Word frequency, predictability, and return-sweep saccades:

Towards the modelling of eye movements during paragraph reading.

Adam J. Parker[1][*]

Timothy J. Slattery[2]

[1]Department of Experimental Psychology

University of Oxford, United Kingdom

[2] Department of Psychology

Bournemouth University, United Kingdom

Word count: 10,407

Corresponding author* at:
Adam J. Parker
Department of Experimental Psychology,
University of Oxford,   Phone: +44 (0) 01865 271444
Anna Watts Building,   Email: adam.parker@psy.ox.ac.uk
Radcliffe Observatory Quarter,
Woodstock Rd,
Oxford, OX2 6GG.

RUNNING HEAD: Word frequency, predictability, and return-sweeps

Abstract

Models of eye movement control during reading focus on the reading of single lines of text. Within these models, word frequency and predictability are important input variables which influence fixation probabilities and durations. However, a comprehensive model of eye movement control will have to account for readers' eye movements across multi-line texts. Line-initial words are unlike those presented mid-sentence; they are routinely unavailable for parafoveal pre-processing. Therefore, it is unclear if and how word frequency and predictability influence reading times on line-initial words. To address this, we present an analysis of the Provo Corpus (Luke & Christianson, 2017) followed by a novel eye-movement experiment. We conclude that word frequency and predictability impact single-fixation and gaze durations on line-initial words. We also observed that return-sweep error (undersweep-fixations) may, among several other possibilities, allow for parafoveal processing of line-initial words prior to their direct fixation. Implications for models of eye movement control during reading are discussed.

Keywords: eye movements; predictability; word frequency; return-sweeps; reading.

Statement of public significance

Readers typically process the word at the point of fixation and the word to the right. When readers move between lines of text they are unable to process information at the start of a new line until it is fixated. This suggests a different time course of processing for information at the start of the line. Indeed, our results suggest that readers compensate for a lack of preprocessing for line-initial words as their fixation times are longer on these words. This is despite the fact that word-based properties (frequency and predictability) influence reading times on line-initial words as they would if they were presented within a line. The observation of predictability effects for line-initial words has implications for accounts of word predictability effects that argue readers only exhibit shorter reading times on predictable words if they are available for parafoveal preprocessing.

Over the last forty years, much has been learned about the cognitive processes involved in reading by examining reader's eye movements (Liversedge & Findlay, 2000; Rayner, 1998, 2009a). One reason for this is that computational models have been developed which yield clear, testable hypotheses about eye movement control during reading (Rayner, 2009b). Yet, models of eye movement control during reading, such as E-Z Reader (Riechle, Pollatsek, Fisher, & Rayner, 1998; Reichle & Sheridan, 2015), and SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005; Laubrock, Kliegl, & Engbert, 2006), have focused on fitting data from single sentence reading studies which are devoid of return-sweeps.

Return-sweeps are a saccadic eye movement that take readers' fixation point from the end of one line to the start of the next line. They are a natural part of the reading process, with approximately 20% of all fixations being preceded or followed by a return-sweep (Rayner, 1998). By excluding these saccades, models are only able to simulate reading of single-line texts, which is far from the experience faced by readers outside of a laboratory setting. When considering an example such as J.K. Rowling's *Harry Potter and the Philosopher's Stone* readers are required to make 30 return-sweeps on the first page alone without any re-reading. This translates as approximately 6,500 return-sweeps across the entire 1997 Bloomsbury paperback edition. The prevalence of return-sweeps during natural reading, highlights the need for models of eye movement control to incorporate mechanisms for modelling return-sweeps if they are to simulate a truly natural reading experience.

To account for return-sweeps within computational models, we must understand eye fixation behavior before and after a return-sweep. The current work focusses on how lexical processing is supported by fixation behavior after a return-sweep. The first fixations on a line fall into two distinct categories (Parker, Kirkby, & Slattery, 2017). The first are fixations that accurately land close to the start of the new line and are immediately followed by a normal left to right progressive saccade. These fixations tend to be longer (~280ms) than other

fixations (Heller, 1982; Hawley, Stern, & Chen, 1974; Parker, Slattery, & Kirkby, 2019; Parker, Nikolova, Slattery, Liversedge, Kirkby 2019; Rayner, 1977; Stern, 1978). Like all saccades, return-sweeps are the result of muscular contractions and are subject to error. A saccade's length is the sum of its intended length plus saccadic range error and random error (Findlay & Walker 1999). These saccadic errors are a function of intended saccade length, being larger for longer intended saccades. Since return-sweeps move the eyes a greater distance than other reading saccades they will be more error prone. The second category of line-initial fixations results from this error. These fixations fall short of the start of the line and are followed by a corrective regression toward the beginning of the line (Hofmeister, Helller, & Radach, 1999; Parker et al., 2017). These fixations have been termed "undersweep" fixations (Parker et al., 2017) and are significantly shorter in duration (130-176ms) than other reading fixations (Heller, 1982; Parker et al., 2019). The fact that undersweep-fixations are more prevalent with longer lines of text further supports oculomotor error as their cause (Beymer, Russell, & Orton, 2005; Hofmeister et al., 1999). It has long been assumed that undersweep-fixations are not involved in on-going linguistic processing (Hawley et al., 1974; Shebilske, 1975). This assumption appears warranted given recent evidence showing that ~140 ms is the earliest point at which lexical effects on fixation durations can emerge (Reingold, Reichle, Glaholt, & Sheridan, 2012; Reingold & Sheridan, 2012). However, it is far from clear how undersweep-fixations may impact the processing of the earlier words on the line of text (those appearing to the left of the fixation). For instance, it is possible that parafoveal preview is obtained for the words to the left of undersweep-fixations during these brief pauses. This is one of the main foci of the current paper.

Parker, Kirkby, and Slattery (2017) have suggested that the longer durations of accurate line-initial fixations result from a lack of preview (though see Kuperman, Dambacher, Nuthmann, & Kliegl, 2010; Rayner, 1977; Stern, 1978 for other potential

explanations). Not only are these accurate line-initial fixations longer, in their supplemental analysis, Parker et al. (2017) reported that predictability effects were absent in first-fixation duration, yet present in later measures (i.e. gaze duration). This contradicts a large body of evidence that has reported predictability effects in the earliest eye-tracking measures with single-line reading (i.e. word skipping, first-fixation duration; see Staub, 2015, for a review). These observations have led computational models of eye movement control to assume that predictability acts on an early stage of processing along with word frequency. Therefore, a delayed effect of predictability for line-initial words may present a challenge for models as they attempt to encompass return-sweeps in their predictions.

To model return-sweep saccades, it is essential to understand lexical processing for line-initial words following return-sweeps. Therefore, the current article presents two investigations (one corpus, one experimental) of the time course of word frequency and predictability effects for line-initial words. First, we present a brief summary of frequency and predictability effects in single-line reading studies and the role that parafoveal preview has in regard to these effects, then we examine how the E-Z Reader model is able to account for these effects.

Frequency and predictability influences on fixation times.

Eye movement studies of reading have shown that fixation durations decrease for high-frequency words in comparison to low-frequency words (Angele, Laishley, Rayner, & Liversedge, 2014; Hand, Miellet, O'Donnell, & Sereno, 2010; Inhoff & Rayner, 1986, Just & Carpenter, 1980; Kliegl, Grabner, Rolfs, & Engbert, 2004; Mielet, Sparrow, & Sereno, 2007; Rayner, Ashby, Pollatsek, & Reichle, 2004; Rayner & Duffy, 1986; Slattery, Pollatsek, & Rayner, 2007; Whitford & Titone, 2014). They also decrease for high-predictability words— that is, words embedded within high-contextual constraint sentences (Altarriba, Kroll, Sholl, & Rayner, 1996; Balota, Pollatsek, & Rayner, 1985; Ehrlich & Rayner, 1981; Gollan,

Slattery, Goldenberg, Van Assche, Duyck, & Rayner, 2011; Kliegl et al., 2004; Miellet et al., 2007; Rayner et al., 2004; Rayner, Slattery, Drieghe, & Liversedge, 2011; Rayner & Well, 1996; Slattery & Yates, 2018). Similar predictability effects have been reported for skipping probability during first-pass reading of a word (Brysbaert, Drieghe, & Vitu, 2005). In contrast, frequency has a smaller influence on skipping with effects being confined to cases in which the eyes were fairly close (3-4 character spaces) to the beginning of the target word (see Rayner et al., 2004 for a discussion).

The emergence of these effects in the earliest possible eye movement measures (word skipping, first-fixation duration) indicates that a word's frequency and predictability influence visual word recognition. Survival analyses have shown that both frequency and predictability have rapid effects on distributions of fixation durations (Reingold & Sheridan, 2018). Divergence points for predictability have been reported at 140 ms (Sheridan & Reingold, 2012) which is comparable to the frequency divergence point of 145 ms (Reingold, Reichle, Glaholt, & Sheridan, 2012). In addition, these results are consistent with event-related potentials studies that have demonstrated both word frequency and predictability effects during the N1 component, from 132-192 ms post-stimulus onset (Sereno, Brewer, & O'Donnell, 2003).

Both predictability and frequency clearly influence early eye movement measures. However, these variables may exert their influence via different mechanisms. Despite the widely-held assumption that the cost of processing a low-frequency word should be largely eliminated when the word is made highly predictable, there is no strong evidence of an interaction on eye movement fixation duration measures of reading (Slattery, Staub, & Rayner, 2012). Instead, the majority of eye movement studies report additive effects of frequency and predictability on fixation times (Altarriba et al., 1996; Ashby, Rayner, & Clifton, 2005; Kennedy, Pynte, Murray, & Paul, 2013; Miellet et al., 2007; Rayner et al.,

2004; Rayner, Binder, Ashby, & Pollatsek, 2001; Slattery et al., 2012; Whitford & Titone, 2014). Based on an additive factors logic (Sternberg, 1969), frequency and predictability are influencing independent mechanisms of word recognition. This picture is complicated by the fact that interactive effects have been reported in word skipping (Gollan et al. 2011; Hand et al. 2010; Rayner et al. 2004) which is a measure that is greatly impacted by parafoveal pre-processing. Therefore, it is possible that, in the absence of parafoveal pre-processing (as is routinely the case with line-initial words), interactions between frequency and predictability will arise in fixation duration measures. Such an interaction would be important to account for in computational models.

Lexical influences and the role of preview.

Once fixated, the time that readers spend on a word, depends on whether the word was visible in the parafovea prior to fixation (Rayner, 1998, 2009; Schotter, Angele, & Rayner, 2012). Evidence from the boundary change paradigm (Rayner, 1975) confirms the importance of parafoveal pre-processing. In this paradigm, an invisible boundary is placed just left of a target word. Before readers cross this boundary, a valid or invalid preview occupies the target location. Once the eyes cross the boundary, the preview is replaced by the target. Fixations on the target are longer following invalid than valid previews. It has been observed that the preview benefit exists only for the word at the target location of the saccade (McDonald, 2006) and that readers do not gain information below fixation (Pollatsek, Raney, Lagasse, & Rayner, 1993). Pollatsek et al. reported that the reading of multiline text was no slower when letters on the line below fixation were masked with strings of Xs, visually similar letter strings, or when the text changed when compared to an unmasked condition. Furthermore, when searching texts for errors, participants very infrequently detected errors below fixation (> 10%), supporting the claim that the preview benefit does not occur below fixation.

Results from studies which have manipulated word frequency and preview validity have shown that frequency effects remain even with invalid preview (Kennison & Clifton, 1995; Choi & Gordan, 2013; Reingold et al., 2012; Risse & Kliegl, 2014). The most powered of these (Reingold et al., 2012), reported a first-fixation duration frequency effect of 20ms (58ms in gaze) with valid previews and a smaller but still significant 9ms (47ms in gaze) with invalid previews.

In contrast, several studies that investigated preview validity and word predictability have reported an interaction, with predictability effects being dependent on a strong orthographic overlap between preview and target (Balota, Pollatsek, & Rayner, 1985; Choi, Lowder, Ferriera, Swaab, & Henderson, 2017; Juhasz, White, Liversedge, & Rayner, 2008; Schotter, Lee, Reiderman, & Rayner, 2015; Veldre & Andrews, 2018; White, Rayner, & Liversedge, 2005). For example, Balota et al. found high rates of skipping for predictable words only when preview was valid or an orthographically similar non-word. This was paralleled in first-fixation and gaze duration where predictability benefits were present only with valid or orthographically similar previews.

The dissociation between effects of frequency and predictability with invalid preview suggests that these variables act on different stages of processing. Staub and Goddard (2019) presented data further supporting the decoupling of word frequency and predictability effects. They orthogonally manipulated target word frequency, predictability and preview validity and found that predictability effects were eliminated by invalid previews, while frequency effects were not. Staub and Goddard reasoned that predictability effects require ambiguity in the perceptual evidence that is available during early orthographic processing. When early processing takes place foveally (i.e. after a direct fixation following an invalid preview), the perceptual evidence outweighs prediction. Meanwhile, frequency effects remain because they act on both early and late stages of processing.

The absence of preview for line-initial words prior to a return-sweep is crucially different from the lack of valid preview in studies that utilize the boundary paradigm. With the boundary paradigm, a lack of valid preview necessitates the presence of an invalid preview. Recent research indicates that "preview benefit" effects in such boundary paradigm studies are actually a complex mixture of valid preview benefits and invalid preview costs (Hutzler et al., 2013; Kliegl, Hohenstein, Yan, & McDonald 2013; Marx, Hawelka, Schuster, & Hutzler, 2015; Vasilev, Slattery, Kirkby, & Angele, 2018). Therefore, exploring how the absence of preview during return-sweeps influences lexical processing will also advance our understanding of "preview benefit" effects within the boundary paradigm.

This is highlighted in recent research (Parker et al., 2017) which reported predictability effects in gaze duration but not in first fixation duration for line-initial words following a return-sweep when preview was unavailable (i.e. undersweep-fixations were excluded) but not invalid. This finding is difficult to explain based on orthographic ambiguity present during parafoveal processing as these words were not processed parafoveally. However, the greater percentage of non-optimal landing positions on line-initial words in these cases may have contributed to foveal orthographic ambiguity which would be capable of reconciling these apparently contradictory findings. Though there exists a second explanation for reconciling these effects based on shifts of belief in sentence meanings. In this account, readers maintain multiple hypotheses about the underlying meaning of a sentence and update the strength of their belief in each hypothesis as new bottom-up information becomes available (see Levy, 2008). Parker et al. (2017) posit that this belief shift happens as soon as new foveal or parafoveal information becomes available. In the case of a predictable target word with an invalid preview, this belief shift happens during parafoveal processing and shifts away from belief in the predictable word—thereby removing the predictability effect in first pass reading measures. With line-initial words, there is no

invalid information to cause such a shift away from the predictable word which is why the predictability effect remains in gaze duration.

Staub and Goddard (2019, Exp 2.) explored a version of the Parker et al. (2017) explanation above that they termed the "semantic suppression" hypothesis. They reasoned that a belief shift away from a predictable word required inconsistent semantic activation from a preview stimulus. Therefore, with non-word previews which have no semantic representations, there would be no belief shift and predictability effects of the target would should remain. As with their Experiment 1 which used word invalid previews, they again reported null effects of target word predictability when invalid previews were non-words. This version of a "semantic suppression" hypothesis is perhaps consistent with Parker et al. (2017), however it is more limited than the one originally envisioned. Here we specify our version of a semantic suppression account as it pertains to predictability effects in reading and how it differs from Staub and Goddard (2019).

Our hypothesis on the emergence of the predictability effect in reading agrees with many of the theoretical assertions from Staub (2015) and Staub and Goddard (2019). That is, we believe that lexical prediction during reading involves graded activation of multiple representations rather than just a single representation. Additionally, since language is not always predictable, we agree that for comprehension to be faithful to the input, predictions should be easily overwritten by bottom-up information. Staub and Goddard assert that predictability effects on eye movements during reading require that weak semantic priors (created from sentence context) be combined with weak (ambiguous) bottom-up evidence from parafoveal pre-processing. Once a word is in foveal vision, they assert that there is too little bottom-up ambiguity in the visual input for weak semantic priors to significantly impact processing. Our explanation of the predictability effect differs from Staub and Goddard in the following ways:

1. Prediction overwriting can happen during parafoveal processing of invalid previews.

2. There exists enough ambiguity in foveal vision that weak semantic priors can still benefit foveal processing in the true absence of parafoveal pre-processing.

3. The absence of semantic activation during parafoveal processing of non-word previews indicates that the next word may be anomalous and is still useful for prediction overwriting[2].

First fixation durations on words during reading are a mixture of single-fixation cases and the first of multiple fixation cases. Often, the duration of the first of multiple fixations will be located further from the optimal viewing position and shorter in duration than single-fixation cases. This is referred to as the inverted optimal viewing position (IOVP) effect (Nuthman, Engbert, Kliegl, 2005, Vitu, McConkie, Kerr, O'Regan, 2001, Vitu, O'Regan, Mittau, 1990). Due to the importance placed on initial fixation durations for evidencing early processing effects, and the variability in these durations due to the IOVP effect, we will be exploring single-fixation durations here for evidence of the earliest predictability effects. That is, these are cases where the readers fixated the target word close enough to an optimal position to avoid the need for a refixation. If predictability effects are found in this measure following a return-sweep, it would be strong evidence in favor of our belief shift explanation of predictability effects over the orthographic ambiguity interpretation offered by Staub and Goddard (2019; see also Staub, 2015). That is, these single-fixation durations following an accurate return-sweep will have provided readers with accurate orthographic information during fixation and without any preview having been available prior to the return-sweep.

Computational models of eye movement control during reading: An overview.

A number of computational models have emerged in attempt to explain the eye movement patterns observed during reading (Legge, Klitz, & Tjan, 1997; Reichle, Warren, &

McConnell, 2009; Reilly & Radach, 2006; Schad & Engbert, 2012; Snell, van Leipsig; Grainger, & Meeter, 2018; see also Reichle, 2006, for an overview of 5 models). However, there is currently no model of eye movements during reading that can account for return-sweep saccades.

Next, we describe one successful model of eye movements during reading, E-Z Reader (Reichle et al, 1998; Reichle & Sheridan, 2015). This model, like all others, does not currently attempt to account for return-sweeps. We then discuss one way it may be extended with a simplifying assumption. Finally, we derive a handful of predictions based on the extended model.

E-Z Reader (Reichle et al, 1998; Reichle & Sheridan, 2015) is built on the assumption that word identification drives the eyes through the text in a strictly serial manner such that the lexical processing of a word cannot begin until all the prior words have been fully processed. Within E-Z Reader, lexical access occurs in two stages. The duration of these stages is an additive function (Rayner et al., 2004; see also Slattery et al., 2012) of the word frequency and contextual predictability[1] of the word being processed. The completion of the first stage, called the "familiarity check", is the signal to start planning a new saccade to the next word (n+1). The completion of the second stage, or "lexical access", is the signal to shift attention to word n+1 and begin the familiarity check on that word in the parafovea. If the labile stage of saccade planning to this parafoveal word completes before the familiarity check on it does, a saccade will be executed to it. If the familiarity check finishes first, the saccade being planned will be cancelled and replaced by a new saccade program to word n+2. In this case, the word n+1 will be skipped. Thus, the main assumptions within E-Z Reader are that lexical processing is the engine driving the eyes, and that attention is allocated to word objects in a strictly serial manner—lexical processing of the next word does not begin until processing of the current word is completed. Moreover, E-Z Reader assumes

that readers gain access to frequency and predictability information about a word at the same point in time—when serial attention is shifted to the word.

Within E-Z Reader, attention for lexical processing is decoupled from the location of the reader's fixation. That is, the eyes can be fixated on word n while attention has moved on to lexically process word n+1. Moreover, because of error in saccade execution (Findlay & Walker, 1999), this decoupling can also occur when the eyes intend to fixate word n+1 but accidentally over or undershoot it resulting in a mislocated fixation (Drieghe, Rayner, & Pollatsek, 2008; Engbert, Nuthmann, Richter, & Kliegl, 2005) on word n or word n+2. With large oculomotor error, the eye's will land far enough away from their intended target and E-Z Reader will implement an immediate corrective saccade to bring the eye's closer to an optimal position. Here we assume that this mechanism would be enough for E-Z Reader to account for undersweep-fixations (i.e. their short duration and subsequent corrective saccade). Note, that while E-Z Reader simulates error in the movement of the eyes, it assumes that there is no error in the movement of attention.

We make one simplifying assumption about return-sweeps: 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 (see Parker, Kirkby, & Slattery, 2017). 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 not result in the start of parafoveal pre-processing of this word, due to it being located in the periphery. Instead, lexical processing (L1) of line-initial words must wait for these words to be both attended and located in the fovea or parafovea (i.e. after execution of the return-sweep). Such a model would predict:

1. The duration of the line-initial fixation following an accurate return-sweep (i.e. no intervening undersweep-fixation) should be longer compared to words fixated during the normal left to right (for English) reading pass within a line.

2. That 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.

3. That the effects of word frequency and predictability would remain the same as for other words. As already mentioned this may not be the case.

The current study.

The E-Z Reader model uses word frequency and predictability as language input variables to simulate eye movements during the reading of a single line of text. Additionally, the model posits that a significant amount of a word's lexical processing happens prior to that word being fixated (i.e. in parafoveal vision). This makes sense for normal left to right reading within a single line of text. However, the target of a return-sweep will lie in the periphery, far outside of the parafovea. In order for models to be advanced to account for return-sweeps, it will be necessary to determine how word frequency and predictability effects emerge following a return-sweep under conditions where parafoveal preview is largely absent—except in cases where there are intervening undersweep-fixations.

The purpose of this article is hence two-fold. First, to investigate the influence of word frequency and predictability for line-initial words so as to inform computational modelling efforts. Given both the findings of Parker et al (2017) and Staub and Goddard (2019) we expect to find frequency effects across all reading time measures. If predictability effects are due to orthographic ambiguity during parafoveal processing, then they should be absent in single-fixation durations but may be present in gaze durations which include refixations from non-optimal initial fixation locations. In contrast, if predictability effects are

the result of Bayesian updating of beliefs in sentence meanings, then they should be present in single-fixation durations and gaze durations.

The second purpose is to examine the potential for undersweep-fixations to provide preview for line-initial words and examine the effect that acquiring preview would have for line-initial predictability effects. Based on the results of Parker et al. (2017), we expect that when return-sweeps land on a line-initial word, these initial fixations will be longer in duration than normal reading fixations and will not exhibit a predictability effect. However, when the return-sweep lands short of the line-initial word and is followed by an immediate regression to the line-initial word, the intervening undersweep-fixation may provide preview of the line-initial word. This would result in shorter first fixation durations on the line-initial word as well as the emergence of a predictability effect in first fixation duration for line-initial words. Such an effect would be crucial for models to account for.

To accomplish these goals, we first present a linear mixed modelling analysis using the Provo Corpus (Luke & Christianson, 2018) followed by a controlled experimental study which aims to verify aspects of the corpus analysis results while also addressing its limitations.

Eye Movement Corpus Analysis

To explore word frequency and predictability effects for line-initial words, we first conducted a corpus style analysis using the Provo Corpus (Luke & Christianson, 2018). The Provo Corpus consist of 55 short multi-line passages, with an average of 50 words (range: 39-62), which were silently read by 84 participants. The data were collected with an SR Research EyeLink 1000 plus eye-tracker sampled at 1000 Hz (for the full description of the method and stimuli, see Luke & Christianson, 2017). This corpus was ideal for our purpose as it contained return-sweeps in the eye movement record and predictability[3] measures were included for all the words in the corpus.

Means for word length, frequency, and predictability are show in Table 1. Fixations shorter than 80 ms and longer than 800 ms were removed (4% of the data) prior to analysis. To determine first-pass reading time measures appropriately for line-initial words in cases of undersweep-fixations, we calculated the reading times as if the undersweep-fixation had not occurred. That is, undersweep-fixations did not end the first pass reading time of the words to their left.

Table 1. Word length, frequency, and predictability means, standard deviations, and correlation coefficients for all words across items, line-initial target words used in the present study, and words entering analyses.

Word Type	Linguistic Variable	Mean	l	f	p
All words in passage	Length	4.8 (2.55)	-		
	Frequency	5.8 (1.41)	-.80***	-	
	Predictability	.41 (0.23)	-.26***	.31***	-
Line-initial target words	Length	6.9 (1.99)	-		
	Frequency	4.5 (0.98)	-.62***	-	
	Predictability	.34 (0.21)	-.22***	.21***	-
Line-initial analyzed words	Length	7.5 (2.17)	-		
	Frequency	4.3 (1.07)	-.62***	-	
	Predictability	.33 (0.21)	-.20***	.21**	-
Line-internal analyzed words	Length	6.3 (2.00)			
	Frequency	4.7 (1.02)	-.59***		
	Predictability	.35 (0.20)	-.13***	-.07***	-

Note. Frequency: Zipf frequency. Predictability: Cloze values. Standard deviations in parentheses. * Correlation is significant at the .05 level. ** Correlation is significant at the .01 level. *** Correlation is significant at the .001 level.

Results

To explore the influence of word frequency and predictability on line-initial words, we considered line-initial words for all lines of text except for the first (10,920 data points). These words were then filtered to exclude function words and punctuated words (23.1%; following Miellet et al., 2007; Whitford & Titone, 2014). When a blink occurred on these words or during an immediately adjacent fixation, it was removed prior to analysis (14.1% of words). We present three different sets of analyses. The first explores the impact that lexical

characteristics of the line-initial word may have on return-sweep landing positions. The second explores the impact of lexical characteristics on single-fixation and gaze durations for line-initial words as well as the role that undersweep-fixations play in the processing of line-initial words. The final set of analyses compares the effect of lexical variables for line-initial words with the other line-internal words.

Eye movement data was analyzed using Bayesian linear mixed effects (BLME) models, constructed using the stan_lmer function from the rstanarm package (version 2.17.4; Goodrich, Gabry, Ali, & Brilleman, 2018) in R (version 3.5.0; R Development Core Team, 2018). To examine the joint effects of word frequency and predictability on the eye movement data, centered word frequency (Zipf) and predictability (Cloze probability) and their interaction were included as fixed effects. As our modeling assumptions are based on E-Z Reader which uses log-transformed word frequency and non-transformed cloze values, analyses reported in the main article were conducted using Zipf frequencies (*log10(frequency per billion words)*) and non-transformed cloze probabilities. However, given that the effect of word predictability on fixation times has been reported to be logarithmic (Smith & Levy, 2013), we include supplemental materials in which cloze probabilities are log-transformed prior to analysis. In these models, centered word length as a control variable and was allowed to interact with word frequency and predictability. In all analyses of the Provo Corpus, we restricted analyses to 4- to 12-letter words (removing 9.9% of line-initial words and 18.7% of line-internal words). All models adopted a full random structure, treating both subjects and items (word token, passage) as random factors, with random intercepts and slopes (Barr, Levy, Scheepers, & Tily, 2013). In all analyses, we assumed Cauchy's prior for effect size; see Abbott and Staub (2015) for discussion. From the Bayesian models, we also computed the 2.5^{th} and 97.5^{th} percentile of the coefficient's posterior distribution. All models had four chains, a warm-up (burn-in) of 500 and 10,000 iterations. Convergence was checked visually

and by using the Gelman-Rubin convergence diagnostic. Converged models had a Rhat statistic of 1 ± 0.1 if the model had converged. The distributions for the lexical variables of analyzed data points are shown in Figure 1.
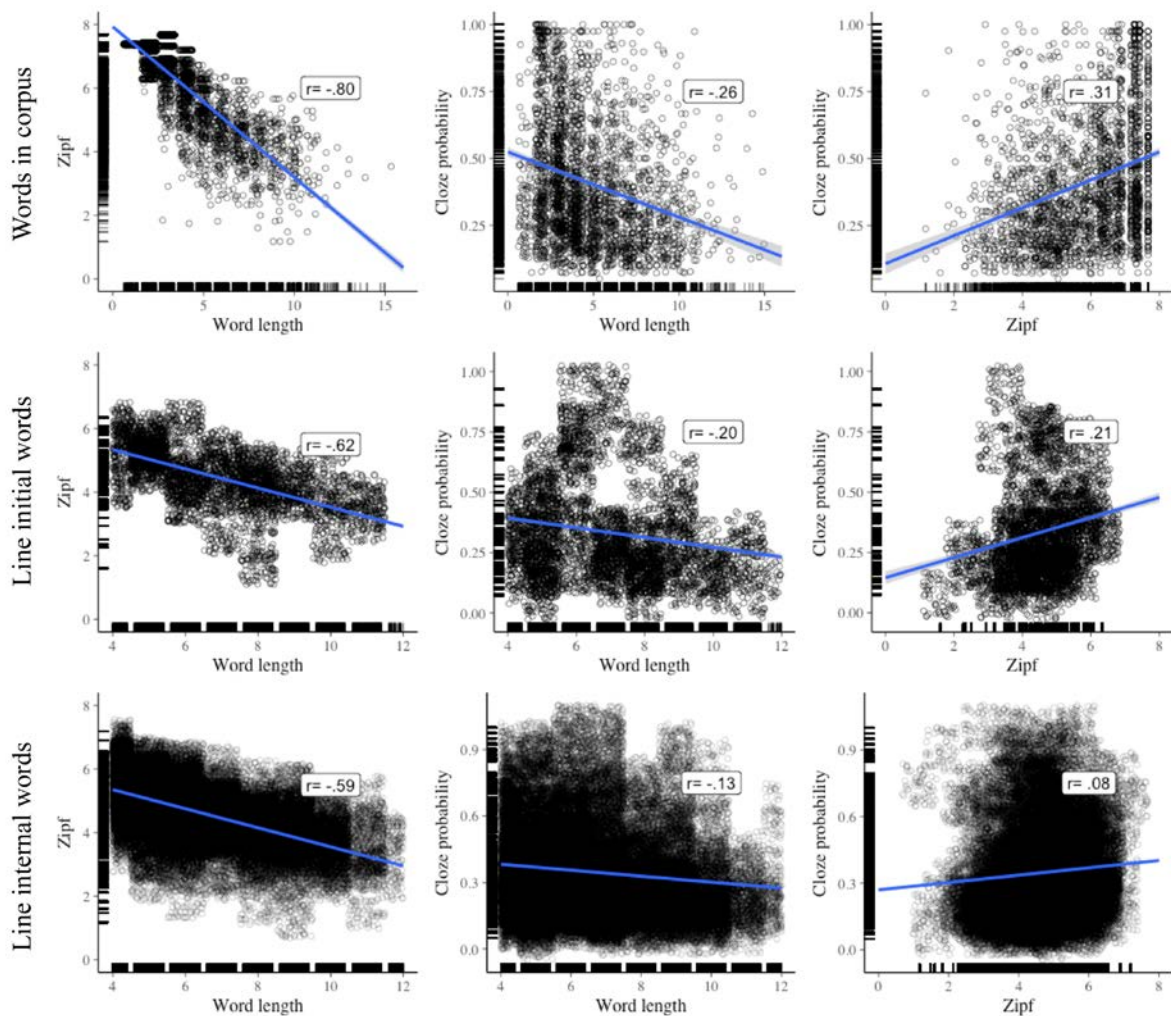

Figure 1. Scatterplots with line of best fit and correlation coefficients for word length, frequency, and predictability for all words in the Provo Corpus, analyzed line-initial words and analyzed line-internal words.

Return-sweep landing position. As shown in Figure 2, the majority of return-sweeps were launched approximately 5.3 characters from the end of the line. Similarly, readers tend to target their return-sweeps towards the start of a line, with the majority of return-sweeps landing approximately 8.3 characters[4] or less from the start of the line. Recent evidence

suggests that readers do not target the centers of line-initial words with their return-sweeps but rather target an area relative to the left margin (Slattery & Vasilev, 2019). However, to examine the possibility that lexical characteristics of the line-initial word may influence return-sweep targeting, we fit a BLME model to landing position data. The model included word frequency, predictability, length, and all possible interactions in the fixed effects structure. Examination of the credibility intervals indicates that return-sweep landing position was uninfluenced by lexical properties of the line-initial word (see Table 2).
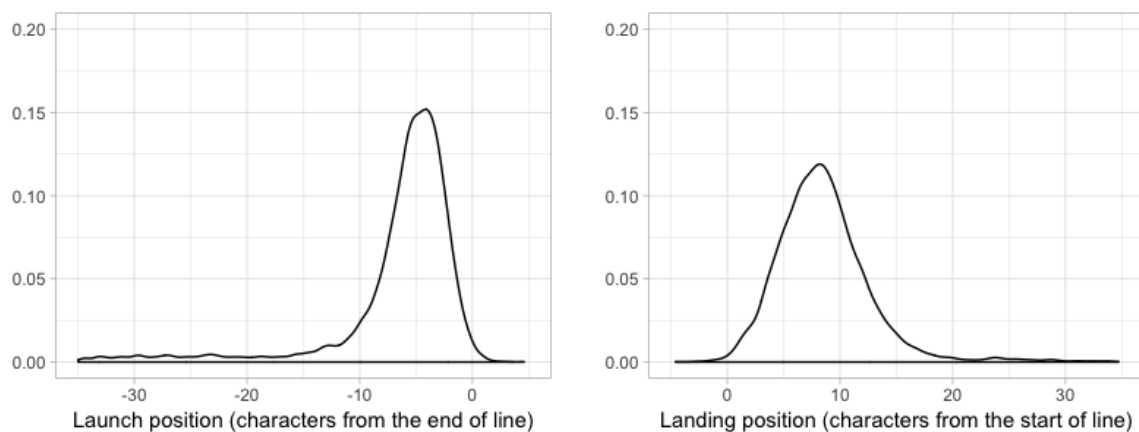

Figure 2. Provo: Density plot for return-sweep launch (left panel) position and landing position (right panel). For return-sweep launch position, zero represents the right margin. Return-sweeps launched from a negative value occurred prior to the end of a line. For return-sweep landing position, zero represents the left margin. As landing position increased, return-sweeps were made further into the line. Approximate average line length: 83 characters.

Table 2. Provo: Bayesian linear mixed effects model results.

| | Return-sweep landing position | | | |
| | | | Percentiles | |
Predictor	b	SD	2.5^{th}	97.5^{th}
(Intercept)	**150.82**	**7.15**	**136.56**	**164.29**
Word frequency (WF)	6.39e-3	.217	-4.21e-1	4.22e-1
Cloze probability (CP)	-20.58	29.69	-79.96	37.84
Word length (WL)	4.96	3.25	-1.66	11.02
WF x CP	6.15e-1	1.26	-1.99	3.06
WF x WL	4.28e-2	1.01e-1	-1.45e-1	2.49e-1
WL x CP	-11.75	16.77	-43.97	23.19
WF x CP x WL	-3.33e-1	5.29e-1	-1.36	7.30e-1

Note. A negative coefficient of posterior mean b (in pixels) indicates landing positions further to the right. The percentiles show 95% credible intervals computed from a Bayesian linear mixed effects models. ***Bold*** values indicate that, given our observed data, there is a 95% probability that the true value of $|b|$ is greater than *0*.

First-pass reading time and undersweep-fixations. We consider two eye movement measures: 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). For reference, single-fixations occurred on 63.6% of the analyzed cases. To assess the impact of word frequency and predictability on these line-initial words, we fit BLME models to each measure. Prior to analysis, measures were log-transformed for normality. Each model included fixed effects for word frequency and predictability. An additional categorical fixed effect was included to code whether participants had made an undersweep-fixation on a given line. Cases in which readers had accurately landed on the line-initial word were coded as 0. Cases in which readers had made an undersweep-fixation were coded as 1. Under this coding scheme, cases in which readers accurately fixated the line-initial word represent the intercept to which undersweep cases were compared. For single-fixation and gaze duration data, undersweep-fixations intervened return-sweeps and fixations on line-initial words on 64.6% and 47.8% of cases respectively.

Due to the uncontrolled nature of lexical variables in corpus data, word length was also included as a fixed effect along with all possible interactions.

Table 3 shows test statistics for reading time BLME models. Model estimates are shown in Figure 3. Examination of credibility intervals indicated that single-fixation durations on line-initial words were shorter when preceded by an undersweep-fixation. However, for the lexical effects and higher-level interactions, there was no reliable effect on single-fixation duration. Turning to gaze duration, there was evidence that frequency exerted an influence, as gaze was shorter for high-frequency words. An effect of undersweep-fixation was also observed, indicating that gaze durations were shorter when readers were able to parafoveally pre-process the line-initial word prior to direct fixation. Additionally, credibility intervals indicated that gaze durations increased with increasing word length. Since the credibility intervals for predictability effects and higher-level interactions included 0, it indicates that these fixed effects had no modulatory effect on gaze durations.

Table 3. Provo: Bayesian linear mixed effect model results for reading time measures.

Predictor	Single-fixation duration				Gaze duration			
			Percentiles				*Percentiles*	
	b	*SD*	*2.5*th	*97.5*th	*b*	*SD*	*2.5*th	*97.5*th
(Intercept)	**2.39**	**1.09e-2**	**2.37**	**2.41**	**2.49**	**1.12e-2**	**2.48**	**2.51**
Word frequency (WF)	-5.30e-3	9.46e-3	-2.38e-2	1.34e-2	**-2.68e-2**	**7.67e-3**	**-4.18e-2**	**-1.19e-2**
Cloze probability (CP)	-2.62e-2	3.76e-2	-9.96e-2	4.78e-2	1.93e-2	2.69e-2	-3.27e-2	7.29e-2
Word length (WL)	1.13e-3	4.48e-3	-7.69e-3	9.93e-3	**3.83e-3**	**3.86e-3**	**3.74e-3**	**1.14e-2**
Undersweep-fixation (UF)	**-5.54e-2**	**1.02e-2**	**-7.55e-2**	**-3.55e-2**	**-1.06e-1**	**7.79e-2**	**-1.21e-1**	**-9.03e-2**
WF x CP	-1.19e-2	5.13e-2	-1.18e-1	8.44e-2	2.72e-2	3.31e-2	-3.85e-2	9.14e-2
WF x WL	-4.55e-3	4.74e-3	-1.40e-2	4.65e-3	-3.13e-3	8.85e-3	-2.05e-2	1.43e-2
CP x WL	3.37e-3	2.95e-2	-5.51e-2	6.04e-2	-3.29e-2	3.21e-2	-9.60e-2	2.94e-2
WF x UF	-7.97e-3	1.11e-2	-3.00e-2	1.36e-2	-2.34e-3	4.05e-3	-1.04e-2	5.51e-3
CP x UF	4.23e-2	4.40e-2	-4.35e-2	1.30e-1	-1.30e-2	2.11e-2	-5.51e-2	2.82e-2
WL x UF	-1.92e-3	5.37e-3	-1.24e-2	8.61e-3	-8.20e-4	4.42e-3	-9.46e-3	7.91e-3
WF x CP x WL	-1.08e-2	3.22e-2	-7.42e-2	5.17e-2	3.16e-2	3.84e-2	-4.36e-2	1.07e-1
WF x CP x UF	4.54e-3	5.81e-2	-6.67e-2	1.65e-1	-1.77e-2	2.31e-2	-6.29e-2	2.74e-2
WF x WL x UF	-3.51e-4	5.70e-3	-1.16e-2	1.08e-2	-7.20e-3	4.74e-3	-1.67e-2	2.03e-2
CP x WL x UF	-1.28e-2	3.44e-2	-8.11e-2	5.48e-2	2.17e-2	2.45e-2	-2.65e-2	6.84e-2
WF x CP x WL x UF	1.06e-2	3.64e-2	-6.10e-2	8.25e-2	1.00e-2	2.65e-2	-4.21e-2	6.17e-2

Note. For single-fixation and gaze duration a negative coefficient of posterior mean *b* (in log milliseconds) indicates shorter reading times. The percentiles show 95% credible intervals computed from a Bayesian linear mixed effects models. **Bold** values indicate that, given our observed data, there is a 95% probability that the true value of *|b|* is greater than *0*.
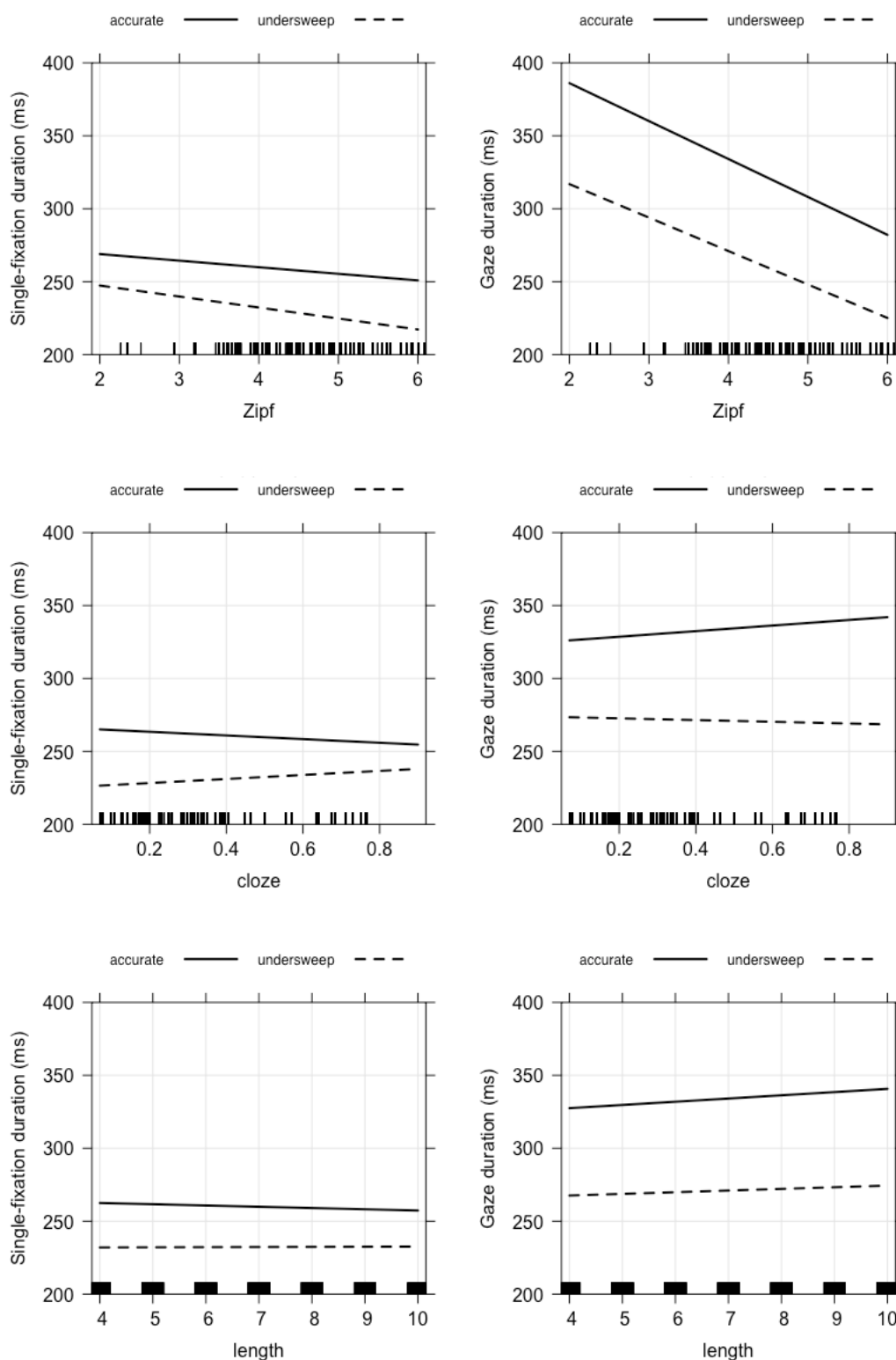
Figure 3. Fixed effects estimates from Bayesian Linear Mixed Effects Models for eye movement measures on line-initial words in the Provo Corpus. The solid line represents the estimate for cases in which line-initial words were fixated following a return-sweep. The dashed line represents cases in which the line-initial word was fixated following an undersweep-fixation.

Comparison of line-initial with line-internal words. In order to inform models that have so far been built for and fit to single line reading data, it will be necessary to compare the effects of lexical variables for line-internal words (those the models currently account for) with line-initial words. Therefore, we fit BLME models to single-fixation duration and gaze duration using word frequency, predictability and length as centered numerical predictors as in the prior analysis. However, this analysis was not restricted to only the line-initial words. Instead, we created a categorical variable to compare line-initial word reading following accurate return-sweeps with line-internal words (all other words except for line-initial and line final). Here line-internal words were coded 0 and line-initial words were coded 1. This meant that line-internal words represent the intercept to which line-initial words were compared. We exclude undersweep-fixation cases from this analysis for two reasons. The first is that we have evidence consistent with the notion that these cases allow for preview of line-initial words and we want to assess lexical effects in the absence of such preview. Second, the inclusion of the undersweep cases would also obscure the fixation time measures for the line-internal words they occur on.

Table 4 shows test statistics for the initial vs. internal word comparison BLME models for the Provo Corpus data. Examination of credibility intervals indicated that single-fixation durations were longer when words were longer, less frequent, and less predictable. The credibility intervals also indicated that single-fixations were longer on line-initial words than on line-internal words. However, there were no higher-level modulatory interactions.

Turning to gaze duration, there was again evidence that, in the Provo Corpus, all three lexical variables (frequency, predictability, and length) exerted an influence, as they had in the single-fixation measure. An effect of line-initial words was again observed, indicating that gaze durations were longer for line-initial words compared to line-internal words. These effects all agree with the single-fixation analysis. However, in gaze there was also evidence

of a higher order interaction. In the Provo Corpus, gaze durations on line-initial words were more influenced by word frequency than line-internal words were. For all other higher-level interactions, there was no modulatory effect on gaze duration.

Table 4. Provo: Bayesian linear mixed effect model results for line-initial and line-internal reading time measures.

Predictor	Single-fixation duration				Gaze duration			
			Percentiles				Percentiles	
	b	SD	2.5th	97.5th	b	SD	2.5th	97.5th
(Intercept)	**2.33**	**5.86e-3**	**2.31**	**2.34**	**2.37**	**7.20e-3**	**2.36**	**2.39**
Word frequency (WF)	**-7.35e-3**	**2.60e-3**	**-1.24e-2**	**-2.17e-3**	**-1.22e-2**	**2.96e-3**	**-1.78e-2**	**-6.33e-3**
Cloze probability (CP)	**-2.92e-2**	**1.06e-2**	**-5.04e-2**	**-8.01e-3**	**-3.40e-2**	**1.51e-2**	**-6.32e-2**	**-3.87e-3**
Word length (WL)	**4.10e-3**	**1.38e-3**	**1.44e-3**	**6.72e-3**	**1.24e-2**	**1.75e-3**	**9.00e-3**	**1.59e-2**
Line-initial words (LIW)	**2.54e-2**	**7.47e-3**	**1.12e-2**	**4.07e-2**	**5.15e-2**	**1.16e-2**	**2.91e-2**	**7.45e-2**
WF x CP	1.17e-2	1.05e-2	-8.48e-3	3.25e-2	1.87e-2	1.43e-2	-1.01e-2	4.52e-2
WF x WL	1.92e-4	9.32e-4	-1.66e-3	2.00e-3	-7.64e-4	1.15e-3	-2.97e-3	1.46e-2
CP x WL	8.95e-3	6.45e-3	-3.10e-3	2.16e-2	1.59e-2	7.87e-3	-1.05e-4	3.13e-2
WF x LIW	-4.86e-3	6.12e-3	-1.72e-2	6.95e-3	**-1.74e-2**	**8.08e-3**	**-3.23e-2**	**-1.15e-3**
CP x LIW	4.38e-2	2.66e-2	-8.92e-3	9.61e-2	5.46e-2	3.70e-2	-1.52e-2	1.28e-1
WL x LIW	-4.37e-3	3.27e-3	-1.06e-2	2.14e-3	-7.33e-3	4.21e-3	-1.54e-2	6.54e-4
WF x CP x WL	5.48e-3	4.44e-3	-3.06e-3	1.42e-2	5.58e-3	5.73e-3	-5.18e-3	1.70e-2
WF x CP x LIW	-2.02e-2	3.37e-2	-9.01e-2	4.65e-2	-2.81e-2	4.28e-2	-1.11e-1	5.76e-2
WF x WL x LIW	-4.54e-3	2.09e-3	-1.07e-2	1.65e-3	7.34e-4	3.67e-3	-6.60e-3	7.94e-3
CP x WL x LIW	-2.07e-2	2.02e-2	-5.94e-2	1.96e-2	-5.02e-2	2.64e-2	-1.01e-1	4.37e-3
WF x CP x WL x LIW	-3.88e-3	1.94e-2	-4.12e-2	3.44e-2	1.05e-2	2.46e-2	-3.80e-2	5.88e-2

Note. For single-fixation and gaze duration a negative coefficient of posterior mean *b* (in log milliseconds) indicates shorter reading times. The percentiles show 95% credible intervals computed from a Bayesian linear mixed effects models. **Bold** values indicate that, given our observed data, there is a 95% probability that the true value of |*b*| is greater than *0*.

Discussion: Eye Movement Corpus Analysis

We had two main goals in analyzing the eye-movement data following return-sweeps of the Provo Corpus. The first of these was to assess the extent to which undersweep-fixations (the short pauses between a return-sweep and a corrective saccade) may have enabled significant preview of line-initial words. Here the evidence was clear in indicating that single-fixation and gaze durations on line-initial words were shorter when preceded by an undersweep than when the return-sweep landed on these words without such intervening pauses. While there may be other explanations of this effect (e.g. the length of the incoming and outgoing saccade, contextual knowledge (by landing on a word ahead of its attention target readers may be able to use this information to facilitate their processing of the line-initial word), and a reduced cost of processing parafoveal information in cases following an undersweep-fixation when the line-initial word was fixated), we believe the most parsimonious is to assume that these undersweep-fixations provided readers with parafoveal preview of the line-initial words. This explanation, is also consistent with our predictions from our extended version of the E-Z Reader model. Furthermore, after statistically controlling for the lexical effects of frequency, length, and predictability, when the return-sweep landed accurately on the line-initial word, both single-fixation duration and gaze duration were longer on line-initial words than for line-internal words fixated during the normal left to right reading pass of the line. This finding is predicted by our modified E-Z Reader framework due to a lack of parafoveal preview for line-initial words following an accurate return-sweep.

Our other main goal in our analysis of the Provo Corpus was to investigate frequency and predictability effects for line-initial words following return-sweeps. While analysis of the data did not entirely coincide with our predictions, several interesting findings were observed

which have potential implications for computational models of eye movements during reading.

In the Provo Corpus, we found little evidence of a predictability effect for line-initial words in single fixation duration or gaze duration. This finding stands in contrast with previous experimental evidence showing that predictability influences gaze durations for line-initial words (Parker et al., 2017). These null findings are also inconsistent with our predictions for E-Z Reader. Based on this, one may reason that word predictability effects emerge from ambiguous orthographic information obtained parafoveally (Staub, 2015; Staub & Goddard, 2019). However, this conclusion cannot firmly be reached as the analysis included very few high-cloze line-initial words after filtering. It may be possible that we did not observe this effect because of the limited range of cloze values for line-initial words in the Provo Corpus. However, we did observe predictability effects for line-internal words which had a similar distribution of cloze values. Therefore, the null effect in the line-initial word analysis may simply be due to a lack of power in this more selective analysis. The analysis that compared line-initial words to line-internal words contained far more observations, and here predictability influenced first pass reading measures. Moreover, this predictability effect did not interact with the factor comparing line-initial to line-internal words.

Despite the equivocal predictability findings, an effect of word frequency was consistently observed on line-initial words in gaze duration which decreased as frequency increased. Moreover, the credibility intervals indicate that this frequency effect was larger for line-initial words than for line-internal words—a finding that would be inconsistent with our modified E-Z Reader framework. Note that the line-initial words in the Provo which entered our analysis were more than a character longer on average than the line-internal words. Others have reported interactions of word length and frequency in corpus analyses of eye-

movements during reading (Kliegl, Nuthmann, & Engbert, 2006). Therefore, the apparently stronger frequency effect for line-initial words compared to line-internal ones may be related to the uncontrolled nature of the lexical predictors.

The credibility intervals did not support a frequency effect in single-fixation duration in the analysis which was limited to line-initial words. However, in the analysis that included the comparison with line-internal words, there was support for a main effect of frequency on single-fixation durations and no evidence that this main effect differed between line-initial and line-internal words. Therefore, the single-fixation data provide an ambiguous picture of the influence of word frequency on line-initial words. One likely reason for this is the large variability in word length for line-initial words within the analyzed portion of the Provo Corpus (range: 4-12). As word length increases, words are more likely to be refixated (less likely to receive a single fixation) in first pass reading. This relationship reduces the number of observations in single-fixation analyses relative to gaze duration.

As an interim summary, we found consistent evidence that undersweep-fixations provide an opportunity for lexical processing to take place for line-initial words prior to direct fixation. Additionally, we found consistent evidence that single-fixation durations and gaze durations are longer on line-initial words following accurate return-sweeps to them then for line-internal words. We interpret this as the result of a lack of preview for line-initial words following accurate return-sweeps. The evidence for frequency effects in gaze durations indicated that these effects on line-initial words are at least as large as those on line-internal words. However, there remain questions about frequency effects in single fixation duration and the relative size of the effects in gaze duration. These questions require additional research with line-initial words controlled on word length. Moreover, in contrast to Parker et al. (2017), we found no effect of predictability on line-initial words in any of our fixation duration measures. This null pattern is consistent with Staub and Goddard's (2019)

predictions that when words are unavailable for processing in the parafovea, only frequency effects should remain. Rather than conclude that predictability effects are absent for line-initial words, we consider the possibility that these null findings result from a lack of highly predictable line-initial words entering our statistical analyses. To address these limitations of the corpus analysis we report an eye movement experiment in which we manipulated target word frequency and predictability while controlling word length to provide a stronger test of our predictions.

Eye Movement Experiment

In order to examine our predictions in a more controlled manner, we present a novel eye movement reading study in which word frequency and predictability of line-initial words were experimentally manipulated. In conducting this experiment, we were able to examine the effect that a full range of cloze-probabilities would have on the processing of line-initial words.

Method

Participants. Forty-Eight English monolingual members of the Bournemouth University community (12 males; mean age 20.5 years, SD= 3.20 years) participated in return for course credit. The sample was selected to be comparable to other studies that have investigated predictability effects for line-initial words (Parker et al., 2017). Based on Parker et al.'s analysis of predictability effects for line-initial words, we estimated Cohen's d to be .21 using Westfall, Judd, and Kenny's (2014) formula for effect size calculation (see Brysbaert & Stevens, 2018, for discussion). Following Westfall et al.'s power analysis for linear mixed effects models[5], it was approximated that 16 participants would yield sufficient power (β= .808) to test predictability effects. Given that we were also interested in testing the interaction between predictability, frequency and undersweep-fixation likelihood, we included 48 total participants in the current study for eight complete counterbalanced groups.

From the final data set we used the "simR" package (version 1.0.4; Green & MacLeod, 2016) to estimate the power of linear mixed effects models to detect predictability effects in log-transformed gaze duration on the basis of 200 random samples. The simulation indicated that β= 93.5% [89.1, 96.5]. In addition, a recently published study by Johnson, Oehrlein, and Roche (2018), investigated the role of parafoveal preview and predictability effects in 48 children reading 60 items. This provided sufficient power for Johnson et al. to detect predictability effects in single-fixation and gaze durations (the measures of interest in the current work) in addition to individual differences. Participants were naïve to the purpose of the experiment, had normal or corrected-to-normal vision, and no self-reported history of reading impairment.

Apparatus. Eye movements were recorded using an SR Research EyeLink 1000 system with a sample rate of 1000 Hz. While viewing was binocular, data was recorded for the right eye only. Black text was displayed on a white background on a BenQ XL2410T LCD monitor with a 1900 × 1080 pixel resolution. Viewing distance was 80 cm, with 1° of visual angle being occupied by 3.76 characters of monospaced Consolas font.

Materials. Stimuli consisted of 64 passages of text adapted from Rayner et al. (2004) and Hand et al. (2010). Each adapted item consisted of one-to-three sentences displayed across two-to-three lines. Target words were presented as the line-initial word on the final line of text. Each passage was designed to accommodate a high- and low-frequency target. Context was varied so that for half of the passages the high-frequency target was high-cloze probability and the low-frequency target was low-cloze probability. For the other half, context was varied so that high-frequency target was low-cloze probability and the low-frequency target was high-cloze probability (see Figure 4). This led to a 2 (frequency) by 2 (predictability) design. Participants viewed 16 items per condition with items being divided into two sets and counterbalanced over participants. A paired-sample *t*-test indicated that the

distance between the final word in the n-1 line and the line-initial target word did not differ

between sentence frames in targets were embedded, $t(31)= .51$, $p= .611$.

`Jamming all my laundry into the washing machine, I`

`simply ignored the fact that it could potentially`

`break` `because I had overloaded its capacity.` *(HP, HF)*

`Jamming all my laundry into the washing machine, I`

`simply ignored the fact that it could potentially`

`erupt` `because I had overloaded its capacity.` *(LP, LF)*

`The geologists hurried away from the volcano.`

`Their measurements suggested that it could`

`break` `at any moment.` *(LP, HF)*

`The geologists hurried away from the volcano.`

`Their measurements suggested that it could`

`erupt` `at any moment.` *(HP, LF)*

Note: HP= high—predictability, LP= low—predictability, HF= high—frequency, LF= low—frequency.

Figure 4. Example stimuli with the target words ***break*** and ***erupt*** shown in bold. Text in the experiment was double spaced.

The targets varied from 5 to 8 letters (mean= 5.7, *SD*= .97). Frequencies were based on the Zipf scale (van Heuven et al., 2014). Target words differed significantly in their Zipf frequency (high frequency: mean= 5.1, *SD*= 0.38; low frequency: mean= 3.54, *SD*= .0.45),

t(31)= 14.23, d= 3.65, p<.001, and were non-overlapping between conditions. A Cloze

norming study (n= 48) confirmed the appropriateness of our stimuli for the current study (see

Table 5). A repeated-measures ANOVA, with frequency (2 levels) and predictability (2

levels) as factors, revealed cloze accuracy was higher in the predictable condition, F(1,35)=

899.79, MSE= 12.26, ηp^2= .97, p<.001). There was no main effect of word frequency on

cloze accuracy, nor was there an interaction between these two factors (both Fs < 1).

Table 5. Cloze probabilities per experimental condition.

Experimental condition		
Predictability	Frequency	Cloze probability
Predictable	High	.64
	Low	.63
Unpredictable	High	.02
	Low	.01

Procedure. After providing informed consent, participants were seated in front of the

monitor and a head rest was used for stability. A 9-point calibration grid was used, with an

acceptance criterion of an average error below 0.4 degrees. At the start of each trial,

participants had to fixate on a black square (2° by 2°) which triggered stimuli presentation

once a stable fixation was detected. The 64 experimental stimuli were presented in random

order intermixed with 60 stimuli from a separate experiment and participants were instructed

to read silently for comprehension. True or False questions appeared after a third of trials and

required participants to respond by pressing one of two buttons on a VPixx response box.

Participants answered 88% of comprehension questions correctly.

Results

Trials with track loss and trials that contained a blink on the target word were

removed, as were trials in which there were five or more blinks. In total, 6.25% of trials were

removed. Short fixations (<80 ms) within a character of a previous or subsequent fixation

were combined with that fixation. Other fixations less than 80 ms in duration were removed

(0.56% of fixations). A logistic regression fit to individual fixations indicated that fewer line-initial fixations were shorter than 80 ms relative to non-line-initial fixations, *b*= -1.16, *SE*= 3.21e-2, *z*= -3.62. Again, we constructed BLME models for eye movement data using the stan_lmer function from the rstanarm package in R. The measures examined in the previous eye movement study were analyzed. For reference, single-fixations occurred on 77.5% of the analyzed cases (13.9% more frequently than in the Provo Corpus). For all models, we adopted a full random structure, treating participants and items as random factors, with random intercepts and slopes. All fixed effects were centered.

Return-sweep landing position. As shown in Figure 5, the majority of return-sweeps were launched approximately 8 characters from the end of the line, with the majority of return-sweeps landing approximately 5 characters or less from the start of the next line. To examine how the lexical characteristics of the line-initial word influenced return-sweep targeting, we again fit a BLME model to landing position data. The model included the experimentally manipulated variables word frequency, predictability, and their interaction in the fixed effects structure (see Table 6). As with our analysis of the Provo Corpus, there was no evidence that the lexical characteristics of the line-initial words influenced the landing position of return-sweeps.
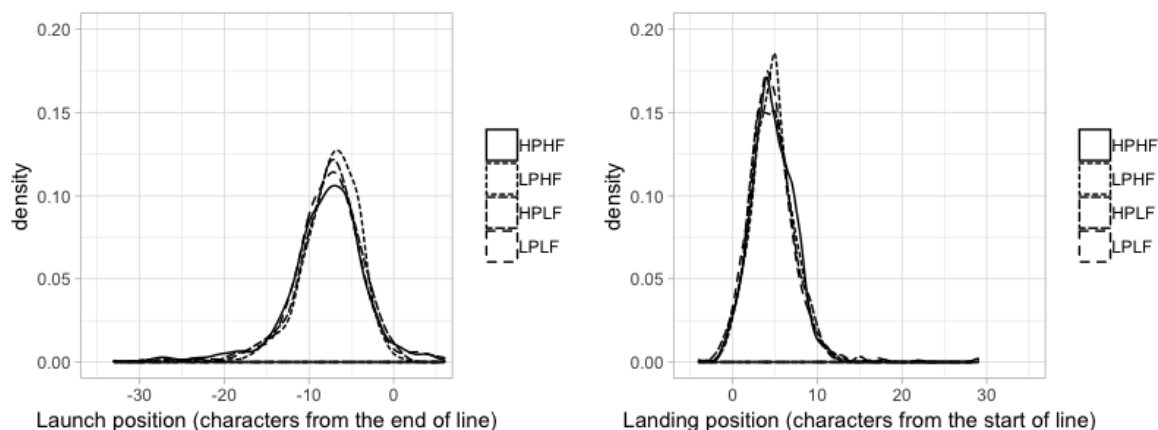
Figure 5. Experiment: Density plot for return-sweep launch (left panel) position and landing position (right panel). For return-sweep launch position, zero represents the right margin. Return-sweeps launched from a negative value occurred prior to the end of a line. For return-sweep landing position, zero represents the left margin. As landing position increased, return-sweeps were made further into the line. Average line length: 60 characters.

Table 6. Experiment: Bayesian linear mixed effects model results.

| | Return-sweep landing position | | | |
| | | | *Percentiles* | |
Predictor	*b*	*SD*	2.5th	97.5th
(Intercept)	**4.60**	**2.30e-1**	**4.16**	**5.08**
Word frequency (WF)	-3.51e-3	7.47e-2	-1.41e-1	1.43e-1
Cloze probability (CP)	3.07e-1	1.85e-1	-5.53e-2	6.56e-1
WF x CP	2.25e-1	2.86e-1	-3.32e-1	7.89e-1

Note. A negative coefficient of posterior mean *b* (in character spaces) indicates landing positions further to the right. The percentiles show 95% credible intervals computed from a Bayesian linear mixed effects models. ***Bold*** values indicate that, given our observed data, there is a 95% probability that the true value of *|b|* is greater than *0*.

First-pass reading time and undersweep-fixations. Pirate plots showing measures of reading time were constructed using the 'yarrr' package (version 0.1.5; Phillips, 2017) and are shown in Figure 6. As with our analysis of the Provo Corpus, appropriate first-time reading measures for line-initial words in the cases of undersweep-fixations, we calculated the reading times as if the undersweep-fixation had not occurred. That is, undersweep-fixations did not end the first pass reading time of the words to their left. Prior to analysis, measures were log transformed for normality. Each model included fixed effects for word frequency and predictability. An additional categorical fixed effect was included to code

whether participants had made an undersweep-fixation on a given line. For single-fixation and gaze duration data, undersweep-fixations intervened return-sweeps and fixations on line-initial words on 36.1% and 29.6% of cases respectively. All higher-level interactions were included between word frequency, predictability, and undersweep-fixation.
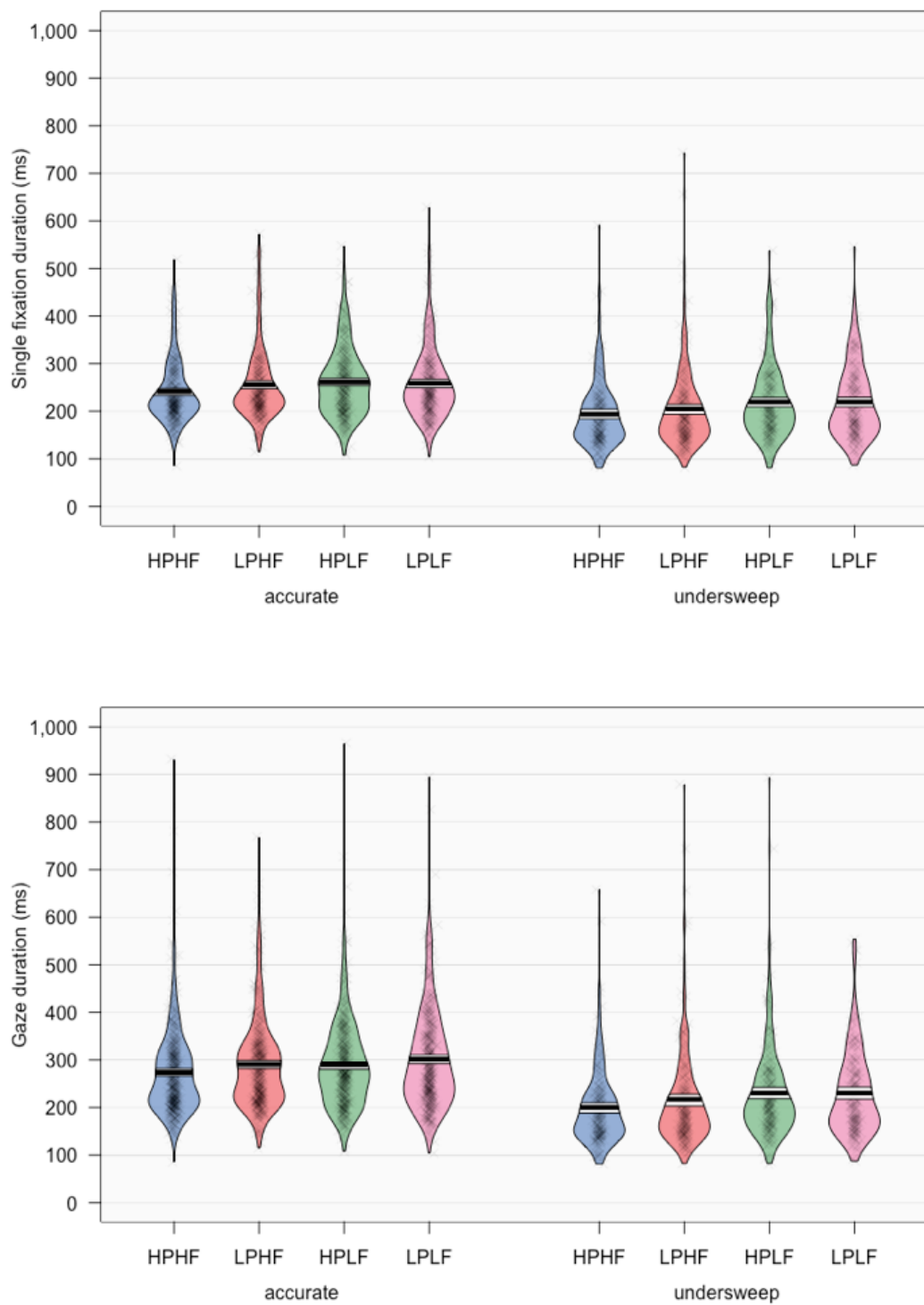
Figure 6. Pirate plots for reading time measures. Shown are the means per condition. Around the means are 95% Highest Density Intervals (HDIs). The distribution is then shown for each condition with the raw data behind.

Test statistics are shown in Table 7. Unlike the analysis of the Provo Corpus, the analysis of single-fixation and gaze duration showed the same pattern of effects in our controlled experiment. Similar to our Provo Corpus analysis, the durations on line-initial words were shorter following an undersweep. The credibility intervals also indicated main effects of frequency and predictability. As these variables increased, line-initial word fixation durations decreased. However, inspection of the credibility intervals indicated that there were no influences of higher-level interactions on these durations.

Table 7. Experiment: Bayesian linear mixed effect model results for reading time measures.

Predictor	Single-fixation duration				Gaze duration			
			Percentiles				*Percentiles*	
	b	*SD*	*2.5*th	*97.5*th	*b*	*SD*	*2.5*th	*97.5*th
(Intercept)	**2.38**	**9.80e-3**	**2.36**	**2.40**	**2.43**	**1.11e-2**	**2.40**	**2.43**
Word frequency (WF)	**-1.10e-2**	**4.62e-3**	**-2.00e-2**	**-1.84e-3**	**-1.32e-2**	**4.40e-3**	**-2.19e-2**	**-4.62e-3**
Cloze probability (CP)	**-2.54e-2**	**1.09e-2**	**-4.69e-2**	**-430e-3**	**-3.66e-2**	**1.02e-2**	**-5.66e-2**	**-1.64e-2**
Undersweep-fixation (UF)	**-1.10e-1**	**8.24e-3**	**-1.26e-1**	**-9.35e-2**	**-1.44e-1**	**8.02e-3**	**-1.60e-1**	**-1.28e-1**
WF x CP	-1.15e-2	1.49e-2	-4.04e-2	1.78e-2	4.09e-3	1.62e-2	-2.76e-2	3.59e-2
WF x UF	-1.43e-2	7.75e-3	-2.95e-2	9.79e-4	-1.31e-2	8.79e-3	-3.03e-2	4.21e-3
CP x UF	3.11e-3	1.65e-2	-2.95e-2	3.53e-2	7.49e-3	1.78e-2	-2.74e-2	4.29e-2
WF x CP x UF	-1.79e-2	2.20e-2	-6.13e-2	2.52e-2	-4.27e-2	2.33e-2	-8.84e-2	2.88e-3

Note. A negative coefficient of posterior mean *b* (in log milliseconds) indicates shorter reading times. The percentiles show 95% credible intervals computed from a Bayesian linear mixed effects models. ***Bold*** values indicate that, given our observed data, there is a 95% probability that the true value of |*b*| is greater than *0*.

Comparison of line-initial with line-internal words. We wanted to assess whether fixation times were longer on line-initial words than they are on line-internal words as they had been in the Provo Corpus. Therefore, we fit a BLME model to single-fixation and gaze duration data using word frequency as a centered numerical predictor, a categorical variable to compare line-initial word reading following accurate return-sweeps with line-internal words (all other words except for line-initial and line final), and their interaction. As before, we excluded undersweep-fixation cases from this analysis. We applied the following constraints to observations entering analysis: line-initial target words were only those that were highly predictable, midline targets were then matched with line-initial words in terms of length and frequency. By analyzing only predictable line-initial target words, we removed the likelihood that longer single-fixation and gaze durations on line-initial words were the result of them being unpredictable. Controlling for word length and frequency between midline and line-initial words meant that comparisons were made on two equal N sets of words that were identical in terms of their length and frequency. The Zipf frequency between line-initial targets and matched midline words did not differ, $t(63)= -.34$, $d= .01$, $p= .736$.

Table 8 shows test statistics for the initial vs. internal word comparison BLME models for the experimental data. Examination of credibility intervals indicated that single-fixation durations were longer when words were less frequent. The credibility intervals also indicated that single-fixations were longer on line-initial words than on line-internal words. However, no higher-level interactions, had modulatory effects.

Turning to gaze duration, there was again evidence that, like in our comparison of line-initial with line-internal words in the Provo Corpus, lexical frequency exerted an influence, as they had in the single-fixation measure. A main effect of line-initial words was again observed, indicating that gaze durations were longer for line-initial words compared to

line-internal words. These effects all agree with the single-fixation analysis. For all higher-

level interactions, there was no modulatory effect on gaze duration.

Table 8. Experiment: Bayesian linear mixed effect model results for line-initial and matched line-internal reading time measures.

Predictor	Single-fixation duration				Gaze duration			
			Percentiles				*Percentiles*	
	b	*SD*	2.5^{th}	97.5^{th}	b	*SD*	2.5^{th}	97.5^{th}
(Intercept)	2.31	7.36e-3	2.29	2.32	2.35	8.90e-3	2.33	2.37
Word frequency (WF)	**-4.68e-3**	**4.70e-3**	**-1.37e-2**	**-1.50e-3**	**-1.72e-2**	**6.01e-3**	**-2.92e-2**	**-5.49e-3**
Line-initial words (LIW)	**7.41e-2**	**8.29e-3**	**5.75e-2**	**8.99e-2**	**7.95e-2**	**8.99e-3**	**6.24e-2**	**9.66e-2**
WF x LIW	-6.24e-3	7.92e-3	-2.15e-2	9.30e-3	4.90e-3	8.57e-3	-1.16e-2	2.17e-2

Note. For single-fixation and gaze duration a negative coefficient of posterior mean b (in log milliseconds) indicates shorter reading times. The percentiles show 95% credible intervals computed from a Bayesian linear mixed effects models. ***Bold*** values indicate that, given our observed data, there is a 95% probability that the true value of $|b|$ is greater than *0*.

Discussion: Eye Movement Experiment

Our eye movement experiment was designed to address specific limitations of our analysis of the Provo Corpus, thereby better informing models of eye movement control during reading. The findings from this experiment are consistent across both gaze and single-fixation duration. We found evidence that these durations increase with decreasing frequency and predictability. Moreover, the means of these measures across the experimental conditions are in line with those reported by Rayner et al. (2004) and Hand et al. (2010) which our items were closely adapted from (see Table 9). The close adaption of our items from these studies comes with at least two important benefits. The first is that we know the size of the effects reported by the studies, which used these items in line central locations. This offers us a good comparison for the line-initial word location in our adapted items. The second benefit is that the findings from Rayner et al. (2004) directly influenced the way in which frequency and predictability influence word processing in the current version of E-Z Reader. Therefore, we can be confident that E-Z Reader would reproduce the effects in single-fixation and gaze duration.

Table 9. Mean single-fixation duration (SFD) and gaze duration (GD) measure comparisons

Study	Current (Undersweep)		Current (Accurate)		Hand et al. (2010)		Rayner et al. (2004)	
Condition	SFD	GD	SFD	GD	SFD	GD	SFD	GD
HPHF	194	200	242	274	259	273	258	274
LPHF	205	217	256	292	269	286	268	290
HPLF	218	229	262	292	285	306	269	288
LPLF	220	231	259	304	294	318	287	310
Predictability	-7	-10	-6	-15	-10	-13	-14	-19
Frequency	-19	-22	-12	-15	-26	-33	-15	-17

Note. In the current experiment, single-fixations occurred on 70.2% of trials in which the line-initial word was accurately fixated and 94.9% of trials in which the line-initial word was fixated following an undersweep-fixation. Single-fixations accounted for 62.8% of observations in Hand et al. (2010). The percentage of trials receiving a single-fixation was not reported in Rayner et al. (2004).

The data from this experiment also help to inform on the cause of the word predictability by word preview interactions reported in boundary change studies. Crucially, single-fixations durations were longer following accurate return-sweeps despite the lack of preview available for these words and this effect did not interact with undersweep-fixations. So, with or without potential parafoveal preview, we found evidence of a predictability effect. That this effect is present in single-fixation duration is important because these represent cases where the line-initial word was efficiently processed without needing to resolve any orthographic ambiguity via refixations. This is difficult data for Staub and Goodard's explanation of predictability effects arising from orthographic ambiguity gleaned during parafoveal processing. It is however, in line with our assertion that predictability effects arise from incrementally updating multiple hypothesis of the underlying meaning of a sentence (Levy, 2008; Parker et al. 2017).

The most consistent finding from the Provo Corpus analysis was that undersweep-fixations appear to provide an opportunity to gain preview of line-initial words prior to their direct fixation. This effect was again found in our controlled experiment in both single-fixation duration and gaze duration. This provides corroborating support for these potential preview benefits effects following undersweep-fixations. Within our modified E-Z Reader framework, this finding follows logically from our assumption that lexical processing of the first word on a new line not begin until that line is fixated so that the word will fall within the fovea or parafovea—which happens during undersweep-fixations. However, this interpretation should be verified with the use of a boundary change experiment in which the boundary is placed to the right of line initial words.

General Discussion

Our primary goal was to investigate the time course of word frequency and predictability effects for line-initial words to provide benchmark return-sweep findings for

models of eye movement control during reading. We focused only on lexical processing following a return-sweep and derived predictions based on a modified E-Z Reader framework. Our modified framework was based on a single simplifying assumption, that no lexical processing of a line-initial word would occur prior to a fixation on the line that brought placed the initial word within the fovea or parafovea. Based on this framework, we predicted that word frequency and predictability effects would be present in single-fixation duration and gaze duration.

Across studies, there was consistent evidence to suggest that word frequency influenced the gaze durations on line initial words. Within the Provo Corpus, this effect of frequency on line-initial words was larger than it was on line-internal words. However, in our controlled experiment, the effect of frequency was statistically similar for line-initial and line-internal words. The gaze duration difference between the high and low frequency words in our experiment (22ms) was close to the differences obtained by Rayner et al. (2004: 17ms), and Hand et al. (2010: 33ms) who used the same target words embedded at line-internal positions in very similar items.

In our eye movement experiment, predictability effects were present in single-fixation and gaze duration, indicating that predictability effects for line-initial words are emerge at an early stage of processing. This contrasts Parker et al. (2017) which reported no effect of predictability in the earliest measure of processing (first-fixation duration). We argue that these differences are the result of the measures chosen between studies. As mentioned previously, single-fixation represent those cases where readers are able to process a word without having to refixate towards a more optimal viewing location. First-fixation, however, includes cases in which readers may have landed non-optimally on a line-initial word and required a refixation. As a result, these fixations may have been terminated at a point to afford readers with more optimal processing before the point at which lexical influences on

fixation occur. We attribute the failure to find predictability effects in the Provo Corpus to a lack of power when considering line-initial words. The is likely given the observation that predictability effects did not differ between line-initial and line-internal words.

Staub and Goddard (2019) have reported that when preview is denied, frequency effects remain while predictability effects vanish. This differs from our experimental findings for single-fixation duration for accurate line-initial fixations. One crucial difference between these studies is how preview was denied. Staub and Goddard, used the boundary technique (Rayner, 1975) meaning that when valid previews were denied, invalid previews were present. In our experiment, preview was denied because the target word was located far in the periphery on the prior fixation. That is, there was no invalid preview used in our experiment. Finding a predictability effect in single-fixation durations following accurate return-sweeps is therefore consistent with our account that readers maintain multiple hypotheses about the underlying meaning of a sentence and update the strength of their belief in each hypothesis as new bottom-up information becomes available (see Levy, 2008). This potentially explains why the predictability effect disappears when using invalid previews in a boundary change experiment—these invalid previews would be inconsistent with the predictable sentence meaning. This position is bolstered by recent research suggesting that an invalid parafoveal preview in a boundary change study can be processed to the point of being integrated into the sentence (Schotter, von der Malsburg, & Leinenger, 2019) It also explains why predictability effects remain for line-initial words following an accurate return-sweep (Parker et al. 2017). Since there was no invalid preview, whatever expectation had been built during reading of the prior words would remain. However, given the inconsistent findings in the Provo Corpus, it may be that our explanation of predictability effects may only apply when words are highly predictable.

Our second goal was to investigate the effect an undersweep would have on later processing of line-initial words. Across both studies, there was a clear consistent reduction in gaze durations following for line-initial words following an undersweep-fixation. This benefit was also observed for single-fixation durations in our experimental data. As undersweep-fixations are viewed as oculomotor error (Hofmesiter et al., 1999; Parker et al., 2017), it is assumed that these fixations are uninvolved in on-going linguistic processing (Hawley et al., 1974; Shebilske, 1975). However, we provide evidence to suggest that these fixations enable the encoding of bottom-up visual input when return-sweeps fail to reach their target. This is likely to occur as a result of attention being only loosely coupled with eye movements. When an undersweep-fixation lands on word *n+1,* it is likely that attention will be on or will rapidly shift to word *n*, and lexical processing of word *n* will begin prior to direct fixation[6]. This same argument is made for the existence of mislocated fixations (Drieghe, Rayner, & Pollatsek, 2008) based on the decoupling of attention and fixation location due to oculomotor error. Indeed, E-Z Reader allows for fixations and attention to be located on different words due to oculomotor error, which is consistent with the current argument that undersweep-fixations provide preview benefit of line-initial words.

Of course, these interpretations of reduced fixation durations on line-initial words following an undersweep-fixation are not the only interpretation of the data. Other interpretations include the length of the incoming and outgoing saccade, contextual knowledge (by landing on a word ahead of its attention target readers may be able to use this information to facilitate their processing of the line-initial word), and a reduced cost of processing parafoveal information in cases following an undersweep-fixation when the line-initial word was fixated. To validate our interpretation of these results, that undersweep-fixations provide preview for line-initial words, future work utilizing the boundary paradigm is needed.

Implications for models of eye movement control

As they are currently implemented, models of eye movement control during reading can only simulate reading single lines of text. During single line reading E-Z Reader uses word frequency and predictability as language input variables, and predicts strong additive effects of both frequency and predictability with fixation times decreasing as these variables increase. With single line reading, E-Z Reader gains access to frequency and predictability information about a word for lexical processing at the same point in time—when serial attention is shifted to the word. However, with multi-line reading we made the additional assumption that the shift of attention to the next line's initial word would not start that word's lexical processing. Instead, lexical processing of this word would need to wait until the return-sweep to that line was completed so that the word would fall into foveal or parafoveal vision. With this added assumption, we predicted the following:

1. The duration of the line-initial fixation following an accurate return-sweep (i.e. no intervening undersweep-fixation) should be longer compared to words fixated during the normal left to right (for English) reading pass within a line.

2. That fixation times on line-initial words would be reduced if preceded by an undersweep-fixation due to the potential availability of preview benefit provided by these fixations.

3. That the effects of word frequency and predictability would remain the same as for other words. As already mentioned this may not be the case.

In our analysis of the Provo Corpus, we found evidence for the first two of these predictions with mixed results for the third prediction. Specifically, we found evidence that word frequency influences gaze duration on line initial words. However, the effect of frequency was actually stronger for line-initial words than for line-internal words. Moreover, while we found evidence that predictability influenced gaze duration which did not differ between line-

initial and line-internal words, in the analysis of only line-initial words there wasn't strong evidence of an influence of predictability.

In our experiment, we found evidence to support all three of these predictions. Line-initial fixation durations following accurate return-sweeps were longer than fixations during the normal left to right reading pass of a line. Single-fixations and gaze durations on line initial words were also shorter if they had been proceeded by an undersweep-fixation. This finding is consistent with our hypothesis that undersweep-fixations provide preview of line-initial words. Finally, the reported effects of word frequency and predictability for line-initial words was similar to those of line-internal words consistent with E-Z Reader.

E-Z Reader may be able to explain the current study's findings related to the processing of line-initial words following return-sweeps with only one modifying assumption—that lexical processing (L1 & L2) of words on a line not occur until that line had been fixated. However, there are other potential explanations of findings in the current study. As outlined by Parker et al. (2017), if the link between lexical processing and saccade planning were moved to a later point for accurate line initial fixations within E-Z Reader, it may be able to simulate longer line-initial fixations. However, this would predict differential effects for frequency and predictability at the very start of the line which would be inconsistent with our observations form the current eye movement experiment.

Conclusions

Two investigations (one corpus, one experimental) present evidence that fixation times on line-initial words are influenced by frequency and predictability in a manner similar to line-internal words despite being unavailable for parafoveal pre-processing on the fixation prior to the return-sweep. We also present data consistent with the notion that undersweep-fixations provide the opportunity to gain bottom-up visual information about line-initial words in a similar manner to parafoveal preview of words during normal left-to-right reading

within a line of text. These findings are the first benchmark findings on lexical processing following return-sweeps, and will be essential in future modelling of multi-line text reading. The findings related to lexical processing of line-initial words are also consistent in principle with our modified E-Z Reader framework. However, it remains to be seen if E-Z Reader or other models of eye movements during reading can simulate these effects while also capturing other aspects of return-sweep saccades during reading such as launch and landing sites.

It has been 20 years since the initial E-Z Reader model was published. The model has been updated 10 times (Reichle, 2011) but is still only capable of reading single lines of text. The current study is the first to provide crucial information about lexical processing following return-sweeps which is necessary for E-Z Reader (and other models of eye movements during reading) to embark on the modelling of multi-line text reading.

Notes

1. Word frequency is estimated as a word's number of occurrences per million words (e.g. van Heuvan, Mandera, Keuleers, & Brysbaert, 2014) and contextual predictability is estimated as its Cloze probability as tabulated by the mean proportion of subjects that correctly guess a word from its preceding context (Taylor, 1953).

2. According to our account, predictability effects should remain for line initial words that received no parafoveal preprocessing while the Staub and Goddard account predicts an absence of a predictability effects. Prior research on predictability effects for such line initial words was equivocal on this matter. While Parker et al. (2017) found the effect in gaze duration, it was absent in first fixation duration. In our account, the null predictability effect for line-initial words in first fixation duration would be explained as either a type 2 error while according to Staub and Goddard, this could be due to orthographic ambiguity resulting from non-optimal landing positions. That is, the bottom-up input for first fixation durations is ambiguous enough to act similar to a parafoveal fixation.

3. Luke and Christianson estimated predictability with a Cloze task in which 470 participants provided cloze estimates for each word in a passage, with 40 participants providing a response for each passage.

4. Text was presented in proportional font. Character spaces estimated to be 15 pixels each.

5. Westfall et al. (2014) published a theoretical analysis of mixed effects models and made a website (https://jakewestfall.shinyapps.io/two_factor_power/) which allows researchers to calculate the power of an experiment and the number of items/participants required for a well-powered experiment. We estimated the sample size using the following values: mean difference: .03, contrast code: .5, residual variance: .0122, participant intercept variance: .0079, item intercept variance: .0007; participant-by-target variance: 0; participant slope variance: .0011; item slope variance: .0004.

6. Slattery and Parker (2019) reported of inhibition of return effects following undersweep-fixations suggesting that attention is at least initially located on the word receiving the undersweep-fixation. However, these inhibition of return effects were significantly weaker following undersweeps than with intraline reading fixations.

Acknowledgments

Declarations of interest

None.

References

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

Altarriba, J., J. F. Kroll, A. Sholl, & K. Rayner. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: evidence from eye fixations and naming times. *Memory & Cognition* 24. 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. https://doi.org/10.1037/a0036396

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

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17*, 364–390. http://dx.doi.org/10.1016/0010- 0285(85)90013-1

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Keep it maximal Appendix. *Journal of Memory and Language*, *68*(3), 1–5. https://doi.org/10.1016/j.jml.2012.11.001

Beymer, D., Russell, D. M., & Orton, P. Z. (2005). Wide vs. narrow paragraphs: an eye tracking analysis. In *IFIP Conference on Human-Computer Interaction* (pp. 741-752). Springer, Berlin, Heidelberg.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word

frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–

990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models:

a tutorial. *Journal of Cognition, 1*(1), 1-20.

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of

eye movement control in reading. In G. Underwood (Ed.), Cognitive Processes in Eye

Guidance. Oxford: University Press, pp. 53-77.

Choi, W., & Gordon, P. C. (2013). Coordination of word recognition and oculomotor control

during reading: The role of implicit lexical decisions. *Journal of Experimental

Psychology: Human Perception and Performance, 39*, 1032–1046.

http://dx.doi.org/10.1037/a0030432

Choi, W., Lowder, M. W., Ferreira, F., Swaab, T. Y., & Henderson, J. M. (2017). Effects of

word predictability and preview lexicality on eye movements during reading: A

comparison between young and older adults. *Psychology and Aging, 32*, 232–242.

http://dx.doi.org/10.1037/ pag0000160

Drieghe, D., Rayner, K., & Pollatsek, A. (2007). Mislocated fixations can account for

parafoveal-on-foveal effects in eye movements during reading. *Quarterly Journal of

Experimental Psychology*, *61*(8), 1239–1249.

https://doi.org/10.1080/17470210701467953

Ehrlich, S. F., & K. Rayner. 1981. Contextual effects on word perception and eye movements

during reading. *Journal of verbal learning and verbal behaviour*, *20*, 641–655.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (n.d.). SWIFT: A Dynamical Model

of Saccade Generation During Reading. https://doi.org/10.1037/0033-295X.112.4.777

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel

processing and competitive inhibition. *Behavioral and Brain Sciences*, *22*(4), 661-

674.

Gollan, T. H., T. J. Slattery, D. Goldenberg, E. Van Assche, W. Duyck, & K.

Rayner. 2011. Frequency drives lexical access in reading but not in speaking: the

frequency-lag hypothesis. *Journal of Experimental Psychology: General, 140*, 189–

209.

Goodrich B, Gabry J, Ali I & Brilleman S. (2018). rstanarm: Bayesian applied regression

modeling via Stan. R package version 2.17.4. http://mc-stan.org/.

Green P, MacLeod CJ (2016). "simR: an R package for power analysis of generalised linear

mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493-498. doi:

10.1111/2041210X.12504

Hand, C. J., Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2010). The frequency-

predictability interaction in reading: It depends where you're coming from. *Journal of

Experimental Psychology: Human Perception and Performance*, *36*(5), 1294–1313.

https://doi.org/10.1037/a0020363

Hawley, T. T., Stern, J. A., & Chen, S. C. (1974). Computer analysis of eye movements

during reading. *Literacy Research and Instruction, 13*, 307-317.

Heller, D. (1982). Eye movements in reading. In R. Groner, & P. Fraisse (Eds.), *Cognition

and Eye Movements* (pp. 487-498). Berlin, Deutscher Verlag der Wissenschaften.

Hofmeister, J., Heller, D., & Radach, R. (1999). The return sweep in reading. In W. Becker,

H. Deubel, & T. Mergner (Eds.), *Current Oculomotor Research* (pp. 349-357). United

States: Springer.

Hohenstein, S., Matuschek, H., & Kliegl, R. (2017). Linked linear mixed models: A joint analysis of fixation locations and fixation durations in natural reading. *Psychonomic bulletin & review*, *24*(3), 637-651. DOI: 10.3758/s13423-016-1138-y

Hutzler, F., Fuchs, I., Gagl, B., Schuster, S., Richlan, F., Braun, M., & Hawelka, S. (2013). Parafoveal X-masks interfere with foveal word recognition: evidence from fixation-related brain potentials. *Frontiers in Systems Neuroscience*, *7*(July), 1–10. https://doi.org/10.3389/fnsys.2013.00033

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

Johnson, R. L., Oehrlein, E. C., & Roche, W. L. (2018). Predictability and parafoveal preview effects in the developing reader: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 44*(7), 973.

Juhasz, B. J., White, S. J., Liversedge, S. P., & Rayner, K. (2008). Eye movements and the use of parafoveal word length information in reading. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1560–1579. http://dx.doi.org/10.1037/a0012319

Just, M. A., & C. P. A. (1980). A Theory of Reading: From Eye Fixations to Comprehension. *American Psychologist*, *87*(4), 713–713. https://doi.org/10.1037/h0021465

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

Kennison, S. M., & Clifton, C., Jr. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 68–81. http://dx.doi .org/10.1037/0278-7393.21.1.68

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.

   https://doi.org/10.1080/09541440340000213

Kliegl, R., Hohenstein, S., Yan, M., & McDonald, S. A. (2013). How preview space/time

   translates into preview cost/benefit for fixation durations during reading. *Quarterly*

   Journal of Experimental Psychology, *66*(3), 581–600.

   https://doi.org/10.1080/17470218.2012.658073

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.

Laubrock, J., Kliegl, R., & Engbert, R. (2006). SWIFT explorations of age differences in eye

   movements during reading. *Neuroscience and Biobehavioral Reviews*, *30*(6), 872–

   884. https://doi.org/10.1016/j.neubiorev.2006.06.013

Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr . Chips : An Ideal-Observer Model of

   Reading, *104*(3), 524–553.

Liversedge, S., & Findlay, J. (2000). Saccadic eye movements and cognition. *Trends in*

   Cognitive Sciences, *4*(1), 6–14. https://doi.org/http://dx.doi.org/10.1016/S1364-

   6613(99)01418-7

Luke, S. G., & Christianson, K. (2018). The Provo Corpus: A large eye-tracking corpus with

   predictability norms. *Behavior Research Methods*, *50*(2), 826–833.

   https://doi.org/10.3758/s13428-017-0908-4

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.

https://doi.org/10.1080/20445911.2015.1008494

McDonald, S. A. (2006). Parafoveal preview benefit in reading is only obtained from the

saccade goal. *Vision Research*, *46*(26), 4416-4424.

https://doi.org/10.1016/j.visres.2006.08.027

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 and*

Review, *14*(4), 762–769. https://doi.org/10.3758/BF03196834

Phillips, N. (2017). yarrr: A Companion to the e-Book "YaRrr!: The Pirate's Guide to R". R

package version 0.1.5. https://CRAN.R-project.org/package=yarrr

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.

https://doi.org/10.1080/20445911.2017.1340303

Parker, A. J., Nikolova, M., Slattery, T. J., Liversedge, S. P., & Kirkby, J. A. (2019).

Binocular coordination and return-sweep saccades amongst skilled adult readers.

Journal of Vision, 19, 1-19. DOI:10.1167/19.6.10.

Parker, A. J., Slattery, T. J., & Kirkby, J. A. (2019). Return-sweep saccades during reading in

adults and children. *Vision research*, *155*, 35-43.

https://doi.org/10.1016/j.visres.2018.12.007

Pollatsek, A., Raney, G. E, Lagasse, L., & Rayner, K. (1993). The use of information below

fixation in reading and in visual search. Canadian Journal of Experimental

Psychology, 47, 179-200. DOI: 10.1037/h0078824

Pollatsek, A., Treiman, R., Reichle, E. D., & Sheridan, H. (2015). E-Z Reader. *The Oxford Handbook of Reading*, (January 2015). https://doi.org/10.1093/oxfordhb/9780199324576.013.17

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85*(3), 618–660.

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* (Vol. 62). https://doi.org/10.1080/17470210902816461

Rayner, K. (2009). Eye Movements in Reading: Models and Data. *Journal of Eye Movement Research*, *2*(5), 1-10.

Rayner, K., & A. D. Well. 1996. Effects of contextual constraint on eye movements in reading: a further examination. *Psychonomic Bulletin & Review, 3*, 504–509.

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. https://doi.org/10.3758/BF03197692

Rayner, K., Ashby, J., Pollatsek, A., & Reichte, 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. https://doi.org/10.1037/0096-1523.30.4.720

Rayner, K., Binder, K. S., Ashby, J., & Pollatsek, A. (2001). Eye movement control in

     reading: Word predictability has little effect on initial landing positions in words.

     Vision Research, 41, 943–954.

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.

Reichle, E. D. (2006). Theories of the "eye-mind" link: Computational models of eye-

     movement control during reading. *Cognitive Systems Research*, *7*, 2-3.

Reichle, E. D. (2011). Serial-attention models of reading. In S. P. Liversedge, I. D. Gilchrist,

     & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 767-786). New

     York, NY, US: Oxford University Press.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye

     movement control in reading. *Psychological Review*, *105*(1), 125–157.

Reichle, E. D., Warren, T., & Mcconnell, K. (2009). NIH Public Access, *16*(1), 1–21.

     https://doi.org/10.3758/PBR.16.1.1.

Reilly, R. G., & Radach, R. (2002). Glenmore: An interactive activation model of eye

     movement control in reading. *ICONIP 2002 - Proceedings of the 9th International*

     Conference on Neural Information Processing: Computational Intelligence for the E-

     Age, *3*(June 2014), 1194–1200. https://doi.org/10.1109/ICONIP.2002.1202810

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.

     https://doi.org/10.1016/j.cogpsych.2012.03.001

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, *7*(1), 206-271.

https://doi.org/10.1080/17470218.2017.1310262

Risse, S., & Kliegl, R. (2014). Dissociating preview validity and preview difficulty in

parafoveal processing of word n 1 during reading. *Journal of Experimental*

Psychology: Human Perception and Performance, 40, 653–668.

http://dx.doi.org/10.1037/a0034997

Schad, D. J., & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus

normal text reading using the SWIFT model. *Visual Cognition*, *20*(4–5), 391–421.

https://doi.org/10.1080/13506285.2012.670143

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention,*

Perception, and Psychophysics, *74*(1), 5–35. https://doi.org/10.3758/s13414-011-

0219-2

Schotter, E. R., Lee, M., Reiderman, M., & Rayner, K. (2015). The effect of contextual

constraint on parafoveal processing in reading. *Journal of Memory and Language, 83*,

118–139. http://dx.doi.org/10.1016/j.jml .2015.04.005

Schotter, E. R., von der Malsburg, T., & Leinenger, M. (2019). Forced fixations, trans-

saccadic integration, and word recognition: Evidence for a hybrid mechanism of

saccade triggering in reading. *Journal of Experimental Psychology: Learning,*

Memory, and Cognition, *45*(4), 677.

Sereno, S. C., Brewer, C. C., & O'Donnell, P. J. (2003). Context effects in word recognition:

Evidence for early interactive processing. *Psychological Science, 14,* 328–333.

Shebilske, W.L., 1975. 'Reading eye movements from an information-processing point of

view'. In: W. Massaro (ed.), *Understanding language: An information-processing*

analysis of speech perception, reading, and psycholinguistics. New York: Academic

Press.

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. https://doi.org/10.1103/PhysRevE.96.063110

Slattery, T.J. & Parker, A.J. (2019). Return-sweeps in reading: Processing implications of undersweep-fixations. *Psychonomic Bulletin & Review,* *https://doi.org/10.3758/s13423-019-01636-3*

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.

Slattery, T. J., Pollatsek, A., & Rayner, K. (2007). The effect of the frequencies of three consecutive content words on eye movements during reading. *Memory and Cognition*, *35*(6), 1283–1292. https://doi.org/10.3758/BF03193601

Slattery, T. J., Staub, A., & Rayner, K. (2012). Saccade launch site as a predictor of fixation durations in reading: Comments on Hand, Miellet, O'Donnell, and Sereno (2010). *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 251-261.

Slattery, T. J., & Vasilev, M. R. (2019). An eye-movement exploration into return-sweep targeting during reading. *Attention, Perception, & Psychophysics*, 81(5), 1197-1203. https://doi.org/10.3758/s13414-019-01742-3

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading times is logarithmic. *Cognition*, *128*, 302-319. DOI: 10.1016/j.cognition.2013.02.013

Snell, J., & Grainger, J. (2018). OB1-Reader: A Model of Word Recognition and Eye Movements in Text Reading Strategy analysis in probabilistic categorization View project Retrograde amnesia View project, *125*(6), 969–984. https://doi.org/10.1037/rev0000119

Stanovich, K. E., & West, R. F. (1983). On priming by sentence context. *Journal of Experimental Psychology: General, 112,* 1–36.

Staub, A. (2015). The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation. *Linguistics and Language Compass*, *9*(8), 311–327. https://doi.org/10.1111/lnc3.12151

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*, 110-127. https://doi.org/10.1037/xlm0000561

Stern, J.A. (1978). Eye Movements, Reading and Cognition. In J.W. Denders, D.F. Fisher, & R.A. Monty (Eds.), Eye Movements and the Higher Psychological Functions (pp. 145-155). Hillsdale, New Jersey: Lawrence Earlbaum Associates.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57,* 421–457.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, *30*, 415-433.

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. https://doi.org/10.1080/17470218.2013.850521

Vasilev, M. R., Slattery, T. J., Kirkby, J. A., & Angele, B. (2018). What are the costs of degraded parafoveal previews during silent reading? *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(3), 371–386. https://doi.org/10.1037/xlm0000433

Veldre, A., & Andrews, S. (2018). Parafoveal preview effects depend on both preview

    plausibility and target predictability. *Quarterly Journal of Experimental Psychology,*

    71, 64–74. http://dx.doi.org/10.1080/17470218.2016.1247894

West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of

    sentence context on word recognition. *Journal of Experimental Psychology:*

    Learning, Memory, and Cognition, 8, 385–399.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in

    experiments in which samples of participants respond to samples of stimuli. *Journal*

    of Experimental Psychology: General, 143(5), 2020-2-45.

White, S. J., Rayner, K., & Liversedge, S. P. (2005). The influence of parafoveal word length

    and contextual constraint on fixation durations and word skipping in reading.

    Psychonomic Bulletin & Review, 12, 466–471. http://dx.doi.org/10.3758/BF03193789

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.

    https://doi.org/10.1080/17470218.2013.848216