our knowledge) no fast or hard rule for integrating the two forms of inference but, taken together, they may suggest that these interactive effects are small with plausible values being centred very close to zero and carry little practical significance. Although this does not completely reconcile our findings, it does suggest that a tightly controlled experimental study of frequency and predictability for line-final words is necessary before strong claims can be made.

A final point of discussion regarding our corpus analysis is that across two eye movement studies (the current work and Parker & Slattery, 2019), we have found mixed results when examining predictability effects for words at the location of line boundaries. This suggests that the corpus may not be appropriate for examining the influence of lexical variables in these spatial locations when focusing solely on line-initial and line-final words. There are a number of possibilities for why this might be. The first is that there are relatively few high-predictability words in the Provo Corpus and this makes it difficult to detect predictability effects when analyses are restricted to the lower end of the cloze scale without a sufficiently high number of observations. This would explain why we are able to detect predictability effects in our comparative analyses with approximately 20 times more observations. A second possible explanation for the absence of predictability effects in these restrcited analyses may be that there was poor calibration in these locations and, as a result, fixation locations are mislocated. Carr et al. (2022) illustrated that during paragraph reading there is often noise that occurs during data acquisition resulting in fixations being inappropriately assigned to the wrong line. Luke and Christianson (2018) do not report whether the eye movement records in the Provo Corpus were adjusted for noise or drift that occurred during data acquisition. A lack of adjustment may explain null effects in these locations given that there is often a downward slope during recording that results in fixations further to the right being assigned to a line below where the reader was looking. To be clear, we do not go as far as to say that the corpus is inappropriate for examining predictability effects given that we have found effects of cloze probability during intra-line reading as have other authors (e.g., Luke & Christanson, 2016), but the apparent lack of effects during the analysis of words at the start and end of the line may suggest that these words might not be suitable for these very specific analyses.

As an interim summary, there was consistent evidence that reading times were shorter on line-final words relative to intra-line words. While the precise explanation for this pattern of results remains unclear, it is evident that frequency effects do emerge for line-final words. The presence of frequency effects indicates that fixations on line-final words are driven by lexical processing and the reduction in fixation durations cannot be attributed to a complete lack of lexical processing in these locations in preparation for a return-sweep. Predictability effects are a little less clear. There may be several possible reasons for the absence of an effect: a lack of control over lexical properties of words entering analysis (e.g., word length) resulting in spurious effects between eye movement measures, or misestimation of true effects. Without a further eye movement study to address the proposed limitations of the Provo Corpus, it is difficult to draw firm conclusions about the effects of word frequency and predictability for line-final words.

Eye Movement Experiment

Pre-registered predictions

Our analysis of the Provo Corpus provided evidence that reading times were shorter on line-final words and that frequency reliably influenced reading times on line-final words. The effects of predictability were a little less clear with the results being mixed between eye movement measures. What makes these predictability-related effects difficult to interpret are the speculated shortcomings of our corpus analysis. Therefore, for our eye movement experiment, we derived predictions based on our extended E-Z Reader framework that was outlined in the *Introduction* and, as such, we predicted additive effects of frequency and predictability that did not vary as a function of position. These predictions are plausible under the paraofveal processing and integration accounts of shorter line-final fixations. Below we specify predictions for (1) intra-line target words, (2) the comparison of intra-line and line-final words, and (3) line-final words.

(1) Intra-line target words

(a) There will be a main effect of frequency on reading time measures, where reading times are shorter for high-frequency words.
(b) There will be a main effect of predictability on reading time measures, where reading times are shorter for highly predictable words.
(c) There will be no evidence of an interaction between frequency and predictability, i.e. an additive effect of frequency and predictability.

(2) Comparison between intra-line and line-final words.

(a) Reading times on line-final words will be shorter than on intra-line words.
(b) Frequency effects will not differ in magnitude for intra-line and line-final words.
(c) Predictability effects will not differ in magnitude for intra-line and line-final words. However, if the predictability by position interaction within the Provo Corpus was indeed reliable, we might expect a significant interaction here, where predictability effects were smaller for line-final words.
(d) As with intra-line reading, there will be additive effects of frequency and predictability for line-final words and, as such, the three-way interaction between frequency, predictability, and position will not reliably influence reading times.

(3) Line-final target words.

(a) There will be a main effect of frequency on reading time measures, where reading times are shorter for high-frequency words.
(b) There will be a main effect of predictability on reading time measures, where reading times are shorter for highly predictable words. Note that if the lack of predictability effects for line-final words within the Provo Corpus was reliable, then we might expect an absence of predictability effects for line-final words.

(c) There will be no evidence of an interaction between predictability and frequency, i.e. an additive effect of predictability and frequency.

Method

Transparency and Openness

To address the limitations of our analysis of the Provo Corpus, we conducted a controlled eye movement experiment between October 2021 and September 2022. The experiment was pre-registered on the Open Science Framework prior to the commencement of data collection. The registration form can be found at https://doi.org/10.17605/OSF.IO/6B8HM and the materials, data, and R scripts can be found at https://doi.org/10.17605/OSF.IO/E4R2H. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Participants

A priori power analyses were conducted for all fixed effects of interest within a frequentist linear-mixed modelling framework (i.e., main effects of predictability and frequency, and a simple main effect of position). To begin, we simulated a data set with known properties for gaze duration; that is a 15 ms effect of frequency, a 15 ms effect of predictability, and a 25 ms effect of position. These estimates were determined to be our minimal effect sizes of interest and are substantially smaller than previously reported effect sizes (see Staub, 2015, Table 2) meaning that our required sample size would be somewhat conservative. We then set all estimates for interactions to zero. The data set contained 104 observations per participant (13 per experimental condition). This number of observations took the 128 experimental items and removed three per experimental condition to build in an arbitrary skipping rate of ~19% across each condition (similar to skipping rates reported by Rayner et al., 2004). For further details see: https://osf.io/8a543/. One thousand simulations were run for 1 to 10 statistical subjects per counterbalance list. We then fitted linear mixed-effects models to examine our simulated data. Within this framework, each hypothesis is mapped directly onto a fixed effect of interest. As shown in Figure 2, approximately 32 participants in total would provide 80% at an alpha level of $|t| > 1.96$ to detect the main effects of frequency and predictability and a simple main effect of position. Approximately 32 participants would provide a scenario where the 95% confidence intervals were above 0.80 (i.e., 80%) power[7].

[7] Note that due to poor visualization which treated statistical subjects as a continuous variable rather than a discrete variable (https://osf.io/zt6we; version 1), we pre-registered that 36 participants would be required to achieve 95% that did not overlap

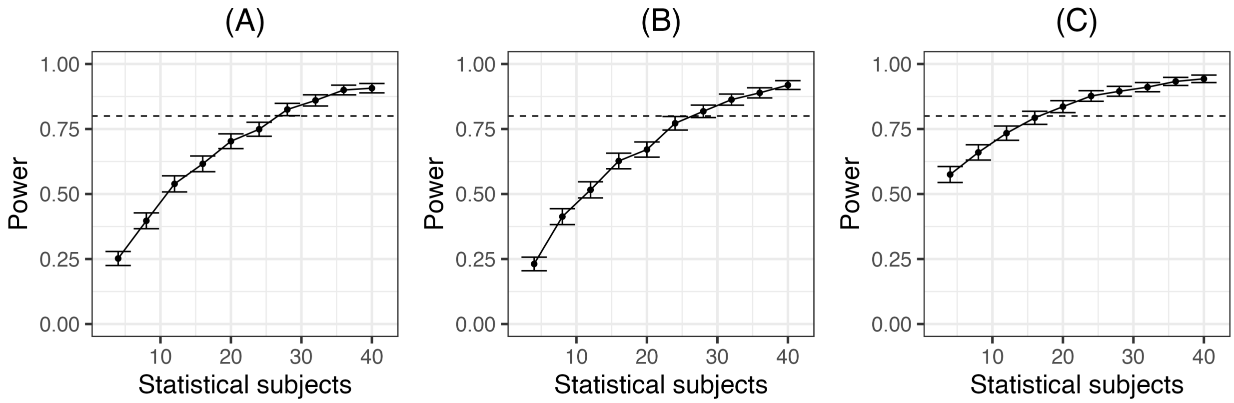

Figure 2: Power curves for effects where we predicted a significant difference: (A) frequency, (B) predictability, and (C) position. The error bars represent the 95% confidence intervals around the mean at each point.

However, as we also set out to assess evidence for a series of null interactions, conducting and powering our study within a null hypothesis testing framework seemed sub-optimal. Thus, we used Bayes Factors to make inferences about critical interactions (e.g., the frequency by predictability interaction) and implemented an *open-ended sequential Bayes Factor design* (Schönbrodt & Wagenmakers, 2018). Under this approach, we specified that we would first collect data from 32 participants. At this point, we fitted Bayesian linear mixed-effects models to the data and derived Bayes Factors. If the Bayes Factors were decisive for all fixed effects, we would stop recruitment. If Bayes Factors were ambiguous (i.e., $1/3 < BF10 < 3$), then we would continue recruiting participants in runs of four (one per counterbalance list) until the Bayes Factors were decisive. The advantage of using the Bayesian stopping rule is that we would not have to adjust significance thresholds for sequential testing. Following our open-ended sequential Bayes Factor design, we stopped data collection when we had usable data for 32 participants as all Bayes Factors for our pre-registered analyses were decisive. The final sample is described below.

Native English speakers were recruited via the UCL PALS SONA Participant Pool. Participants were aged between 18-45 years, had no language, hearing, or visual impairments, and had no history of neurological illness. Participants were reimbursed at a rate of £8.00/hour or received course credit for their participation. The experimental procedure was granted ethical approval by the UCL Department of Experimental Psychology's Ethics Chair, ethics application number: EP_2021_015. Of the 36 readers initially recruited, data was removed from three readers due to poor calibration and low data quality and one further participant's data was removed due to excessive blinks. The final sample of 32 readers (22 female, 10 male) had a mean

with zero. However, as can be seen in Figure 2 (and verified by running the power analysis code), it is indeed confirmed that 32 participants are sufficient to achieve adequate power for the main effects of interest. Furthermore, as Bayes Factors were all decisive at this point, further recruitment seemed uneconomical.

Table 3: Cloze probabilities per experimental condition

Experimental condition			
Frequency	Predictability	Position	Cloze probability
HF	HP	Line-final	0.71 (0.17)
LF	HP	Line-final	0.71 (0.15)
HF	LP	Line-final	0.01 (0.03)
LF	LP	Line-final	0.01 (0.03)
HF	HP	Intra-line	0.70 (0.15)
LF	HP	Intra-line	0.71 (0.16)
HF	LP	Intra-line	0.01 (0.04)
LF	LP	Intra-line	0.01 (0.02)

age of 22.3 years (SD= 5.53 years; range: 18 - 40 years).

Materials

Sixty-four high- and low-frequency target words were selected for the experiment. High-frequency words had a mean Zipf frequency of 4.8 (SD= 0.36) and low-frequency words had a mean Zipf frequency of 3.6 (SD= 0.58). For the experiment, a high-frequency word was paired with a low-frequency word matched on length. The mean length across all words was 6.0 characters (SD= 1.30).

For each word pairing, four passages of text were created (each with two lines). The context was varied so that two passages would highly constrain the high-frequency target word and the target would appear either intra-line or line-final. Low-frequency words were also embedded in these passages so that they were low-predictability. Two passages would highly constrain the low-frequency word and the target would be intra-line or line-final. High-frequency words were also embedded in these passages so that they were low-predictability (see Figure 3). This led to a 2 (frequency: high vs low) $\times$ 2 (predictability: high vs low) $\times$ 2 (position: intra-line or line-final) design. Participants viewed each passage for the 64-word pairing (128 stimuli in total). That is, 16 items per experimental condition with items being divided into four sets and counterbalanced over participants. On average, items in the line-final condition had a mean line length of 81.6 characters (SD= 4.92 characters) and items in the intra-line condition had a mean line length of 80.7 characters (SD= 5.00 characters).

A cloze norming study (n= 48) confirmed the appropriateness of our stimuli for the current experiment. Cloze probabilities are shown in Table 3. A repeated-measures ANOVA, with frequency, predictability, and position as factors, revealed that cloze accuracies were higher in the predictable condition, $F(1, 504)$= 4904.05, $p<$ .001. All other main effects and interactions were non-significant (Fs $<$ 1).

Jamming all my laundry into the washer, I ignored the fact that it could ***break***
because I had overloaded its capacity. *(HP, HF, Line-final)*

Jamming all my laundry into the washer, I ignored the fact that it could ***erupt***
because I had overloaded its capacity. (LP, LF, Line-final)

Jamming all my laundry into the washing machine, I had simply ignored the fact
that it could ***break*** because I had overloaded its capacity. (HP, HF, Intra-line)

Jamming all my laundry into the washing machine, I had simply ignored the fact
that it could ***erupt*** because I had overloaded its capacity. (LP, LF, Intra-line)

The geologists hurried from the volcano. The measurements suggested that it could ***erupt***
at any moment. (HP, LF, Line-final)

The geologists hurried from the volcano. The measurements suggested that it could ***break***
at any moment. (LP, HF, Line-final)

The skilled geologists hurried to get away from the volcano. Their measurements
suggested that it could ***erupt*** at any moment. (HP, LF, Intra-line)

The skilled geologists hurried to get away from the volcano. Their measurements
suggested that it could ***break*** at any moment. (LP, HF, Intra-line)

Figure 3: Example stimuli with the target words *break* and *erupt* shown in bold. Text in the experiment was 2.5 spaced across lines.

Apparatus

An SR Research EyeLink 1000+ desktop-mounted system with a sampling rate of 1000 Hz was used to track monocular eye movements. Stimuli were presented on a Dell UltraSharp U2414H 23.8-inch monitor with 1920 × 1080 resolution at a viewing distance of 74 cm. Each character was presented in black 18-point Courier New font and 2.5 line spacing was used[8]. Responses to comprehension questions were recorded via a button press on the keyboard.

Procedure

Participants were tested in a laboratory room at University College London. Participants were first asked to read an information sheet before providing written informed consent. Participants were informed that they would be reading short passages of text for comprehension and answering occasional *TRUE/FALSE* comprehension questions (appearing after 25% of trials). Participants were instructed to press *SPACE* when they had finished reading a passage. When answering comprehension questions, participants were instructed to press the *S* key for *TRUE* and the *K* key for *FALSE*. Before completing the reading experiment participants completed a 9-point calibration and validation procedure. The average error of the calibration and validation procedure had to be below 0.40 or the procedure was repeated. For the passages to appear on the screen participants first had to first fixate a point that was positioned slightly left of the first word in the passage. Participants were presented with four practice trials before the experimental items. Items were presented in random order. The entire experiment lasted approximately 45 minutes. Participants were debriefed at the end of the experiment.

Data analysis

Our analyses of the experimental data mirrored our analysis of the Provo Corpus. Again, we analysed single fixation durations and gaze durations on line-final words regardless of whether readers' return-sweeps were made from these words or not. Predictions for (1) intra-line reading and (2) a comparison of intra-line and line-final words were assessed via models fitted to both intra-line and line-final reading data. Predictions for (3) line-final words were examined via models fitted to reading times for line-final words.

[8]This line spacing is larger than readers are typically exposed to when reading natural texts, where single line spacing is used. A recent study conducted by Christofalos et al. (2023) empirically examined the effect of line spacing on return-sweep behaviour. While they reported that return-sweeps were launched from closer to the end of the line with large spacing (i.e., double- and triple-spaced) and that fications were longer overall, these manipulations did not influence the durations of return-sweep fixations. For a comprehensive discussion of these results, interested readers should see Christofalos et al.

Data cleaning

We pre-registered that all participants scoring less than 70% correct on the comprehension questions would be removed from the analysis; however, no participants were excluded for this reason. Fixations shorter than 80 ms and longer than 1200 ms were removed prior to analysis[9]. Of the 4096 experimental trials, 12 were removed for excessive track loss. We pre-registered that we would remove trials that contained a blink on or adjacent to a target word leading to the removal of 6.0% of trials. The resulting data set of 3,850 observations had 1,881 target words with single fixation durations and 2,687 target words with gaze durations. We then applied a Hoaglin and Iglewicz (1987) outlier removal procedure to reading time data to identify outliers individually for each participant across each experimental condition. For our analysis of single fixation duration, there were 1,815 observations following cleaning, indicating that the Hoaglin and Iglewicz procedure led to the removal of 3.6% of observations. For our analysis of gaze durations, there were 2,585 observations following cleaning, indicating that the Hoaglin and Iglewicz procedure led to the removal of 3.8% of observations.

Linear mixed-effects analysis To address our experimental predictions, we again analysed *single fixation durations* and *gaze durations* on target words. As in our analysis of the Provo Corpus, data were analysed by fitting LMMs to the data with the *lmer()* function from the lme4 package and Bayes Factors were calculated using the *hypothesis()* function from the brms package[10]. To assess our first two hypotheses, we compared reading times on intra-line and line-final words within a single model with an identical fixed effects structure: *dv~ frequency × predictability × position + (1 | participant) + (1 | item)*, where participant and item are random factors. Word frequency and predictability were both deviation coded as -.5 and .5 within each model. Position was coded so that intra-line words corresponded to the intercept to which line-final words were compared (i.e., treatment coding). As with our analysis of the Provo Corpus, our coding scheme meant that main effects of each categorical predictor were first assessed for the intercept (i.e., intra-line words). Any interaction with position indicated whether the main effect of lexical variables differed for line-final words relative to intra-line words. A model fitted to reading times on line-final words was then used to assess hypothesis 3: *dv~ frequency × predictability + (1 | participant) + (1 | item)*. For all dependent variables, we applied a log transformation to remove the rightwards skew of the distribution. Inspection of the skewness values indicated that the log-transformation reduced the skew in the data as skewness fell from 0.947 to

[9]Note that our Corpus analysis and Experiment used different fixation duration cutoffs. We additionally conducted an analysis removing fixations shorter than 80 ms and longer than 800 ms for our eye movement experiment. With these cleaning procedures, the overall pattern of results and the conclusions we draw remain unchanged. In the article we report data analysis following our pre-registered cutoffs of 80 ms to 1200 ms for our eye movement Experiment.

[10]We again used *buildmer()* to optimise the fitting of linear mixed-effects models and noninformative priors for the calculation of Bayes Factors.

-0.072 for single-fixation duration and from 1.526 to 0.247 for gaze duration.

Results

The mean accuracy on comprehension questions was 85.6% (*SD*= 35.16%; range: 70.6 - 97.1). Below we report our pre-registered analysis of reading times on target words, followed by an exploratory analysis of fixation and refixation likelihood. Approximately 74.0% of intra-line words were fixated during first-pass reading while 65.7% of line-final words were fixated during first-pass reading. Return-sweeps were made from line-final words at a rate of 55.5% during first-pass reading (52.4% when the target words was intra-line and 58.6% when the target word was line-final). Reading times on target words are visualised in Figure 4.

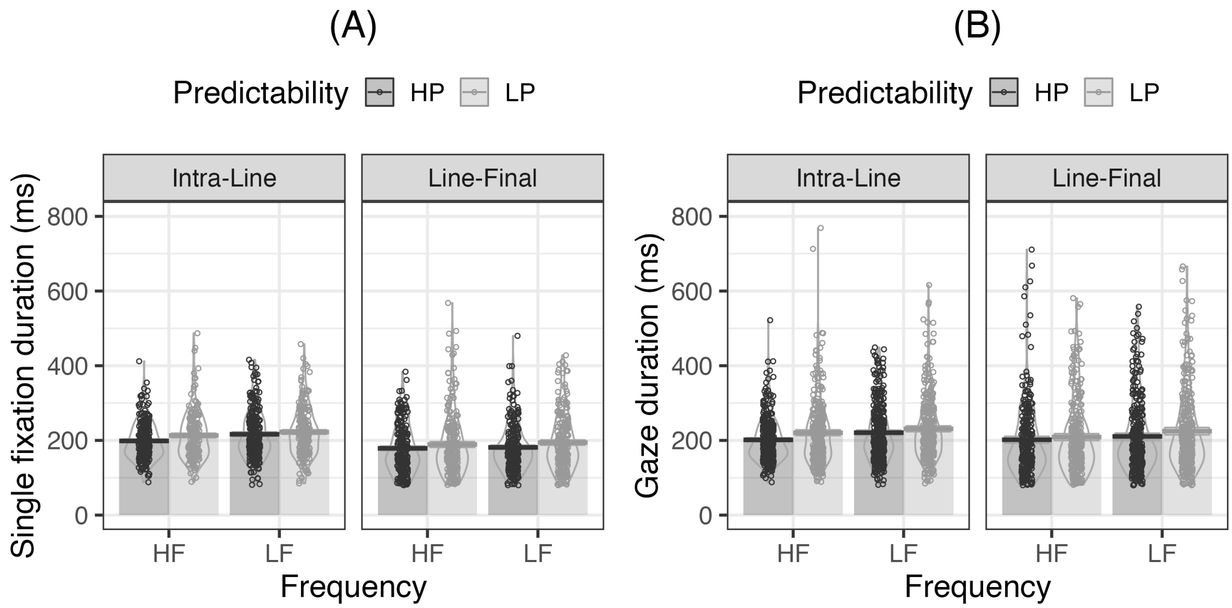

Figure 4: (A) Single fixation durations and (B) gaze durations per experimental condition. Reading times are shown in black for high-predictability targets and in grey for low-predictability targets. Horizontal bars present the mean, and error bars represent the 95% confidence intervals. Individual points present individual data points. HP: high-predictability; LP: low-predictability; HF: high-frequency; LF: low-frequency.

Single fixation duration To compare single fixation durations for intra-line and line-final words, we fitted a linear mixed-effect model to 1,815 data points *(lmer(dv~ frequency × predictability × position + (1 + position | participant) + (1 | item)))*. As indicated in Table 4, there were significant main effects of frequency and predictability at the reference level (intra-line words) indicating that intra-line single fixation durations were shorter for high-frequency and high-predictability words. The simple main effect of position significantly impacted single fixation duration, indicating that line-final words received shorter single fixa-

Table 4: Linear mixed-effects model coefficients for the eye movement experiment.

		Single fixation duration				Gaze duration			
Model	Fixed effect	b	SE	t	BF10	b	SE	t	BF10
Comparison model	(Intercept)	5.316	0.020	271.93	-	5.328	0.020	271.98	-
	Frequency (F)	-0.053	0.020	-2.58	7.008e+02	-0.061	0.019	-3.20	2.221e+03
	Predictability (P)	-0.059	0.020	-2.90	2.152e+02	-0.061	0.019	-3.25	1.666e+03
	Position	-0.152	0.031	-4.88	4.000e+04	-0.084	0.032	-2.59	1.139e+02
	F × P	-0.033	0.042	-0.78	0.056	-0.021	0.042	-0.51	0.048
	F × Position	0.006	0.029	0.21	0.029	0.011	0.028	0.38	0.030
	P × Position	0.013	0.029	0.43	0.031	-0.021	0.028	-0.75	0.036
	F × P × Position	0.028	0.058	0.49	0.063	0.020	0.055	0.37	0.060
Line-final words	(Intercept)	5.163	0.030	171.54	-	5.245	0.034	155.79	-
	Frequency (F)	-0.046	0.023	-2.05	4.617e+01	-0.050	0.022	-2.24	4.443e+03
	Predictability (P)	-0.046	0.023	-2.03	4.802e+01	-0.082	0.022	-3.66	7.774e+01
	F × P	-0.004	0.046	-0.09	0.046	-0.001	0.048	-0.02	0.047

tion durations than intra-line words. Higher-level interactions did not significantly impact single fixation durations, indicating a null interaction for frequency and predictability for intra-line words and that effects of frequency and predictability did not differ between intra-line and line-final words.

We then fitted a model to single fixation durations for line-final words. The final model *(lmer(dv~ frequency × predictability + (1 | participant) + (1 | item)))*, fitted to 905 data points, indicated significant main effects of frequency and predictability where single fixation durations on line-final words were shorter for words of a higher frequency and those that were highly predictable. There was no significant interaction between frequency and predictability.

Gaze duration To compare gaze durations for intra-line and line-final words, we fitted a linear mixed-effect model to 2,585 data points *(lmer(dv~ frequency × predictability × position + (1 + position | participant) + (1 | item)))*. There were significant main effects of frequency and predictability at the reference level (intra-line words) indicating that intra-line gaze durations were shorter for high-frequency and high-predictability words. The simple main effect of position significantly impacted gaze durations, indicating that line-final words received shorter gaze durations than intra-line words. No higher-level interactions were observed, indicating a null interaction for frequency and predictability for intra-line words and the effects of frequency and predictability did not differ between intra-line and line-final words.

We then fitted a model to gaze durations for line-final words. The final model *(lmer(dv~ frequency × predictability + (1 | participant) + (1 | item)))*, fitted to 1,226 data points, indicated significant main effects of frequency and predictability where gaze durations on line-final words were shorter for words of a higher frequency and those that were highly predictable. There was no significant interaction between frequency and predictability.

Exploratory analyses

Our pre-registered analyses focused exclusively on first-pass reading times on target words. However, E-Z Reader additionally specified how frequency and predictability influence fixation, refixation, and regression out likelihood. Therefore, we conducted formal analyses of these measures for target words. We chose to only explore these measures in our experimental work as fixation likelihoods are heavily influenced by word length (Rayner, 1998, 2009)– a variable that was not controlled in the Provo Corpus. Fixation, refixation, and regression probabilities are visualised in Figure 5.

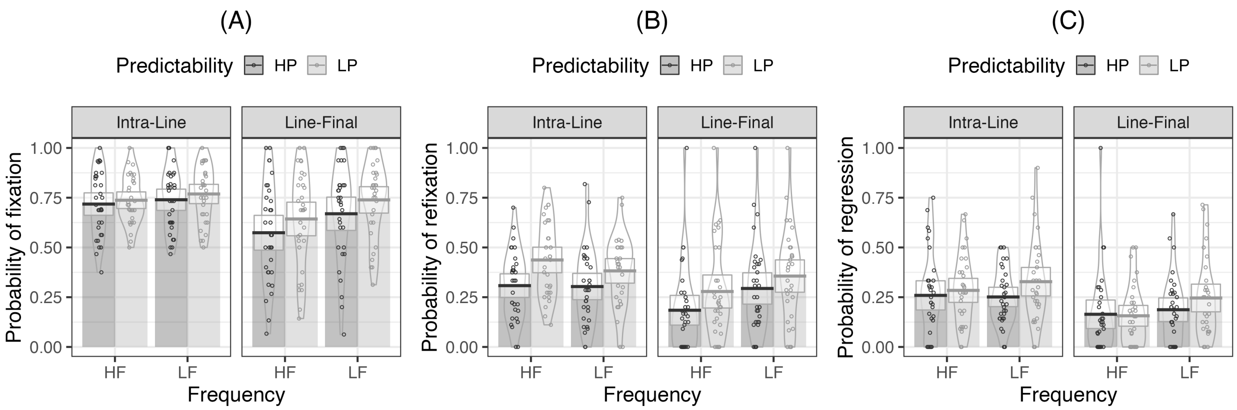

Figure 5: Aggregated probability of (A) fixation, (B) refixation, and (C) regression per experimental condition. Probabilities are shown in black fo high-predictability targets and in grey for low-predictability targets. Horizontal bars present the mean, and error bars represent the 95% confidence intervals. Individual points present participant means per condition. HP: high-predictability; LP: low-predictability; HF: high-frequency; LF: low-frequency.

To statistically assess the effect of our experimental manipulations on fixation, refixation, and regression likelihood, we first fitted generalised linear mixed-effects models, using a binomial function, to the dependent variables for intra-line and line-final target words: *glmer(dv~ frequency × predictability × position + (1 | participant) + (1 | item), family=binomial)*. We then fitted a model to line-final target word data *(glmer(dv~ frequency × predictability + (1 | participant) + (1 | item), family=binomial))* following our pre-registered analyses. All models reported included only random intercepts for participants and items and no random slopes due to a lack of convergence.

Coefficients for our exploratory analyses are included in Table 5. First, for fixation likelihood, our comparison model (3,850 data points) indicated that participants were significantly more likely to fixate intra-line words. Bayes factors indicated that there was insufficient evidence to conclude a null effect of frequency on fixation likelihood, while there was moderate evidence for a null effect of predictability on fixation likelihood. While frequentist results indicated a lack of evidence for higher-level interactions influencing fixation likelihood,

Table 5: Generalised linear mixed-effects model coefficients for our exploratory analyses of the experimental data set.

		Fixation likelihood				Refixation likelihood				Regression likelihood			
Model	Fixed effect	b	SE	z	BF10	b	SE	z	BF10	b	SE	z	BF10
Comparison model	(Intercept)	1.199	0.145	8.27	-	-0.653	0.113	-5.76	-	-1.130	0.152	-7.45	-
	Frequency (F)	-0.163	0.110	-1.47	0.341	0.122	0.117	1.04	0.186	-0.055	0.128	-0.43	0.147
	Predictability (P)	-0.127	0.110	-1.15	0.221	-0.490	0.117	-4.17	1.042e+03	-0.239	0.128	-1.87	0.755
	Position	-0.428	0.076	-5.65	1.332e+15	-0.383	0.090	-4.27	3.016e+02	-0.610	0.103	-5.91	-7.389e+82
	F × P	0.108	0.277	0.39	0.285	-0.161	0.271	-0.59	0.315	0.354	0.327	1.08	0.509
	F × Position	-0.290	0.151	-1.92	0.959	-0.642	0.178	-3.62	9.033e+01	-0.593	0.205	-2.89	1.145e+01
	P × Position	-0.214	0.151	-1.42	0.404	0.052	0.177	0.29	0.176	-0.015	0.205	-0.07	0.206
	F × P × Position	-0.099	0.302	-0.33	0.304	0.079	0.355	0.22	0.348	-0.300	0.410	-0.73	0.470
Line-final words	(Intercept)	0.883	0.222	3.98	-	-1.142	0.148	-7.74	-	-1.771	0.168	-10.56	-
	Frequency (F)	-0.536	0.111	-4.83	4.988e+28	-0.557	0.136	-4.09	2.217e+02	-0.649	0.160	-4.05	3.636e+02
	Predictability (P)	-0.400	0.111	-3.61	9.422e+01	-0.475	0.136	-3.50	4.808e+01	-0.256	0.160	-1.60	0.553
	F × P	0.024	0.304	0.08	0.301	-0.083	0.303	-0.27	0.307	0.008	0.363	0.02	0.356

Bayes factors typically clustered around 1/3 indicating that any evidence for the null was weak. By contrast, our model fitted to fixation likelihood for line-final words (1937 data points) indicated that both frequency and predictability significantly impacted fixation likelihood. There was a null effect of the frequency by predictability interaction and there was Bayesian evidence to suggest that the interaction between frequency and predictability did not influence fixation likelihoods for line-final words.

For refixation likelihood, our comparison model (2,687 data points) indicated participants were significantly less likely to refixate line-final words. Participants were also less likely to refixate high-predictability intra-line words. However, there was a null effect of frequency on intra-line words. The only significant interaction to impact refixation probability was the frequency by position interaction, indicating that the effect of frequency on refixation likelihood differed between intra-line and line-final words. Inspection of the model fitted to refixation likelihood for line-final words (1,272 data points) indicated that there was a clear effect of frequency where readers were less likely to refixate high-frequency line-final words. Similarly, participants were less likely to refixate high-predictability line-final words. There were no reliable effects of higher-order interactions on refixation likelihood, as confirmed by Bayes Factor analysis.

For regression likelihood, our comparison model (2,687 data points) indicated participants were significantly less likely to make regressions out of line-final words. The effect of frequency was null, as indicated by Bayes factors, for regression likelihood while there was insufficient evidence to make conclusions about the effect of predictability on regression likelihood. There was a significant interaction between frequency and position, indicating that the effect of frequency on regression likelihood differed between intra-line and line-final words. The remaining higher-level interactions were non-significant with Bayes factors indicating either insufficient evidence to warrant conclusions or evidence for the null. Inspection of the model fitted to regression likelihood for line-final words (12,72 data points) indicated that there was a clear effect of frequency where readers were less likely to make regressions out of high-frequency line-final words. The effects of predictability and the frequency by predictability interactions were non-significant and Bayes factor analysis indicated insufficient evidence to warrant strong conclusions on the effect of these variables on regression likelihood.

Discussion

Following our analysis of the Provo Corpus, we conducted a pre-registered eye movement study to examine the effects of frequency and predictability for line-final words. Crucially, our experiment allowed us to examine frequency and predictability under conditions where word length was controlled with sufficient statistical power. We pre-registered three sets of predictions. These are related to (1) reading times on intra-line words, (2) differences in reading times between intra-line and line-final words, and (3) reading times on line-final words. We consider each set of predictions below.

For (1) intra-line reading, the pattern of results was consistent across eye movement measures and confirmed our predictions. Reading times were shorter on high-frequency and high-predictability words and the interaction between frequency and predictability had no reliable impact on reading times. The outcome of these predictions falls in line with the published literature indicating that the effects of frequency and predictability are additive during intra-line sentence reading (Altarriba et al., 1996; Ashby et al., 2005; Kennedy et al., 2013; Miellet et al., 2007; Rayner et al., 2004; Rayner et al., 2001; Slattery et al., 2012). This replication element adds strength to the novel contributions of our work.

For (2) reading time differences between intra-line and line-final words, the findings were again consistent across eye movement measures. It was clear the reading times on line-final words were shorter than reading times on intra-line words. This finding is of course not novel and had not only been found in our analysis of the Provo Corpus but also in previous studies (e.g., Tiffin-Richards & Schroeder, 2018). In our analysis of the Provo Corpus, we observed an interaction between predictability and position in gaze duration. However, this did not extend to our experiment. In fact, our analysis of (3) reading times on line-final words confirmed that reading times were shorter for high-frequency and high-predictability line-final words. Furthermore, there was no evidence of an interaction between frequency and predictability for line-final words, confirming an additive effect as there is for intra-line reading. Together, these novel findings have important implications for both our theoretical understanding of how line-final words are processed and how computational models could be extended to the reading of line-final words. We defer a discussion of these implications for the *General Discussion.*

In addition to our pre-registered analyses, we explored fixation, refixation, and regression out likelihoods in a formal exploratory analysis. The most striking observation from these analyses is that line-final words are less likely to be fixated, less likely to be refixated, and less likely to be followed by a regression out to earlier words on the line. It has previously been argued that readers tend to avoid fixating extreme locations on a line to minimise the distance traversed by a return-sweep (Parker, Slattery, et al., 2019) and that skilled readers may be able to use parafoveal vision to encode line-final words and avoid fixating them under certain circumstances (Parker & Slattery, 2021). This use of parafoveal vision at line extremes may be able to explain the reduced fixation probability for line-final words reported here. However, the reduction in refixation probability and regressions out of line-final words, compared to intra-line words, may be more parsimoniously explained by the existing assumptions for integration failure within E-Z Reader 10. That is, if word *n* is a line-final word, then one source of integration failure, the identification of word *n+1* before the integration of word *n*, will be all but eliminated. This single mechanism within E-Z Reader 10 predicts reduced rates of refixation and regressions out for line-final words and can also explain the reduction of line-final fixation durations.

Our exploratory analyses also indicated that both frequency and predictability influenced fixation and refixation likelihoods for line-final words and frequency influenced regressions out of line-final words. However, the effects of frequency and predictability were largely equivocal for intra-line words. Given that the effects of lexical variables have been reported to influence fixation, refixation, and regression out likelihoods during intra-line reading (Rayner, 2009), it becomes difficult to interpret how frequency and predictability influence these eye movement measures. It will, therefore, be important to conduct well-powered work to verify how these variables impact fixation, refixation, and regression out likelihoods and determine whether processing difficulty plays a larger role in determining fixation likelihoods and regressions out for line-final words.

General Discussion

For computational models of eye movement control during reading to be able to simulate eye movements across multiline texts, it is essential to first understand how placeholders for lexical processing within the models (i.e., frequency and predictability) influence the processing of line-initial and line-final words. Previous endeavours have shown that consistent with E-Z Reader's assumptions; frequency and predictability have additive effects on the processing of line-initial words (Parker & Slattery, 2019). Our goal here was to examine frequency and predictability effects for line-final words to provide benchmark findings for the next generation of eye movement models that look to simulate eye movements across line boundaries. Our initial linear mixed-effects analysis of the Provo Corpus indicated that line-final words receive shorter reading times than intra-line words. While there was evidence of frequency and predictability effects for intra-line words, results were mixed for line-final words and likely confounded by a potential lack of power to detect small effects due to increased noise and uncertainty around estimates or experimental control over variables, such as word length. To address these limitations, we conducted a pre-registered eye movement experiment where we manipulated frequency, predictability, and target word position. In line with our Provo analysis, reading times were shorter for line-final words. Furthermore, there were clear additive effects of frequency and predictability for both intra-line and line-final target words. These findings have strong implications for accounts of shorter reading times on line-final words and for expanding models of eye movement control to reading at line boundaries.

The most consistent finding reported across both studies is the observation that both readers' single fixation and gaze durations decrease for line-final words. Shorter fixations have not only been reported for line-final words (Tiffin-Richards & Schroeder, 2018) but also for the final fixation on a line (Abrams & Zuber, 1972; Adedeji et al., 2021; Parker, Nikolova et al., 2019; Parker, Slattery, et al., 2019; Rayner, 1977). It has been suggested that shorter line-final reading times and line-final fixations are the result of readers preparing the oculomotor system to initiate a return-sweep (Mitchell et al., 2008). In its strongest form, the return-sweep planning account would suggest that line-final fixations are uninvolved in language processing. Consistent with this suggestion, Hofmeister (1997) reported that text degradation (i.e., stimulus quality) did not affect line-final fixation duration. The return-sweep planning account is, however, extremely difficult to reconcile with findings from the current study as frequency and predictability effects emerge for line-final words, indicating that fixations on these words are being terminated based on lexical properties of the line-final word[11]. The observation that word-level properties influence reading times on line-final words is not novel to this study. Parker et al. (2023) reported longer line-final fixation durations when low-frequency targets are positioned at the end of the line compared to a condition where low-frequency words are positioned at the start of a line. Echoing Parker et al.'s conclusions on line-final reading times, it is indeed time to abandon the claim that reading times on words appearing at the end of the line are uninvolved in lexical processing.

If return-sweep preparation and reduced lexical processing are not the cause of shorter reading times for line-final words and shorter line-final fixations, then what is? A competing account has been put forward by Rayner (1977) and suggests that shorter line-final fixations may be due to the absence of a word to the right of the current fixation, eliminating the need to process parafoveal information of the upcoming word. Tiffin-Richards and Schroeder (2018) reported evidence consistent with this notion. They reported that beginning readers in Grade 2 did not show the same reduction in fixation durations as did older child readers (e.g., children in Grade 3) when reading at line boundaries. Given the assumption that parafoveal

[11] Line-final fixation durations have also been shown to be influenced by reading skill, further suggesting that language processing terminates fixations on line-final words (Parker & Sattery, 2021).

processing capacity develops with expertise and proficiency (Häikiö et al., 2009; Marx et al., 2015; Pagán et al., 2016; Tiffin-Richards & Schroeder, 2015), the lack of a decrease in reading times on line-final words for the youngest of children may reflect their reliance on foveal processing. Thus, there is no benefit of a reduced need for parafoveal processing when fixating a line-final word for the youngest of reading. Nevertheless, because Tiffin-Richards and Schroeder did not directly manipulate parafoveal load, it is difficult to draw firm conclusions on the matter. There is also evidence that is inconsistent with Rayner's explanation. Parker and Slattery (2021) reported that spelling ability, a measure that is hypothesised to index parafoveal processing (e.g., Slattery & Yates, 2018; Veldre & Andrews, 2015), was unrelated to line-final fixation durations. If Rayner's account holds, then we might expect that better spellers would show shorter line-final fixations as they would benefit from a reduction in parafoveal load but this was not the case. Instead, in the absence of strong evidence, we entertain several other explanations for shorter line-final fixations. We have already stated that shorter reading times on line-final words may reflect a reduction in skipping costs during line-final fixations. Alternatively, it may be that shorter reading times on line-final words reflect reduced effects of lateral masking. Within psycholinguistics, lateral masking refers to the interference that an adjacent letter has on the letter being processed (e.g., Townsend et al., 1997). During reading, words are available in upcoming parafoveal vision and the visual properties of the upcoming word may impact the processing of the foveal word. However, when processing a line-final word, there is no adjacent word to the right of fixation that could interfere with foveal processing. Consequently, this may reduce line-final reading times. A remaining explanation could be that readers have learned to terminate line-final fixations earlier than they would during intra-line reading as they can conduct additional lexical processing during the return-sweep, which is considerably longer than an intra-line reading saccade. We would like to note, however, that this explanation may be difficult to incorporate within the E-Z Reader architecture given that the completion of *L1* triggers saccade execution. It may be possible that an additional mechanism involving new parameters and additional assumptions for E-Z Reader (e.g., sampling from an *L1* or fixation distribution with shorter means in the case of a return-sweep) may be capable of accurately describing the data. Finally, our exploratory analysis of refixation and regression rates points towards a fourth potential account of shorter line-final reading times. This account suggests that shorter reading times on line-final words stem from a reduction in failures of post-lexical integration processes. Here, the pause in the incoming stream of new words that occurs at the end of a line provides additional time for post-lexical integration processes thereby reducing comprehension breakdowns at these locations and avoiding the associated time costs (we expand on this explanation below).

To expand computational models of eye movement control, we examined frequency and predictability effects for intra-line words and line-final words. In our corpus-style and experimental work, we replicated shorter reading times on high-frequency and high-predictability intra-line target words. Furthermore, these had additive effects, replicating much of the published literature (Altarriba et al., 1996; Ashby et al., 2005; Kennedy et al., 2013; Miellet et al., 2007; Rayner et al., 2004; Rayner et al., 2001; Slattery et al., 2012). The consistency between studies diverged when looking at frequency and predictability effects for line-final words. For instance, predictability effects were absent for line-final words in the Provo Corpus, but the eye movement experiment indicated clear evidence of both frequency and predictability effects for line-final words. We attribute the failure to find predictability effects in the Provo Corpus to a lack of power stemming from the increased noise in Provo reading times, a restricted range of cloze values entering the analysis (i.e., few high cloze proability words), or it could reflect a lack of control over variables entering the analysis. As such, we place more emphasis on the interpretation of our experimental work. At face value, this pattern of results may seem to coincide with E-Z Reader's additive assumption on frequency and predictability, but as they are currently implemented, no model can account for the reading of line-final words.

Given that E-Z Reader may be able to account for frequency and predictability effects for line-final words without an additional assumption, the remaining effect it needs to account for is the observation that line-final fixations are shorter than intra-line fixations. That said, two assumptions within E-Z Reader may already be able to account for this observation. First, the reduction in duration for line-final fixations may represent the elimination of the need to process parafoveal information for the upcoming word. Therefore, it may be that readers cannot incur skipping costs during fixations on line-final words. Recall that skipping costs refer to the observation that fixations prior to a skip tend to be longer than fixations that occur on adjacent words and that E-Z Reader predicts inflated fixations on word *n* prior to skipping word *n+1*. In the case of line-final words, readers cannot plan a skip and instead must initiate a return-sweep. With the added assumption that readers cannot incur skipping costs when fixating line-final words, EZ Reader may then be able to at least partially explain the current findings with the following assumption: when a line-final word is identified, a saccade is programmed to the next line-initial word. However, because this lies far outside of the parafovea, there can be no skipping cost incurred as a result of parafoveally processing the line-initial word. Second, E-Z Reader 10 predicts that post-lexical integration will fail if word *n+1* is identified prior to the integration stage completing on word n. However, when word *n* is line-final, word *n+1* won't be identified until after the return-sweep. This will provide considerable additional time for integration to complete making it far less likely to have integration failures associated with the processing of line-final words. Indeed, the current finding of reduced refixations to and regressions from line-final words compared to intra-line words is in line with this modeling assumption. Thus, referring back to Parker and Slattery's

(2019) assumptions about lexical processing for line-initial words, it would seem as though only a single assumption would be required to model both reading at the end and the start of the line if the skipping cost or integration accounts were true: that no lexical processing of the first word on a new line can occur until there is a fixation on this new line that places the first word within the fovea or parafovea.

While assumptions about skipping costs and integration failure within E-Z Reader may be able to explain our findings, that does not mean that our results are incompatible with other models of eye movement control, such as SWIFT (e.g., Engbert et al., 2005) and OB1-Reader (e.g., Snell et al., 2018). We consider our findings in light of each of these models in turn.

While E-Z Reader assumes that lexical processing is serial, SWIFT assumes that multiple words falling within efficient vision can be processed in parallel. While E-Z Reader assumes that fixation times are strictly influenced by the lexical properties of the fixated word, SWIFT takes a more nuanced approach. Saccade timing and, as a consequence, fixation durations are regulated by an autonomous timer that maintains a preferred reading speed, where fixation durations are generated from a Gaussian distribution. The saccades within the model are targeted towards words based on their patterns of lexical activation, which is moderated by word frequency and predictability. Frequency and predictability, however, are assumed to only occasionally influence oculomotor processes. SWIFT assumes that word-skipping is driven primarily by word length. However, a word's frequency and predictability can influence the selection of saccade targets, producing increased skipping rates for words that are frequent or highly predictable. Within the model, predictability is independent of visual input, meaning that effects of predictability can occur earlier than those of frequency and act via a process of foveal inhibition. Thus, SWIFT predicts neither additive or interactive effects of frequency and predictability. Similar to E-Z Reader, SWIFT also predicts skipping costs. However, SWIFT assumes that because longer fixations afford more parafoveal processing of n+1, it is less likely to compete for saccade target selection and more likely to be skipped.

Currently, fixation durations within SWIFT are controlled by a random timer. To allow for line position effects to emerge within the model it may be necessary for this timer to vary as a function of line position. Alternatively, the timer could be impacted by the number of words available for processing within the span of attention. That is, when a reader is fixating the last word or two on the line there may be fewer words within the attentional span which could in turn increase the processing speed and thus decrease the random timer. The SWIFT 3 model (Schad & Engbert, 2012) allows for the attentional window of processing to be modified based on foveal processing difficulty. However, while SWIFT 3 allows difficult foveal processing to reduce parafoveal processing, it does not allow for easy parafoveal processing to increase foveal processing. It would appear then that SWIFT may require additional assumptions to account for reduced fixation durations at the end of lines of text. As with our potential explanations within E-Z Reader, simulations of SWIFT will be required to draw firm conclusions on the matter.

The OB1-Reader model (Snell et al., 2018), which integrates ideas from models of visual word recognition and eye movement control during reading, is not too dissimilar from SWIFT in that it assumes that multiple words falling within the attentional input window can be processed in parallel. Using low-level cues, such as the number of to be recognised words and word length, OB1-Reader maps these words onto a spatiotopic sentence-level representation. Because the model assumes open bigram coding of letters, where the word *page* can activate nodes for *pa*, *pg*, *pe*, *ag*, *ae*, and *ge*, OB1-Reader also assumes parallel processing at the letter level. Much like with E-Z Reader and SWIFT, the activation of a word node within OB1-Reader is influenced by its length, frequency, and contextual predictability and once its activation reaches a recognition threshold, it is identified[12] The effects of frequency and predictability within OB1-Reader are interactive and vary as a function of word length. Within OB1-Reader, saccades are generated based on random sampling of a Gaussian distribution, where the range is larger when a word has been recognised. As such, lexical processing influences when the eyes are moved. Just like in SWIFT, a word's frequency and predictability can influence the selection of saccade targets, producing increased skipping rates for words that are frequent or highly predictable.

Given that OB1-Reader also incorporates a random timer for saccade execution, similar adaptions as we suggest for SWIFT may allow for line position effects to emerge; that is, the timer could vary as a function of position in a line or it could be impacted by the number of words available in the attentional input window. However, OB1-Reader may also predict shorter line-final fixations based on its current implementation. Within OB1-Reader foveal word recognition is hampered by orthographically unrelated information in the parafovea (e.g. Snell, Vitu & Grainger, 2017; Snell & Grainger, 2018). At the end of the line, then there is no (rightward) parafoveal information available to readers and this may speed up foveal processing, thus generating shorter line-final reading times[13].

While there are ways in which SWIFT and OB1-Reader may be able to model differences in line-final reading times, they are somewhat incompatible with the additive effects of word frequency and predictability that we reported in our pre-registered eye movement experiment. SWIFT remains agnostic on the nature of this interaction while OB1-Reader assumes an interaction and neither of these assumptions map onto our empirical findings. However, E-Z Reader assumes additive effects of these two variables and it will likely have a much easier time modeling the observed frequency and predictability effects with minor changes to

[12] Note that the rate of lexical processing within OB1-Reader is also driven by orthographic overlap between parafoveal and foveal words, where overlap has a facilitatory effect on foveal word processing and in turn reduces fixation durations.

[13] We would like to thank Joshua Snell for highlighting this possibility.

its serial architecture if this model is extended to simulate eye movements across line boundaries.

Conclusion

Fifty years' worth of eye movement research has shown that eye-tracking can be used to examine fundamental questions about the cognitive, visual, and perceptual processes underlying reading. These findings have led to the development of sophisticated computational models that make specific and testable predictions about eye movements during reading. Despite models, such as E-Z Reader and SWIFT, dominating the field for approximately 20 years they are still only capable of simulating the reading of single lines of text. The studies reported here aim to inform the next generation of models as they look to simulate eye movements across multiline texts by examining how placeholders for lexical processing (i.e., frequency and predictability) influence fixation behaviour for words occurring prior to a line boundary.

The most consistent finding across our linear mixed-effects analysis of the Provo Corpus and our pre-registered eye movement experiment is that reading times are shorter on line-final words. There exist several potential explanations for this, such as reduced engagement in parafoveal preview, absent skipping costs, additional processing time during a return-sweep, or reduced lateral masking. Future studies will need to investigate these explanations, but our observation that reading times on line-final words are influenced by properties of the fixated word strongly indicates that a lack of engagement in lexical processing is not responsible for the observed shorter reading times on line-final words. Furthermore, the additive effects of frequency and predictability coincide with E-Z Reader's assumptions about these lexical variables and we suggest that an additional assumption that skipping and integration failure costs are either absent or reduced for line-final words may be able to account for the results observed in the current work.

Acknowledgements

We would like to thank Melchi Canlas and Xinxin Yan for their assistance with data collection.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Ethical approval statement

The experimental procedure was granted ethical approval by the UCL Department of Experimental Psychology's Ethics Chair, ethics application number: EP_2021_015.

References

Abrams, S. G., & Zuber, B. L. (1972). Some Temporal Characteristics of Information Processing during Reading. *Reading Research Quarterly*, *8*(1), 40–51.

Adedeji, V. I., Vasilev, M. R., Kirkby, J. A., & Slattery, T. J. (2021). Return-sweep saccades in oral reading. *Psychological Research*.

AlJassmi, M. A., Warrington, K. L., McGowan, V. A., White, S. J., & Paterson, K. B. (2022). Effects of word predictability on eye movements during Arabic reading. *Attention, Perception & Psychophysics*, *84*(1), 10–24.

Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The Influence of Lexical and Conceptual Constraints on Reading Mixed-Language Sentences: Evidence from Eye Fixations and Naming Times. *Memory & Cognition*, *24*(4), 477–492.

Angele, B., Laishley, A. E., Rayner, K., & Liversedge, S. P. (2014). The Effect of High- and Low-Frequency Previews and Sentential Fit on Word Skipping During Reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*(4), 1181–1203.

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

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364–390.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Brysbaert, M., Vitu, F., & Drieghe, D. (2005). Word skipping: Implications for theories of eye movement control in reading. In *Cognitive processes in eye guidance*. Oxford University Press.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28.

Dambacher, M., Slattery, T. J., Yang, J., Kliegl, R., & Rayner, K. (2013). Evidence for direct control of eye movements during reading. *Journal of Experimental Psychology. Human Perception and Performance*, *39*(5), 1468–1484.

Dickey, J. M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*, 204–223.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655.

Engbert, R., & Kliegl, R. (2011). Parallel graded attention models of reading. In *The oxford handbook of eye movements.* Oxford University Press.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A Dynamical Model of Saccade Generation During Reading. *Psychological Review*, *112* (4), 777–813.

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency Drives Lexical Access in Reading but Not in Speaking: The Frequency-Lag Hypothesis. *Journal of Experimental Psychology. General*, *140* (2), 186–209.

Häikiö, T., Bertram, R., Hyönä, J., & Niemi, P. (2009). Development of the letter identity span in reading: Evidence from the eye movement moving window paradigm. *Journal of Experimental Child Psychology*, *102* (2), 167–181.

Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association*, *82* (400), 1147–1149.

Hofmeister, J. (1997). *Tüber Korrektursakkaden beim Lesen von Texten und bei leseähnlichen Aufgaben. (On corrective saccades in reading and reading-like tasks.).* Rheinisch-Westfälische Technische Hochschule, Aachen.

Hofmeister, J., Heller, D., & Radach, R. (1999). The return sweep in reading. *Current Oculomotor Research: Physiological and Psychological Aspects*, 349–357.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal Word Processing during Eye Fixations in Reading: Effects of Word Frequency. *Perception & Psychophysics*, *40* (6), 431–439.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87* (4), 329–354.

Kennedy, A., Pynte, J., Murray, W. S., & Paul, S.-A. (2013). Frequency and predictability effects in the Dundee Corpus: An eye movement analysis. *Quarterly Journal of Experimental Psychology*, *66* (3), 601–618.

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

Kuperman, V., Dambacher, M., Nuthmann, A., & Kliegl, R. (2010). The effect of word position on eye-movements in sentence and paragraph reading. *Quarterly Journal of Experimental Psychology*, *63* (9), 1838–1857.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, *4* (1), 6–14.

Luke, S. G., & Christianson, K. (2017). The Provo Corpus: A large eye-tracking corpus with predictability norms. *Behavior Research Methods*, *50* (2), 826–833.

Marx, C., Hawelka, S., Schuster, S., & Hutzler, F. (2015). An incremental boundary study on parafoveal preprocessing in children reading aloud: Parafoveal masks overestimate the preview benefit. *Journal of Cognitive Psychology*, *27*(5), 549–561.

McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research*, *28*(10), 1107–1118.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578–586.

Miellet, S., Sparrow, L., & Sereno, S. C. (2007). Word frequency and predictability effects in reading French : An evaluation of the E-Z Reader model. *Psychonomic Bulletin & Review*, *14*(4), 762–769.

Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the Selective Reanalysis hypothesis. *Journal of Memory and Language*, *59*(3), 266–293.

Nicenboim, B., Schad, D., & Vasishth, S. (2021). An introduction to bayesian data analysis for cognitive science. *Under Contract with Chapman and Hall/CRC Statistics in the Social and Behavioral Sciences Series*.

Pagán, A., Blythe, H. I., & Liversedge, S. P. (2016). Parafoveal Preprocessing of Word Initial Trigrams During Reading in Adults and Children. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *42*(3), 411–432.

Parker, A. J., Kirkby, J. A., & Slattery, T. J. (2017). Predictability effects during reading in the absence of parafoveal preview. *Journal of Cognitive Psychology*, *29*(8), 902–911.

Parker, A. J., Nikolova, M., Slattery, T. J., Liversedge, S. P., & Kirkby, J. A. (2019). Binocular coordination and return-sweep saccades among skilled adult readers. *Journal of Vision*, *19*(6), 10.

Parker, A. J., Räsänen, M., & Slattery, T. J. (2023). What is the optimal position of low-frequency words across line boundaries? An eye movement investigation. *Applied Cognitive Psychology*, *37*(1), 161–173.

Parker, A. J., & Slattery, T. J. (2019). Word Frequency, Predictability, and Return-Sweep Saccades: Towards the Modeling of Eye Movements During Paragraph Reading. *Journal of Experimental Psychology. Human Perception and Performance*, *45*(12), 1614–1633.

Parker, A. J., & Slattery, T. J. (2021). Spelling ability influences early letter encoding during reading: Evidence from return-sweep eye movements. *Quarterly Journal of Experimental Psychology*, *74*(1), 135–149.

Parker, A. J., Slattery, T. J., & Kirkby, J. A. (2019). Return-sweep saccades during reading in adults and children. *Vision Research*, *155*, 35–43.

Pynte, J., Kennedy, A., & Ducrot, S. (2004). The influence of parafoveal typographical errors on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1-2), 178–202.

Rayner, K. (1975). Parafoveal identification during a fixation in reading. *Acta Psychologica*, *39*(4), 271–281.

Rayner, K. (1977). Visual Attention in Reading: Eye Movements Reflect Cognitive Processes. *Memory & Cognition*, *5*(4), 443–448.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, *124*(3), 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The Effects of Frequency and Predictability on Eye Fixations in Reading: Implications for the E-Z Reader Model. *Journal of Experimental Psychology. Human Perception and Performance*, *30*(4), 720–732.

Rayner, K., & Duffy, S. A. (1986). Lexical Complexity and Fixation Times in Reading: Effects of Word Frequency, Verb Complexity, and Lexical Ambiguity. *Memory & Cognition*, *14*(3), 191–201.

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye Movements and Word Skipping During Reading: Effects of Word Length and Predictability. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(2), 514–528.

Rayner, K., & Well, A. D. (1996). Effects of Contextual Constraint on Eye Movements in Reading: A Further Examination. *Psychonomic Bulletin & Review*, *3*(4), 504–509.

Reichle, E. D. (2011). Serial-attention models of reading. In *The oxford handbook of eye movements*. Oxford University Press.

Reichle, E. D. (2021). *Computational Models of Reading: A Handbook*. Oxford University Press.

Reichle, E. D., & Drieghe, D. (2013). Using E-Z Reader to Examine Word Skipping During Reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(4), 1311–1320.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *The Behavioral and Brain Sciences*, *26*(4), 445–476.

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*(2), 177–206.

Reingold, E. M., & Sheridan, H. (2018). On using distributional analysis techniques for determining the onset of the influence of experimental variables. *Quarterly Journal of Experimental Psychology*, *71*(1), 260–271.

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*(1), 128–142.

Sereno, S. C., Hand, C. J., Shahid, A., Yao, B., & O'Donnell, P. J. (2018). Testing the limits of contextual constraint: Interactions with word frequency and parafoveal preview during fluent reading. *Quarterly Journal of Experimental Psychology*, *71*(1), 302–313.

Sheridan, H., & Reingold, E. M. (2012). The time course of predictability effects in reading: Evidence from a survival analysis of fixation durations. *Visual Cognition*, *20*(7), 733–745.

Slattery, T. J., Pollatsek, A., & Rayner, K. (2007). The effect of the frequencies of three consecutive content words on eye movements during reading. *Memory & Cognition*, *35*(6), 1283–1292.

Slattery, T. J., & Vasilev, M. R. (2019). An eye-movement exploration into return-sweep targeting during reading. *Attention, Perception & Psychophysics*, *81*(5), 1197–1203.

Slattery, T. J., & Yates, M. (2018). Word skipping: Effects of word length, predictability, spelling and reading skill. *Quarterly Journal of Experimental Psychology*, *71*(1), 250–259.

Snell, J., & Grainger, J. (2018). Parallel word processing in the flanker paradigm has a rightward bias. *Attention, Perception, & Psychophysics*, *80*, 1512–1519.

Snell, J., Leipsig, S. van, Grainger, J., & Meeter, M. (2018). OB1-reader: A model of word recognition and eye movements in text reading. *Psychological Review*, *125*(6), 969.

Snell, J., Vitu, F., & Grainger, J. (2017). Integration of parafoveal orthographic information during foveal word reading: Beyond the sub-lexical level? *Quarterly Journal of Experimental Psychology*, *70*(10), 1984–1996.

Staub, A. (2015). The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation. *Language and Linguistics Compass*, *9*(8), 311–327.

Staub, A., & Goddard, K. (2019). The Role of Preview Validity in Predictability and Frequency Effects on Eye Movements in Reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *45*(1), 110–127.

Team, R. D. C. (2020). *R: A language and environment for statistical computing. 82*.

Tiffin-Richards, S. P., & Schroeder, S. (2015). Children's and adults' parafoveal processes in German: Phonological and orthographic effects. *Journal of Cognitive Psychology*, *27*(5), 531–548.

Tiffin-Richards, S. P., & Schroeder, S. (2018). The Development of Wrap-Up Processes in Text Reading: A Study of Children's Eye Movements. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *44*(7), 1051–1063.

Townsend, J., Taylor, S., & Brown, D. (1971). Lateral masking for letters with unlimited viewing time. *Perception & Psychophysics*, *10*(5), 375–378.

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*(6), 1176–1190.

Vasilev, M. R., Adedeji, V. I., Laursen, C., Budka, M., & Slattery, T. J. (2021). Do readers use character information when programming return-sweep saccades? *Vision Research*, *183*, 30–40.

Veldre, A., & Andrews, S. (2015). Parafoveal lexical activation depends on skilled reading proficiency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 586.

Voeten, C. C. (2022). *Buildmer: Stepwise elimination and term reordering for mixed-EffectsRegression.* https://CRAN.R-project.org/package=buildmer

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804.

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