PERGAMON

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Vision Research 43 (2003) 1735–1751

Vision Research

www.elsevier.com/locate/visres

# Low-level predictive inference in reading: the influence of transitional probabilities on eye movements ☆

Scott A. McDonald *, Richard C. Shillcock

*Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ Scotland, UK*

Received 20 November 2002; received in revised form 27 March 2003

**Abstract**

We report the results of an investigation into the ability of transitional probability (word-to-word contingency statistics) to account for reading behaviour. Using a corpus of eye movements recorded during the reading of newspaper text, we demonstrate both the forward $[P(n|n-1)]$ and backward $[P(n|n+1)]$ transitional probability measures to be predictive of first fixation and gaze durations: the higher the transitional probability, the shorter the fixation time. Initial fixation position was also affected by the forward measure; we observed a small rightward shift for words that were highly predictable from the preceding word. Although transitional probability is sensitive to word class, with function words being generally more predictable from their context than content words, the measures accounted equally well for the data for both classes.


*Keywords:* Reading; Eye movements; Fixation duration; Fixation location

## 1. Introduction

The predictability of an element of language, such as a word, from the linguistic context in which it appears, has a major impact on its relative ease or difficulty of processing (e.g., Altarriba, Kroll, Sholl, & Rayner, 1996; Stanovich & West, 1983). Although it is well-established that contextual predictability can influence eye movement behaviour during reading, the source and temporal locus of this influence are not yet clear. Several studies have shown that manipulations of 'high-level' contextual constraint (i.e., requiring the integration of the meanings of individual words in the context) influence 'late' processing measures, such as second-pass reading time and the probability of making a regression (e.g., Calvo & Meseguer, 2002; Rayner & Well, 1996). However, the evidence for predictability effects on 'early' processing measures, such as the duration of the initial fixation made on a word, is mixed, with some studies reporting positive results (Binder, Pollatsek, & Rayner, 1999), and others failing to observe a reliable influence (Altarriba et al., 1996; Balota, Pollatsek, & Rayner, 1985; Drieghe, Brysbaert, Desmet, & De Baecke, in press; Rayner & Well, 1996).

Experimental manipulations of contextual constraint/predictability have been shown to affect both the *where* and *when* components of eye movement control. Whether or not a given word was directly fixated (a where decision) is a good indicator of early influences on the where component. Predictable words are more likely to be skipped than less predictable words (e.g., Brysbaert & Vitu, 1998; Drieghe et al., in press; Ehrlich & Rayner, 1981; Rayner & Well, 1996). However, the ability of predictability to affect a saccade's landing position within a word is controversial (see Lavigne, Vitu, & d'Ydewalle, 2000; Rayner, Binder, Ashby, & Pollatsek, 2001). Evidence for late influences on the where component is provided by analyses of regressive eye movements: predictable words are regressed to less often than less predictable words (Calvo & Meseguer, 2002; Rayner & Well, 1996). The when component of eye guidance is also clearly affected by predictability, with effects reported using a variety of processing measures. For example, predictability has been found to influence 'early' measures such as first fixation duration (Binder et al.,

☆ Part of this research was presented at the 10th Conference on Architectures and Mechanisms for Language Processing (AMLaP), September 19–21, 2002, in Tenerife, Spain.

* Corresponding author. Tel.: +44-131-650-3423; fax: +44-131-650-3461/3423.

*E-mail address:* scott.mcdonald@ed.ac.uk (S.A. McDonald).

0042-6989/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0042-6989(03)00237-2

1999; Morris, 1994), 'mid' measures such as gaze duration (e.g., Balota et al., 1985; Binder et al., 1999; Rayner & Well, 1996), and 'late' measures such as second-pass and total reading time (Calvo & Meseguer, 2002; Rayner & Well, 1996), and the likelihood of making a regression (Calvo & Meseguer, 2002). Accordingly, there is little evidence that the locus of the effects of predictability—at least when operationalised using the Cloze procedure (see below)—is restricted to a particular segment of the time course of processing.

In this paper, we investigate the potential influence on early processing during reading of a simple source of predictability: the transitional probability of a pair of words, as computed from the British National Corpus (Burnage & Dunlop, 1992). Probabilistic information of this sort is conceivably available from one's reading experience, and it represents a computationally inexpensive source of linguistic knowledge that could be exploited by the language processor in reading. We propose that at least some portion of the variance in eye movement behaviour can be explained by word-to-word contingency statistics. [1] Using a corpus of eye movement data collected while subjects read passages of text, we explored the effects of transitional probability on fixation duration and skipping probability measures.

### 1.1. Measuring predictability

An established technique for measuring contextual predictability is the *Cloze* task (Taylor, 1953). In the most popular version of the procedure, subjects are presented with a set of incomplete sentences and are asked to supply the word that best completes each sentence. The proportion of subjects producing a particular response for a given sentence fragment is then calculated, and this quantity is considered as an estimate of the predictability of the completion word given the preceding sentence context. Typically large manipulations of Cloze probability are required in order to observe reliable effects on eye movement variables such as gaze duration and skipping rate (Hyona, 1993), with reported high-/low-predictability condition contrasts of 0.64/0.01 (Binder et al., 1999), 0.78/0.01 (Rayner et al., 2001, Experiment 2), and 0.41/0.04 (Rayner & Well, 1996). Using a regression analysis design and items that varied in Cloze probability from 0 to 0.96, Calvo and Meseguer (2002) found that a word's Cloze value was a unique predictor only of 'late' processing measures such as second-pass reading time and the number of regressions made to the word.

### 1.2. Transitional probabilities

An alternative approach to assessing the predictability of a word in context is to make use of corpus data. For instance, we can determine that on is highly predictable given the occurrence of *rely* simply by comparing the number of times the sequence *rely* on occurs in the corpus to the number of times that *rely* is followed by a different word. This view of predictability—*transitional probability*—does not require human judgements such as those obtained in the Cloze procedure, and so predictability can in principle be quantified for any word in the lexicon occurring in a particular context. We refer to corpus-derived transitional probability as a measure of 'low-level' predictability because no 'high-level' conceptual knowledge about word and context meaning (which is likely a major determinant of subjects' performance in the Cloze task) contributes to the measurement. In this respect, our work is complementary to other investigations of low-level determinants of eye movement behaviour, such as the statistical structure of the lexicon (e.g., Clark & O'Regan, 1999; Legge, Klitz, & Tjan, 1997) and the anatomical properties of the visual system (Shillcock, Ellison, & Monaghan, 2000).

We investigate two different ways of defining the transitional (or *bigram*) probability between two words, which we refer to as the 'forward' and 'backward' transitional probabilities. The conventional 'forward' definition $P(n|n-1)$ captures the predictability (or dependence) of a particular word from the immediately preceding word. This type of statistical knowledge is fundamental to the success of many language engineering applications concerned with automatic speech recognition, natural language understanding, generation and translation (e.g., Jelinek, Mercer, & Roukos, 1992). Statistical learning, specifically of the transitional probabilities between phonemes, has been proposed as a mechanism for the task of word segmentation during language development (Cairns, Shillcock, Chater, & Levy, 1997; Saffran, Aslin, & Newport, 1996). The 'backward' transitional probability $P(n|n+1)$, which measures the statistical dependence of a word on the immediately following word, has less utility in natural language processing tasks. However, Jurafsky, Bell, Gregory, and Raymond (2001) analysed a corpus of spontaneous speech using multiple regression analysis, and report $P(n|n+1)$ to be a significant predictor of certain phonological aspects of language production behaviour, such as the likelihood of vowel reduction and the duration of the word produced.

Transitional probabilities are estimated from the relative frequency information present in a large text

[1] An early precedent for our research is provided by Morton (1964), who investigated the influence of statistical contingencies on oral reading speed and eye movements. He compared various Markov orders of approximation to English, where a zero-order approximation is realised as a random list of words, and a first-order approximation reflects between-word sequential probabilities (i.e., transitional probabilities). Morton found that reading rate was faster and average fixation duration was shorter for the first-order compared with the zero-order approximation.

corpus and are typically adjusted using statistical smoothing techniques in order to deal with sparse data—those words and word sequences that do not occur in the corpus but for which an estimate is required. The basic formulations of the two measures are simple ratios of joint and marginal frequencies:

Forward transitional probability

$$P(n|n-1) = \frac{f(n-1,n)}{f(n-1)}$$

Backward transitional probability

$$P(n|n+1) = \frac{f(n,n+1)}{f(n+1)}$$

## 2. Methods

### 2.1. Corpus collection

We assembled a corpus consisting of excerpts from 10 articles selected from contemporary editions of Scottish and UK national broadsheet newspapers. Our goal was to give participants a representative sample of newspaper content to read; articles were therefore chosen to cover a wide range of topics. Article excerpts were slightly edited and formatted into double-spaced, left-justified pages of a maximum of 65 characters in width; each page consisted of not more than 10 double-spaced lines of text. Excerpts varied in length from 97 to 405 words, filling between one and four pages of display. The total length of the corpus was 2262 words, which occupied 23 display pages.

Twenty-three young adult participants, all native speakers of British English with uncorrected normal vision, each read the entire corpus of 10 excerpts while their eye movements were recorded. Participants were randomly assigned to one of three versions of the corpus; each version contained a different random order of article presentation.

Subjects were seated at a viewing distance of approximately 75 cm from a 15 in. RM VGA monitor. Stimuli were displayed in a monospaced font as light cyan characters on a black background. One character subtended 0.26° of visual angle. Eye movements were recorded from the right eye using a Fourward Technologies Generation 6.3 Dual Purkinje Image eyetracker. The eyetracker has a resolution of less than 1 min of visual arc. In order to minimise subjects' head movements, a bite-bar and forehead rest were employed. Verbal instructions were provided to the subject while setting up and positioning the eyetracker. Once the subject was suitably seated, the instrument was calibrated and calibration was checked by asking the subject to fixate on a pattern of small squares distributed throughout the display area. Once calibration was achieved to the operator's satisfaction, a screen of instructions was presented, in order to familiarise the subject with the procedure.

After the subject had finished reading each screen of text, a pattern of fixation squares was displayed in order that the experimenter could check the calibration accuracy, and re-calibrate if necessary. A yes–no comprehension question followed the end of each article to which the subject was required to respond using a button box; the message "ERROR" was displayed if the question was answered incorrectly.

### 2.2. Data selection

For the analyses presented below, the eyetracking data had to meet a number of criteria in order to be included. First, for the measures first fixation duration (FirstF), gaze duration (Gaze), and the probability of skipping (SkipP), the dataset was restricted as follows. The first fixation on every line, as well as any regressive fixations immediately following this fixation were disregarded, as they are most likely corrective saccades triggered when the return sweep falls short of the beginning of the line (Hofmeister, Heller, & Radach, 1999). All data associated with words preceded or followed by punctuation, as well as the first and last word of every line of text, were also excluded from consideration. Finally, all cases with missing data due to a blink or other signal irregularity were eliminated. After these constraints were applied, the data record consisted of 31,242 cases. For the initial landing position measure (LandPos), these constraints were relaxed slightly to include words followed by punctuation (Hill & Murray (2001) found no effect of following punctuation on this variable), and data for fixations falling on the last word of every line were retained.

## 3. Results

All data points in the figures represent values (e.g., mean durations or proportions) computed by aggregating the data from all 23 participants. However, all inferential statistical analyses were conducted treating subjects as a random effect, in order to partition between-subject variance from within-subject variance. Below, we present the results of testing our predictions about the influence of lexical statistics on a selection of eye movement variables. [2] The primary statistical technique employed was multiple linear regression analysis.

[2] In this article we focus on 'early' eye movement behaviour. Hence, we do not examine measures which reflect re-inspection of the text, such as total fixation time, second-pass reading time, the probability that the saccade leaving the target word was a regression, or the probability of making a regression back to the target word.

Regression techniques for repeated measures data were conducted as recommended by Lorch and Myers (1990, Method 3). For several of the analyses below, the dataset was partitioned in order to simulate the orthogonal manipulation of certain variables; in these instances conventional analysis of variance (ANOVA) methods were employed. All word frequency and transitional probability measures were first (natural) logarithmically transformed.

### 3.1. Global characteristics of the dataset

Before exploring the effects of transitional probability on reading behaviour, we present a global summary of our eye movement dataset. First, 44.3% of the words were skipped during first-pass [3] reading. The mean gaze duration was 266 ms, and the average duration of an individual fixation was 241 ms. The overall likelihood of refixation (the likelihood of making an additional fixation(s) on a word, given that it was fixated at least once) during first-pass reading was 0.093. Considering only saccades launched from and landing on the same line of text, the mean progressive saccade size was 8.58 character spaces. The overall regression probability (returning from a later position in the text to the target word, regardless of whether it was fixated previously) was 0.082. These global characteristics of our dataset are closely comparable to those reported for other text reading studies (e.g., Vitu & McConkie, 2000; Vitu, McConkie, Kerr, & O'Regan, 2001).

Of the 2262 words in the corpus, 52.6% were function (or closed-class) words. Category membership was determined using the word class information available in the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). Fig. 1 displays the relative token frequencies for each class, partitioned by word length. Function words comprised the majority of the short (1–4 letter) words. Function words differed substantially from content words with respect to skipping rate (63.0% compared to 21.5%), gaze duration (246 vs. 277 ms), refixation likelihood (0.040 vs. 0.125), but not regression probability (0.083 vs. 0.080).

### 3.2. Forward transitional probability $P(n|n-1)$

The transitional probabilities $P(n|n-1)$ were computed for every word pair in the newspaper corpus using the CMU-Cambridge statistical language modelling toolkit (Clarkson & Rosenfeld, 1997) in conjunction with the 100-million word British National Corpus (BNC). In order to provide probability estimates for word pairs that did not occur in the BNC, the software's

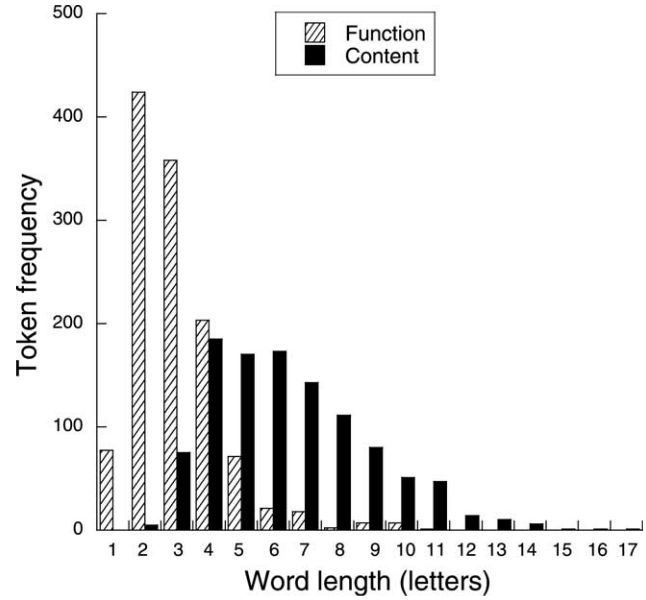


Fig. 1. The distribution of function and content words in the newspaper article corpus.

Good-Turing smoothing option was employed. This technique re-distributes some of the probability mass for more common events to infrequent and unseen events. For words following punctuation marks, a back-off strategy replaced $P(n|n-1)$ with the word's unigram probability $P(n)$. Fig. 2 displays the frequency histogram of the FTP values for all word pairs in the 2262 word corpus. Log-transformed transitional probabilities for our dataset (see above) ranged from −18.61 to −0.10 ($M = -5.63$, SD = 3.24). An example pair from the top end of the range is *undivided attention* with a forward transitional probability (FTP) of 0.4068 (−0.899 in natural log units). We can contrast this situation, where *attention* is quite predictable given *undivided*, with the pair *limited attention*, which has a low-FTP value of 0.0011 (−6.853 in log units).

The goal of the first set of multiple regression analyses was to determine if transitional probability was a significant predictor of reading behaviour when the effects of other factors were held constant. For instance, word frequency is naturally highly intercorrelated with transitional probability ($r = 0.814$ for the current dataset), and it is necessary to assess their independent predictive power. Word length (e.g., Just & Carpenter, 1980) and launch distance (Brysbaert & Vitu, 1998; Vitu et al., 2001) are also important determinants of eye movement behaviour. In all of the regression analyses reported below, the initial variables entered were subjects (dummy-coded), word length (WL) and launch distance (LD).

#### 3.2.1. First fixation duration

After the initial phase of the regression analysis, subjects, word length and launch distance explained a

[3] 'First-pass' refers to the initial fixation(s) made on a word until the eyes exit the word, either to the left or to the right. All regressions made to a word initially skipped or previously fixated are excluded.

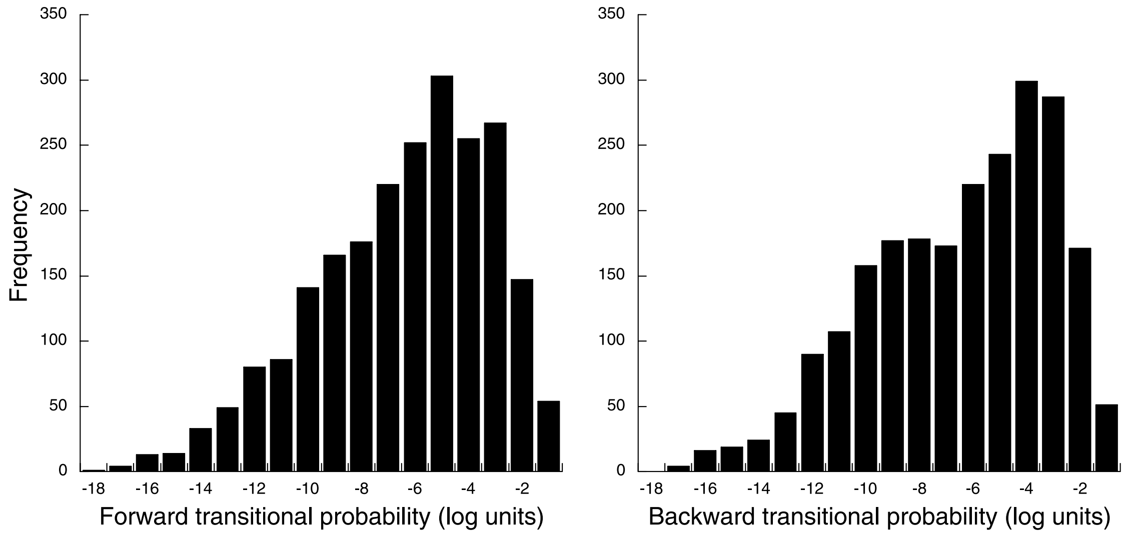


Fig. 2. Frequency histograms of the forward (left panel) and backward (right panel) transitional probability values for all words in the newspaper corpus. Bin size = one log unit.

total of 12.3% of the FirstF variance (WL and LD uniquely accounted for 1.2% and 1.4% of the variance, respectively). After removing variance due to these three variables, a simultaneous multiple regression revealed both frequency and FTP as significant independent predictors of first fixation duration: $\beta = -0.074$, $F(1,22) = 27.88$, $p < 0.001$; $\beta = -0.076$, $F(1,22) = 34.65$, $p < 0.001$, respectively. The negative slope coefficients indicate that FirstF tends to increase as either frequency or FTP decreases. The higher the FTP value, the shorter the duration of the initial fixation. Controlling for the other variables in the equation, first fixation duration increased 1.9 ms for every log unit decrease in FTP.

Because of their high intercorrelation, frequency and FTP compete for variance in the regression analysis, so it useful to provide the relevant figures when only one of the two was included in the regression. With FTP excluded, frequency explained 0.67% of the FirstF variance ($\beta = -0.123$, $F(1,22) = 117.00$, $p < 0.001$). When frequency was excluded, FTP accounted for 0.74% of the variance ($\beta = -0.112$, $F(1,22) = 143.43$, $p < 0.001$). Although the amount of the total variance in the current dataset explained by FTP is very small, in related work using pairs of length and frequency-matched words embedded in sentence materials, we found a reliable 11 ms FTP effect (McDonald & Shillcock, in press).

We next sought to determine if there was any dependence of the FTP effect on launch distance. If predictability effects are conditional on the availability of parafoveal visual information about the identity of word $n$ on the previous fixation, then we should observe larger FTP effects for near compared with far launch distances (all else being equal), because parafoveal preview is more viable the closer the previous fixation. To answer this question, we first partitioned the first fixation data into seven launch distance bins, and then conducted separate multiple regression analyses for each bin. After partialling out the effects of subjects, word length and word frequency, the forward transitional probability measure explained a significant amount of variance for the closest four launch distance bins only: [2,1], $\beta = -0.089$, $F(1,22) = 7.22$, $p < 0.05$; [4,3], $\beta = -0.102$, $F(1,22) = 17.74, p < 0.001$; [6,5], $\beta = -0.113, F(1,22) = 27.66$, $p < 0.001$; [8,7], $\beta = -0.084$, $F(1,22) = 17.57$, $p < 0.001$; respectively ($F$s $< 2$, $p$s $> 0.10$ for the other three bins). Thus, FTP has no reliable effect on FirstF beyond eight character spaces before the space before the target word.

Fig. 3 displays the mean first fixation durations for the high- and low-predictability cases (operationalised using the upper and lower quartiles of the data for each launch distance bin), for words of length four–seven letters. From the plots, it is apparent that mean first fixation duration for the high-FTP cases is generally shorter than for the low-FTP words. [4] There is a trend for the size of the FTP effect to decrease as launch distance increases. These findings suggest that some visual information about the identity of word $n$ is required in order for the transitional probability effect to be manifested. Predictions about the identity of $n$ given that $n - 1$ is currently fixated may need to be corroborated by partial visual information, such as the identity of the first few letters of $n$.

[4] It should be noted that the data points in Fig. 3 are the mean FirstF values for the extreme FTP quartiles; frequency remains a confounding variable.

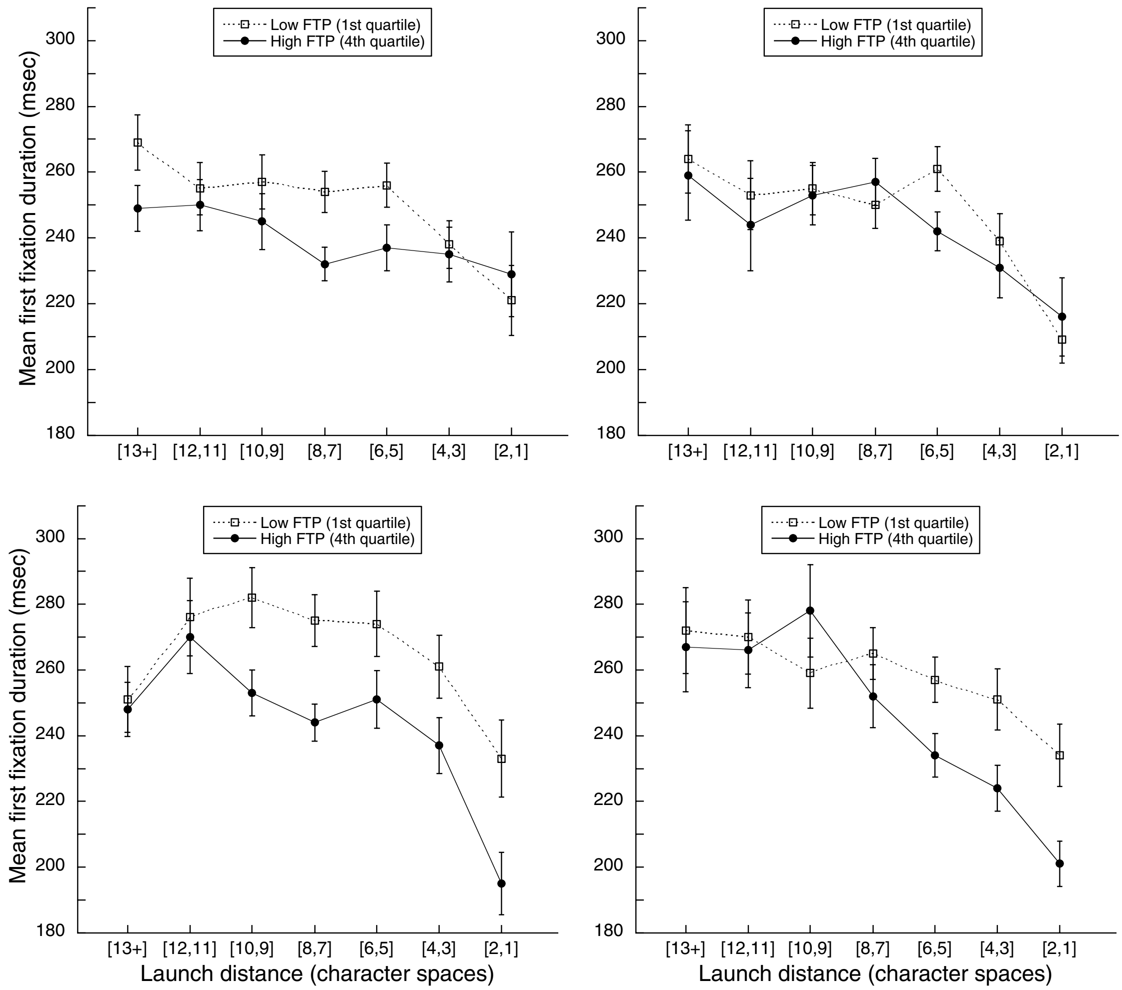


Fig. 3. Mean first fixation duration as a function of launch distance and forward transitional probability (low and high, operationalised as the first and fourth quartiles of the data for each launch distance bin), for 4-letter (upper left panel), 5-letter (upper right panel), 6-letter (lower left panel), and 7-letter (lower right panel) words.

### *3.2.2. Gaze duration*

The analysis procedure was identical to that for FirstF. Subjects, word length and launch distance explained a total of 14.7% of Gaze variance. In a simultaneous multiple regression additionally including frequency and FTP, both variables were significant predictors of Gaze: $\beta = -0.139$, $F(1,22) = 74.75$, $p < 0.001$; $\beta = -0.045$, $F(1,22) = 9.31$, $p < 0.01$, respectively. In contrast to the FirstF results, frequency had the stronger linear relationship with Gaze when FTP was also present in the equation. Gaze duration increased 5.05 ms for every log unit decrement of frequency, and increased 1.48 ms for every log unit decrement of FTP.

We also compared the predictive power of frequency and FTP with respect to the fixation duration of words receiving a single fixation (SingleF), which paralleled the FirstF results. Subjects, word length and launch distance explained 15.7% of SingleF variation. A simultaneous multiple regression adding frequency and FTP to these three variables indicated that both frequency and FTP were significant independent predictors of SingleF: $\beta = -0.084$, $F(1,22) = 25.79$, $p < 0.001$; $\beta = -0.075$, $F(1,22) = 28.84$, $p < 0.001$, respectively. When a single fixation is made on a word, the time the eyes spend on that word decreases as the forward transitional probability increases.

### *3.2.3. Skipping rate*

Brysbaert and Vitu (1998) conducted a meta-analysis of a number of studies manuipulating the degree to which a word was constrained by context, and arrive at an estimate of 11% of skipping rate variance explained by this factor. Furthermore, they note that although most of the skipping variance can simply be accounted for by word length, there is also an independent effect of launch distance. This launch distance effect was also apparent in our dataset; Fig. 4 displays the mean skipping rate as a function of launch distance for 4-, 5-, 6-, 7- and 8-letter words. We tested the hypothesis that an

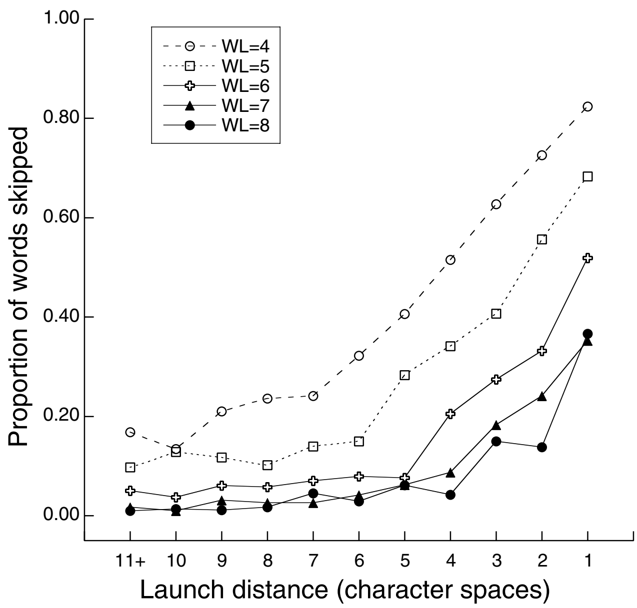


Fig. 4. The probability of skipping words of lengths 4–8 during first-pass reading as a function of launch distance and word length.

independent portion of skipping rate variance may be explained by transitional probability.

Whether or not a word skipped is a binary variable; linear logistic regression analyses were therefore appropriate in order to determine the influence of FTP on SkipP. As done in the FirstF and Gaze duration analyses, subjects, word length and launch distance were entered first into the regression. These three factors explained a moderate amount of the SkipP variance ($R^2 = 0.39$; Cox & Snell, 1981), with a classification accuracy of 80.2%. Adding frequency and FTP to the regression indicated that both variables were significant independent predictors: $B = 0.171$, $\text{Wald} = 230.71$, $p < 0.001$; $B = 0.034$, $\text{Wald} = 13.60$, $p < 0.001$, respectively, and adding either separately to the regression containing the other four variables resulted in significant improvements in model fit: $\chi^2 = 534.75$, $p < 0.001$; $\chi^2 = 13.53$, $p < 0.001$, respectively. However, the addition of FTP did not improve the overall classification accuracy (the percentage of skipped/not skipped cases correctly predicted by the equation) of the regression model over the performance of the model including frequency (80.4%). Although the addition of FTP was a statistically significant addition to the regression equation, the practical significance of its addition was negligible: no more cases were correctly classified. A word's predictability from the preceding word did not explain word skipping above and beyond the effect of its frequency of occurrence.

We next examined the potential role of visual influences on skipping probability, by partitioning the data according to launch distance to the target word. A difference in skipping rate between high- and low-FTP words, once word length and frequency are held constant, may become apparent for close launch sites only, if visual information about the target word is also necessary for a skipping decision to be made. Rayner, Sereno, and Raney (1996) found that the effect of frequency on skip rate was mediated by launch distance; they report a frequency effect for launch sites up to and including three character spaces before the target word only. Our finding of a negligible overall effect of predictability on skipping rate in the regression analysis may be due to the collapsing together of cases where parafoveal information about word $n + 1$ was readily available with cases where it was not.

We partitioned the eye movement data into seven launch distance bins ranging from [2,1] to [13+], and conducted separate logistic regression analyses for each bin, first removing variance from SkipP that was attributable to subjects, word length, and frequency. Consistent with the analysis of the non-partitioned dataset, there were no appreciable FTP effects. In contrast to the results of the FirstF analyses, there was no interaction between FTP and launch distance on the probability of skipping. The biggest improvement in model fit was observed for the [6,5] bin ($\chi^2 = 7.65$, $p < 0.01$); however, the addition of FTP to the regression model resulted in the correct classification of only 12 more cases. We conclude that the failure of FTP to explain an appreciable amount of SkipP variance in the overall regression analysis was not due to the collapsing together of cases where parafoveal visual information about the target word was available with cases where it was not.

To summarise the forward transitional probability findings: all three fixation duration measures tested were influenced by FTP; fixations on words that were predictable from the immediately preceding word (possessing high-FTP values) were shorter than fixations on less-predictable words (possessing low-FTP values), controlling for potentially confounding factors. The likelihood of skipping a given word was not influenced by FTP once other factors such as word frequency were statistically controlled.

### 3.3. Backward transitional probability $P(n|n + 1)$

As mentioned above, the backward transitional probability (BTP) has been shown to be a significant predictor of certain phonological characteristics of language production such as duration and vowel reduction. Based on Jurafsky et al.'s (2001) findings, we tested the hypothesis that the predictability of the current word given the *following* word would influence the eye movement behaviour associated with the current word. To illustrate, we predicted that the first fixation and gaze duration on a word such as *last* when followed by *year* (a high-backward transitional probability of 0.1416) would be shorter than when *last* is followed by *picture* (which has a low-backward transitional probability of 0.0010).

Using the BNC and the same methods employed for the forward transitional probability (see above), we computed the smoothed conditional probabilities for each word in the corpus, given the next word. The frequency histogram of the BTP values for all word pairs in the 2262 word corpus is displayed in Fig. 2. Log-transformed backward transitional probability values for our dataset ranged from −16.65 to −0.02 ($M = -5.59$, $SD = 3.25$). The forward and backward transitional probabilities in the dataset were highly correlated: Pearson's $r$ was 0.719. This correlation likely arises from the fact that word pairs are often equally mutually predictable (e.g., compound nouns such as *shopping mall*). Word frequency and backward transitional probability were also highly intercorrelated: $r = 0.805$. Because of these strong intercorrelations, these three variables would compete for variance in a multiple regression analysis. We therefore decided to control for the confounding effects of FTP and frequency by drastically reducing FTP and frequency variability in the dataset. First, for each of five word lengths (3–7 letters), we truncated the FTP range by retaining only those cases that fell inside the mean FTP value plus or minus 0.5 standard deviations. This step excluded 62.2% of the cases (aggregating the data for all five word lengths), retaining those cases occupying the centre of the distribution. Second, we restricted the word frequency range of the remaining data, by excluding those cases outside of the mean frequency value, plus or minus one standard deviation, which retained 67.5% of the data (25.5% of the original data for these five word lengths). Finally, of the remaining data we designated the upper quartile of the BTP range as the 'high' predictability condition, and the lower quartile as the 'low' condition, and computed the appropriate eye movement variables for both partitions. Table 1 displays the sample sizes, means and standard deviations of the FTP and frequency variables for this restricted dataset.

Table 1
Descriptive statistics for the restricted dataset (words of length 3–7 only) used to test the influence of the backward transitional probability measure on FirstF and SkipP

| Word | | FTP | | | Frequency | | |
|---|---|---|---|---|---|---|---|
| Length | $n^a$ | Mean | SD | $n^b$ | Mean | SD | $n^c$ |
| 3 | 2230 | −4.04 | 2.09 | 638 | 12.76 | 1.75 | 456 |
| 4 | 2863 | −5.78 | 2.09 | 1105 | 11.57 | 1.19 | 779 |
| 5 | 2124 | −6.57 | 2.42 | 893 | 10.22 | 1.13 | 578 |
| 6 | 2174 | −8.38 | 2.61 | 831 | 9.27 | 1.30 | 536 |
| 7 | 1684 | −8.70 | 2.61 | 715 | 8.95 | 0.99 | 472 |

*Note*: SD = standard deviation; FTP = forward transitional probability, in natural log units. Frequency is also in log units.
[a] Sample size before truncating the dataset.
[b] Sample size after elimination of cases outside of the range (mean FTP ±0.5 SDs).
[c] Sample size after elimination of cases outside of the range (mean frequency ±1 SD).

### 3.3.1. First fixation duration

The duration of first fixations on the target word was reliably influenced by the transitional probability of word $n$ given word $n+1$. Fig. 5 displays the mean FirstF values for the upper and lower quartiles of the data, for each word length examined. The mean duration of initial fixations made to words with a high-BTP value was shorter than the durations of fixations made to words with a low-BTP value (238 and 253 ms, respectively, averaged over the mean value for each word length). This difference was apparent for all word lengths for which sufficient data were available. The BTP effect was also statistically reliable: a two-way analysis of variance, with subjects as a random factor and BTP and word length as independent variables, indicated a significant main effect of BTP only: $F(1,22) = 8.29$, $p < 0.01$; $F(4,88) < 1$, respectively, with no interaction: $F(4,88) < 1$.

As reported above, the FTP effect is modulated by launch distance: the further away the previous fixation from the target word, the smaller the predictability advantage. We proposed an explanation based on the availability of parafoveal preview. If fixation durations are also influenced by the predictability of the current word given the *next* word, then it seems possible that this effect will be similarly mediated by the visibility of the conditioning word ($n+1$). However, in this case it is not parafoveally previewed letter information about word $n$ that is relevant, but instead the visibility of word $n+1$ during the current fixation on word $n$. Therefore, rather

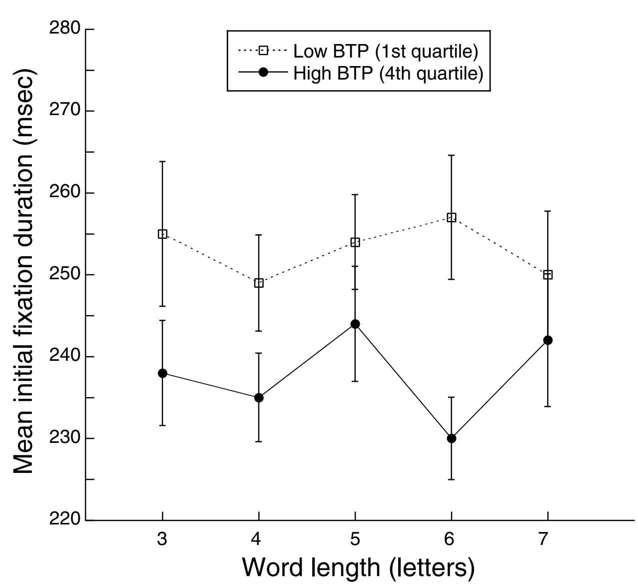


Fig. 5. Mean first fixation duration as a function of the backward transitional probability measure (low and high, operationalised as the first and fourth quartiles of the data for each word length) for words of length four–seven.

than see an interaction between BTP and the *launch distance* to word $n$, we should observe an interaction with the distance between the *fixation position* on $n$ and the beginning of word $n+1$ (i.e., the eccentricity of $n+1$). The predictability advantage should be larger the more of $n+1$ that is visible from the current fixation location on $n$. Because of insufficient data, it was not possible to do this analysis separately for each word length; Instead, we pooled the data for words of length 3–7. Fig. 6 plots the mean first fixation durations for the high- and low-BTP conditions (the fourth and first quartiles of the truncated dataset, determined separately for each word length) as a function of eccentricity. For instance, if the fourth letter of a 5-letter target word was fixated, the eccentricity of word $n+1$ is three character spaces. A clear relationship between the size of the BTP effect and the eccentricity of the conditioning word ($n+1$) is evident. There is no difference between the high- and low-predictability conditions beyond a distance of four characters before the first letter of $n+1$. This observation was confirmed by separate $t$-tests conducted for each eccentricity. The difference between the high- and low-BTP condition was statistically reliable for eccentricities of 3 and 4 only: $t(22) = 3.16$, $p < 0.005$; $t(22) = 3.74$, $p < 0.001$, respectively.

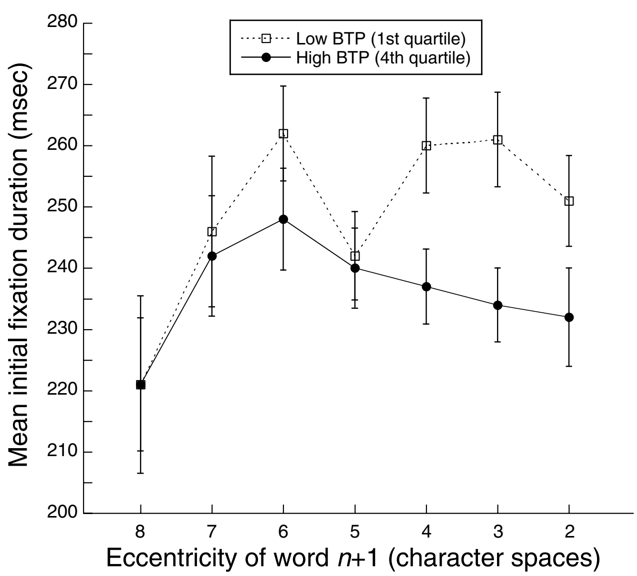


Fig. 6. Mean first fixation duration as a function of the eccentricity of word $n+1$ (collapsing together the data for 4–7 letter words) and backward transitional probability (low and high, operationalised as the first and fourth quartiles of the data for each eccentricity).

To verify that the data representing each eccentricity value were evenly distributed over the high- and low-BTP conditions (i.e., to check that the words occupying a particular eccentricity/BTP quartile cell did not differ markedly in frequency or FTP from the other BTP condition for the same eccentricity), we computed mean FTP and frequency values for each combination. Table 2 displays the sample sizes and mean frequency and FTP values for the cases associated with each eccentricity. Although there was virtually no difference in mean FTP values between the two BTP conditions (mean difference = 0.09), mean log word frequency was slightly higher on average in the high-BTP condition compared with the low-BTP position (mean difference = 0.49 log units). It is very unlikely that this small frequency difference was responsible for the large observed difference in FirstF. The regression equation derived from the model including subjects, word length, launch distance and frequency indicated a predicted increase in FirstF duration of 3.3 ms for every log unit decrease in frequency. As the largest difference in mean log frequency indicated in Table 2 was only 0.78 (when eccentricity = 6 character spaces), this frequency difference would account for at most 2.6 ms of the BTP effect.

### 3.3.2. Gaze duration

The results of comparable analyses of the Gaze values for the restricted dataset were very similar to the results of the FirstF analyses reported above. Fig. 7 displays the mean gaze duration for the 3–7-letter words for the two BTP conditions. From Fig. 7, it is apparent that for all five word lengths, words that were more predictable given the immediately following word had shorter gaze

Table 2
Descriptive statistics for the restricted dataset (aggregating words of length 3–7) used to examine the influence of the eccentricity of the conditioning word ($n+1$) on the BTP effect

| Eccentricity | FirstF (low-BTP) | | | FirstF (high-BTP) | | | Mean frequency | | Mean FTP | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SE | $n$ | Mean | SE | $n$ | Low-BTP | High-BTP | Low-BTP | High-BTP |
| 2 | 251 | 7.41 | 112 | 232 | 8.05 | 100 | 10.99 | 11.35 | −5.91 | −5.79 |
| 3 | 261 | 7.70 | 130 | 234 | 6.05 | 149 | 10.94 | 11.00 | −5.96 | −6.25 |
| 4 | 260 | 7.74 | 149 | 237 | 6.12 | 128 | 10.51 | 11.07 | −6.41 | −6.24 |
| 5 | 242 | 7.21 | 130 | 240 | 6.57 | 112 | 10.32 | 10.91 | −6.72 | −6.31 |
| 6 | 262 | 7.74 | 107 | 248 | 8.31 | 95 | 9.59 | 10.37 | −7.34 | −6.96 |
| 7 | 246 | 12.30 | 56 | 242 | 9.84 | 54 | 9.00 | 9.70 | −7.83 | −7.80 |
| 8 | 221 | 10.86 | 32 | 221 | 14.48 | 27 | 8.76 | 9.12 | −8.41 | −8.65 |

*Note*: SE = standard error; FirstF = first fixation duration; BTP = backward transitional probability; FTP = forward transitional probability, in natural log units. $n$ is the number of cases from which each mean value is computed. Eccentricity values represent the number of character spaces between the current fixation location and the beginning of word $n+1$. The low- and high-BTP conditions correspond to the first and fourth quartiles of the available data for each word length.

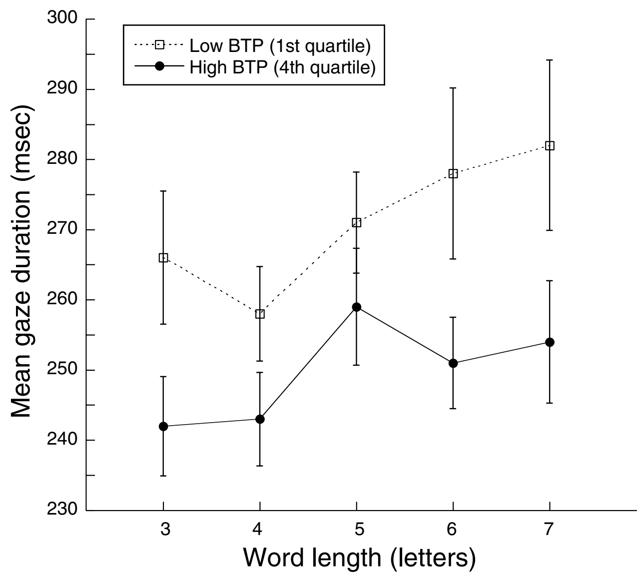


Fig. 7. Mean gaze duration as a function of the backward transitional probability measure (low and high, corresponding to the first and fourth quartiles of the data for each word length) for 4–7-letter words.

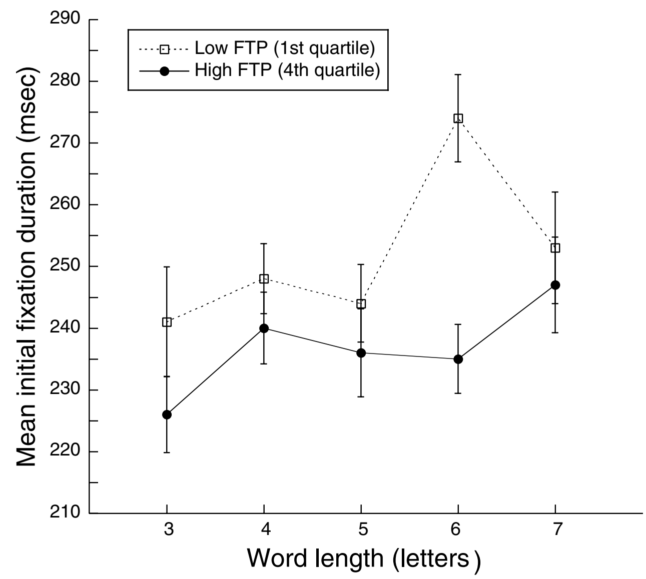


Fig. 8. Mean first fixation duration as a function of the forward transitional probability measure (low and high, operationalised as the first and fourth quartiles of the data for each word length) for words of length four–seven.

durations than words that had low-BTP values (250 and 271 ms, respectively, averaged over the mean value for each word length). A two-way ANOVA confirmed the statistical significance of the BTP effect: $F(1,22) = 11.49$, $p < 0.01$. There was no effect of word length: $F(4,88) = 1.03$, $p = 0.396$, nor an interaction: $F < 1$.

*3.3.3. Skipping rate*

Because the FTP and frequency range constraints were derived from the dataset containing only fixated words, we used the same procedure as described above to create a new truncated dataset containing both fixated and skipped cases. After range restrictions were applied, 25.7% of the cases were retained for analysis. A two-way ANOVA indicated that the probability of skipping a given word was not reliably influenced by its backward transitional probability: $F(1,22) = 1.08$, $p = 0.31$. Word length, as expected was a highly significant factor: $F(4,88) = 97.48$, $p < 0.001$; there was also a reliable BTP $\times$ word length interaction: $F(4,88) = 9.168$, $p < 0.001$. Separate ANOVAs of the data for each word length indicated that the interaction was due to the presence of the anticipated higher skipping rate for the high-BTP condition for the 3- and 6-letter words (*p*s $<$ 0.05), null effects for the 5- and 7-letter words, and a higher skipping rate for the *low*-BTP condition for the 4-letter words ($p < 0.001$).

## 3.4. *Reconciling influences of forward and backward transitional probability*

Although we have taken care to eliminate FTP as a confounding variable in the BTP analyses, in order to show that BTP differences are independently predictive of fixation duration differences, the possibility remains that the FTP effects we have reported for the FirstF, SingleF and Gaze dependent variables were in fact due to a BTP confound. In order to address this issue, we restricted the BTP and frequency ranges of the FirstF data for words of lengths 3–7 using exactly the same methods as we used to test for a BTP effect. After truncation, 23.3% of the original data remained for analysis. If BTP was a confounding variable in the FTP regression results, then we should observe no significant difference in first fixation durations between high-FTP and low-FTP cases. The mean FirstF values for the extreme quartiles of the data for each word length are plotted in Fig. 8. An ANOVA including FTP and word length as independent variables confirmed significant main effects of both FTP and word length: $F(1,22) = 8.14$, $p < 0.01$; $F(1,22) = 3.41$, $p < 0.05$, respectively, with a marginally significant FTP $\times$ WL interaction: $F(1,22) = 2.05$, $p = 0.096$. The mean effect size was 15 ms (estimated by averaging the differences in mean FirstF for each word length), which was identical to the effect size observed in the BTP analysis.

## 3.5. *Function and content words*

Within the lexicon a broad distinction can be drawn between content and function words. 'Content' (or *open-class*) words are members of those grammatical categories to which new words can easily be added: the nouns, verbs, adjectives and adverbs. 'Function' or *closed-class* words, are much less amenable to new members. In English at least, function words are generally more predictable from their immediately preceding context than are content words, in part due to the prevalence of phrasal and particle verbs, e.g., *rely on*, *pick up*, *preside over*. This is true of our newspaper corpus; the function words had a higher mean forward transitional probability than the content words ($M = -3.86$, SD = 1.88; $M = -8.24$, SD = 2.96, respectively). The function words in our corpus were also more predictable from their immediately *following* context than were content words; the mean backward transitional probability for the function words was −3.71 (SD = 1.89) compared with −8.43 (SD = 2.91) for the content items. Examples of high-BTP items include *the same*, *I am*, *to protect*, and *of course*.

Are the function words primarily responsible for the regression results reported above? Partitioning the dataset into function and content words and conducting separate regression analyses for each class did not confirm this suspicion. After partialling out subjects, word length, launch distance, and frequency, FTP was a significant predictor of FirstF in both the function-only and the content-only analyses: $\beta = -0.047$, $F(1,22) = 12.82$, $p < 0.01$; $\beta = -0.068$, $F(1,22) = 29.66$, $p < 0.001$, respectively. A between-group comparison of the regression model fits failed to indicate a class difference in terms of FirstF prediction performance: $Z = 0.75$, n.s. The same partitioning according to word class applied to the Gaze dependent variable gave comparable results: $\beta = -0.048$, $F(1,22) = 11.26$, $p < 0.01$; $\beta = -0.043$, $F(1,22) = 8.98$, $p < 0.01$, for the function and content items, respectively. There was also no reliable difference in the abilities of the two regression models to explain Gaze variance: $Z = -1.49$, $p > 0.05$. Separate logistic regression analyses with SkipP as the dependent variable showed a word class difference, however: the addition of FTP to the regression equation including subjects, word length, launch distance and frequency did not significantly improve model fit for the probability of skipping a content word: $\chi^2 = 0.275$, n.s. Whilst the addition of FTP improved model fit for the function word subset of the data ($\chi^2 = 24.97$, $p < 0.001$), there was virtually no improvement in classification accuracy over the model containing the other variables (an additional nine cases were correctly classified).

Because the number of function and content items was not distributed evenly over the different word lengths in the restricted dataset, it was only feasible to compare the influence of BTP on function and content fixation durations for the 4-letter words (the other four word lengths had distributions that were heavily skewed towards one of the classes, leaving insufficient cases to conduct analyses of both classes). An ANOVA incorporating word class and BTP as factors did not indicate main effects of either word class: $F(1,22) = 1.58$, $p = 0.222$ or BTP: $F(1,22) < 1$, or an interaction: $F < 1$.

### 3.6. Initial landing position

The location of the first fixation made on a word (or *landing position,* LandPos) is largely a function of low-level oculomotor factors: the word's length and launch distance (e.g., McConkie, Kerr, Reddix, & Zola, 1988; Radach & McConkie, 1998). Whether predictability can influence initial landing position is controversial, with one study finding an effect (Lavigne et al., 2000) that other researchers (Calvo & Meseguer, 2002; Rayner et al., 2001; Vonk, Radach, & van Rijn, 2000) have failed to observe. Lavigne et al. (2000) found that the initial landing position in high-frequency words, for near-launch distances of less than 5–7 character spaces, was further rightward when context made the target word predictable than when the context was neutral. We tested the hypothesis that transitional probability would also influence LandPos, predicting that words with high $P(n|n-1)$ values would be fixated further rightwards, consistent with Lavigne et al.'s findings.

As word length and launch distance are the two principle determinants of initial fixation position, we first partitioned the dataset according to these factors, and then plotted mean LandPos as a function of forward transitional probability and launch distance (see Fig. 9), for 4-, 5-, 6- and 7-letter words. High- and low-FTP conditions corresponded to the extreme quartiles of the distribution for each launch distance bin. [5] From Fig. 9, a small effect of FTP (on the order of ~0.15 of a letter space) is apparent for the 4-, 5- and 7-letter words, which does not appear to be restricted to a specific portion of the launch distance range. In order to confirm the presence of a FTP influence on LandPos, we conducted a three-way analysis of variance (predictability × launch distance × word length), including subjects as a random variable. In the cases where no data were available for one or more participants in a particular condition, the mean LandPos value for that condition contributed by the remaining participants was used. Replacement of missing data points was required for 1.8%, 2.2%, 2.5%, and 1.4% of the cells for the 4-, 5-, 6- and 7-letter words, respectively. The ANOVA revealed a reliable influence of FTP on LandPos: $F(1,22) = 11.564$, $p < 0.01$. As expected, there were also significant effects of both launch distance and word length: $F(5,110) = 184.31$, $p < 0.001$; $F(3,66) = 46.59$, $p < 0.001$, respectively. There were no reliable interactions between FTP and either launch distance or word length. We next conducted separate two-way ANOVAs on the data for each of the four word lengths. The separate ANOVA results indicated marginally significant

[5] We are confident that we do not need to control for frequency (e.g., by restricting the range of the data) for this analysis as neither Rayner et al. (1996) nor Lavigne et al. (2000) nor Calvo and Meseguer (2002) found any influence of frequency on initial landing position.

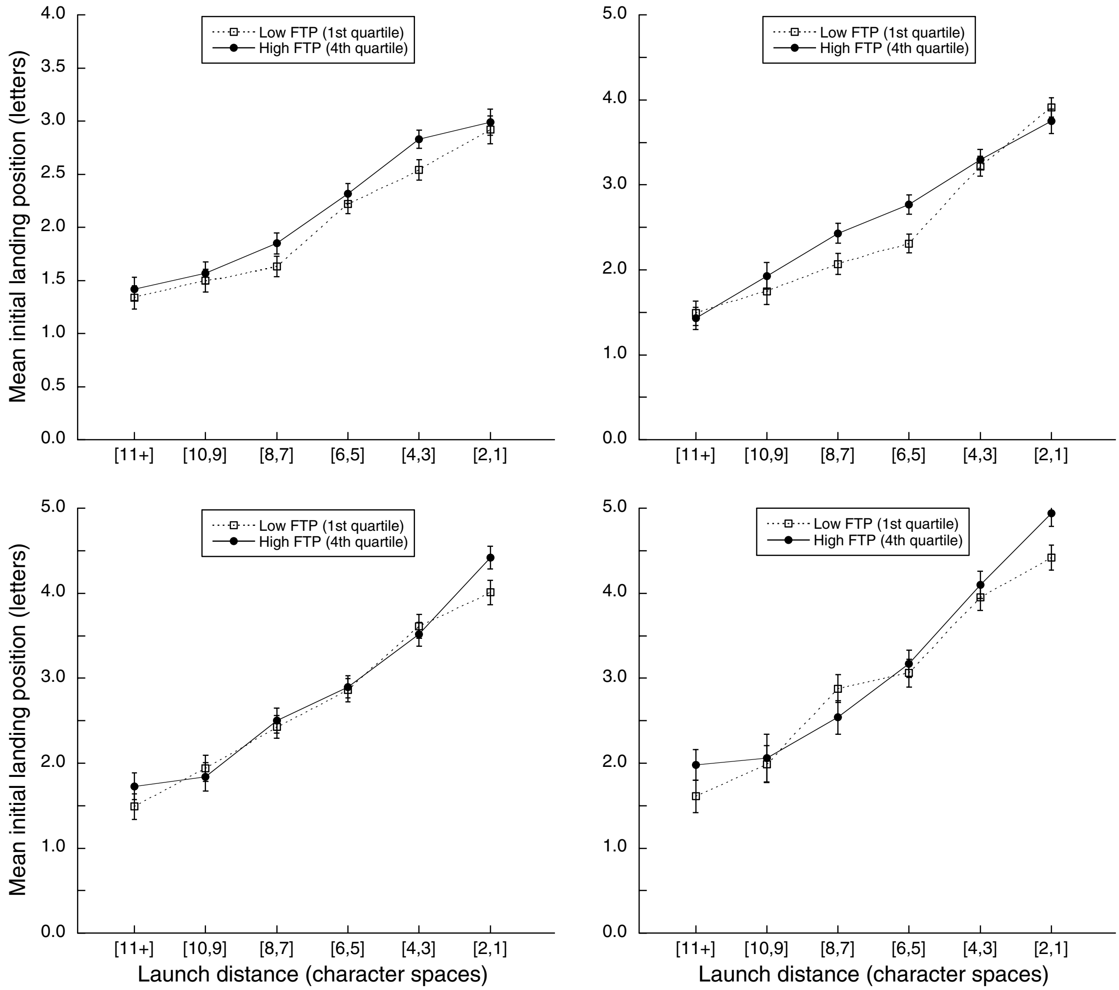


Fig. 9. The proportion of fixations landing on a given letter position of the target word, as a function of forward transitional probability (low and high), for 4-letter (upper left panel), 5-letter (upper right panel), 6-letter (lower left panel), and 7-letter (lower right panel) words.

main effects of FTP for the 4- and 5-letter words: $F(1,22) = 3.95,\ p = 0.059$; $F(1,22) = 3.98,\ p = 0.064$; respectively, but not for the 6- and 7-letter words ($p\text{s} > 0.15$).

We next tested for a corresponding effect of backward transitional probability on LandPos, for the same four word lengths. As done for the fixation time analyses, we first restricted the FTP range of the data for each word length, collapsing over all launch distance. This restriction retained 39.1% of the original data (pooling the data for all four word length). Fig. 10 displays the mean initial landing positions for the extreme BTP quartiles. In contrast to the FTP results, there was no consistent BTP effect. Only the 6-letter words showed an influence: mean LandPos was further rightwards when the word was highly predictable given the subsequent context than when unpredictable: $F(1,22) = 55.51,\ p < 0.001$. This difference was about 0.9 letters. However, for this word length the mean launch distance for the low-BTP condition was $\sim$0.8 character spaces closer than the launch distance for the low-BTP condition. Given that mean LandPos moves leftward $\sim$0.5 characters for every one-character increase in launch distance (McConkie et al., 1988; Radach & McConkie, 1998), this difference could explain less than half of the obtained BTP effect size.

Rayner et al. (2001) propose that the LandPos effect observed by Lavigne et al. (2000) was the result of intended skips of the target word falling short and landing on the last letter of the intended-to-be-skipped word. This suggestion is confirmed by their own data, which show an equal number of fixations on the last letter for both their high- and low-predictability conditions (see Rayner et al., 2001, Fig. 7). Hence, an influence of predictability on LandPos may be due to proportionally more fixations landing on the final one or two letters of predictable target words than would otherwise be expected. This explanation was not supported by our data, however. Fig. 11 displays the proportion of fixations landing on each of the possible fixation positions for the 5- and 6-letter words in our dataset, as a function of FTP. There is no indication that the rightward shift of

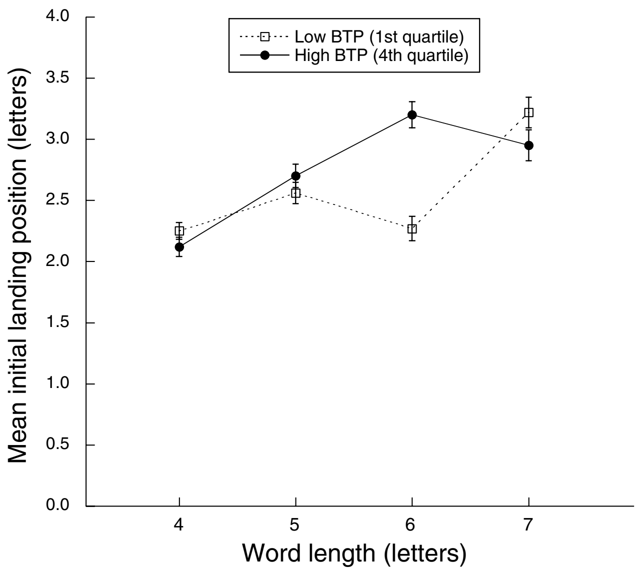


Fig. 10. Mean initial landing positions for 4-, 5-, 6- and 7-letter words, as a function of backward transitional probability (low and high, corresponding to the extreme quartiles).

the mean LandPos is due solely to a disproportionate number of high-FTP fixations landing on the final letter(s) of the target word. Thus, intended skips of word $n+1$ falling short of the intended target ($n+2$) is not a viable explanation of the FTP effect.

## 4. Discussion

The analyses presented here have shown that the transitional probabilities between two words, as computed from a large text corpus, influence several eye movement measures. Transitional probability was a significant predictor of the 'early' processing measures FirstF and SingleF, as well as the 'mid' measure Gaze. Both the forward and the backward transitional probability were unique predictors of first fixation and gaze durations. Our results have confirmed the existence of a new source of statistical information that we have shown to affect eye movement behaviour during reading.

The proportion of FirstF and Gaze variance explained by either transitional probability measure was small, though statistically reliable. The regression equation fit to the FirstF data indicated a decrease of 1.9 ms for every log unit of FTP; this result is in line with the 11 ms FTP effect obtained by McDonald and Shillcock (in press) using tightly controlled sentence stimuli, where the high- and low-predictability conditions differed by an average of 3.5 log units of FTP. Note that the average FTP and BTP effects of 15 ms that we obtained when controlling the other variable represent conservative estimates of effect size. Truncating the dataset to reduce the variability of potentially confounding variables meant that the range of the variable of interest was reduced compared with its range in the original dataset.

The presence of an influence of the backward transitional probability on fixation duration seems, at first, surprising. Because FTP was controlled by severely truncating its range, this BTP effect is independent from the influence of FTP on fixation duration. This result adds to the accumulating evidence for parafoveal-on-foveal effects (e.g., Inhoff, Radach, Starr, & Greenberg, 2000; Kennedy, 2000), but differs slightly from previous findings in that it is not the properties of the subsequent (parafoveal) word that influence the processing of the current (foveal) word, but rather it is the predictability of the current word from the subsequent word that is relevant. The fact that BTP affects the duration of the

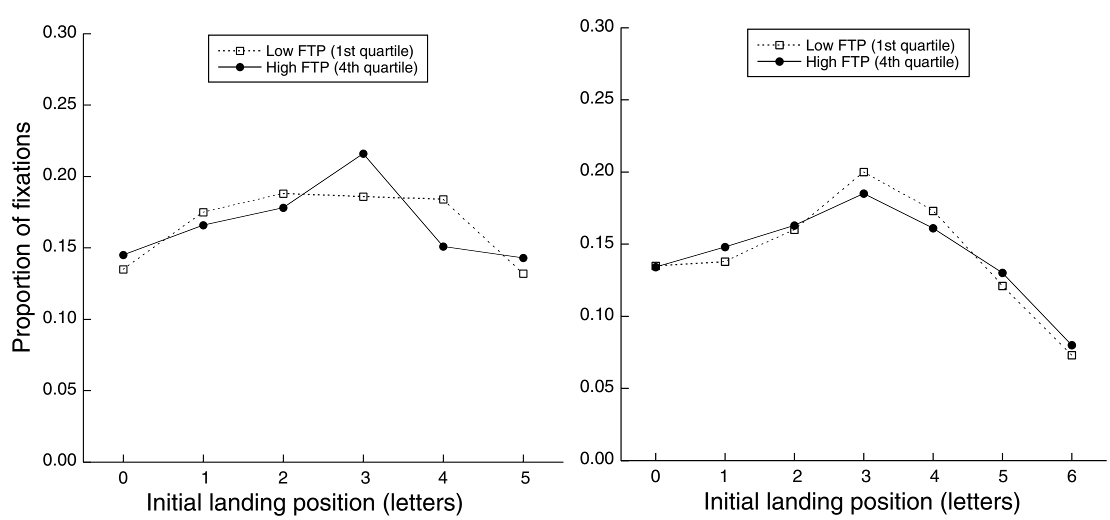


Fig. 11. Proportion of fixations landing on the different letter positions of the 5-letter words in the dataset, as a function of forward transitional probability (low and high, corresponding to the extreme quartiles).

first fixation made on a word strongly suggests that at least partial encoding of the parafoveal word occurs in parallel.

One important finding was that the influence of the forward transitional probability on first fixation duration was contingent on the availability of parafoveal information about the target word from the previous fixation location. The distance between the launch site and the first letter of the target word modulated the presence of the FTP effect, with the advantage for more-predictable words disappearing for far launch distance. At launch distance beyond roughly nine characters before the target word, the FTP effect disappeared. The same result was observed for the presence of the backward transitional probability effect, which also exhibited dependence on the eccentricity of the conditioning word $n+1$. When the position of the fixation on word $n$ was more than four characters in front of $n+1$, the BTP effect disappeared.

One explanation of these findings rests on visibility: parafoveal preview of the target noun is more viable the closer the previous fixation. Studies of eye movement behaviour when no preview is available (Balota et al., 1985; Inhoff & Rayner, 1986) have shown that parafoveal preview increases the efficiency of lexical processing. It appears that partial parafoveal visual information is required in order to corroborate statistical predictions about the identity of the target noun in order to facilitate lexical processing. It is interesting that the FTP effect was maintained at launch distances as far as 10 characters. From this distance, it is very unlikely that actual letters in the target word could be discerned (Underwood & McConkie, 1985); the only information available to corroborate predictions would be low-spatial frequency information such as word length. Further work is necessary to determine the relative importance of letter and length information.

Our results suggest that it is primarily the when, not the where of eye movement control that is influenced by low-level predictability. The failure of either transitional probability measure to explain an appreciable amount of SkipP variance, even when the dataset was partitioned according to the visibility of the target word (for the FTP analyses) or the conditioning word (for the BTP analyses), indicates that the selection of the target for the next saccade may be only minimally sensitive to this information. However, numerous studies have shown that skipping rate is influenced by predictability: highly constrained words (as determined by the Cloze task) are more likely to be skipped than less constrained words (e.g., Balota et al., 1985; Ehrlich & Rayner, 1981; Rayner & Well, 1996). Predictability as measured by Cloze procedure, as we have already mentioned, reflects high-level conceptual knowledge. This SkipP effect is conceivably the result of top-down processing: the next word is narrowed down sufficiently by the previous context that no direct fixation is deemed necessary. In contrast, 'low-level' predictability as measured by transitional probability may represent a form of bottom-up processing. As we have seen from the launch distance-partitioned analyses, at least partial parafoveal preview of the target word appears to be necessary in order to observe a facilitatory effect of a high-FTP on first fixation duration.

We found no word class differences with respect to the ability of the two transitional probability measures to predict any of the dependent variables. Fixation times on function and content words were explained equally well, even though function words are intrinsically more predictable. This result is consistent with a reading study by Schmauder, Morris, and Poynor (2000). Using matched sets of function and content words, Schmauder et al. found equivalent word frequency effects for both classes with first fixation duration and skipping rate as dependent variables; no interactions with word class were observed.

Finally, our analyses have shown a small, though consistent influence of FTP on the initial landing position measure, thus suggesting that saccade planning is also subject to influences from low-level predictability. This effect was not observed in the data for 6-letter words, but was on the order of 0.15 letter spaces for 4-, 5- and 7-letter words. Although saccade target selection was not appreciably influenced by FTP beyond the influences of word length, launch distance, and word frequency (there was no improvement in skipped/not-skipped classification accuracy), the mean landing position was further rightward when the target word $n$ was predictable given $n-1$ than when not. Modulation of the initial fixation position was only affected by the forward transitional probability; the backward measure had no consistent influence on LandPos.

### *4.1. Implications for models of eye movement control*

Is the influence of transitional probability compatible with recently proposed models of eye movement control? We briefly consider how FTP and BTP could be integrated into the most recent version of the EZ-Reader model (Reichle, Rayner, & Pollatsek, 2003) and the SWIFT model (Engbert, Longtin, & Kliegl, 2002). In both models, the 'high-level' notion of predictability is recognised for its influence on eye movements, and the behavioural consequences of predictability differences are implemented by increasing the time required for lexical processing as a linear function of (Cloze) predictability.

EZ-Reader 7 (Reichle et al., 2003) assumes that lexical processing occurs in two stages. The first stage (termed 'familiarity check' in earlier versions of the model) corresponds to identification of a word's orthographic form, and the second stage corresponds to lexical access. The time required to complete the first stage is a linear function of word frequency and high-level predictability. The time to complete stage 2 is also determined by frequency and predictability, with more weight attributed to predictability than in stage 1. If we assume that transitional probability can influence the first stage of lexical processing only, then a reasonable modification to try is to replace the predictability term in the stage 1 lexical processing equation with transitional probability, or to replace both frequency and predictability with a composite variable that estimates a word's probability of occurrence, integrating transitional probability with frequency in a principled way (see McDonald & Shillcock, in press).

A fundamental feature of the SWIFT model (Engbert et al., 2002) is its functional separation between saccade timing and saccade target selection. In order to extend this model to explain our findings, we might assume that transitional probability influences early processing stages; adjusting the parameters of the lexical processing function that determines when the next saccade is launched. Only high-level predictability is capable of affecting target selection; this distinction in locus between the two types of predictability would account for the observed null effect of transitional probability on skipping rate. [6]

Our findings that BTP influenced the fixation duration on word $n$, and the interaction of the BTP effect with eccentricity of $n+1$, pose difficulties for the sequential attention allocation assumption of EZ-Reader. In contrast, SWIFT assumes a spatial distribution of atttention, and thus parallel lexical processing of non-fixated words. It is reasonable to suppose that attention cannot be restricted to the foveal word if the statistical dependence between this word and the subsequent word influences fixation time on the currently fixated word.

### 4.2. Explaining the effects

We have presented the implications of transitional probability statistics on eye fixation behaviour without considering how the language processor would acquire such statistical information. We propose that exposure to written (and spoken) language is sufficient for compiling the necessary contingency statistics. The vast amount of language that one encounters represents a rich source of statistical knowledge, of which word-to-word contingencies are a straightforward example. Experience with reading, then, represents a form of *distributional learning*: because words do not occur in isolation—rather, they are encountered as part of syntactically coherent and semantically meaningful sequences—assimilating a new word into one's mental lexicon may also involve encoding its surrounding context. A distributional learning mechanism has already been put forward to explain how infants learn to segment the speech stream into words (Saffran et al., 1996), and a similar experience-based mechanism must underlie the acquisition of word frequency knowledge. Knowledge of the statistical likelihoods of linguistic entities occuring in context (word $n$ occurring after word $n-1$, and word $n$ occurring before word $n+1$) could be obtained using similar mechanisms that are deployed for acquiring probabilistic knowledge about particular events in the world occurring, given a conditioning event. To use a simple example, one's subjective probabilities of hearing thunder and seeing lightning increase given the presence of black clouds in the sky.

In order to assess the appropriateness of the noun *mice* as the object of a given verb, such as *train*, it is clear that one's experience with mice in the world and their ability to be trained would aid the interpretation of a sentence such as '*The child trained mice to jump at the sound of a bell*'. But no such 'world knowledge' is required for access to information about contingencies between the *words* referring to the entity and the action. Thus, effective comprehension could, at least part of the time, arise from sources that do not require high-level cognitive processing (cf. Fodor, 1983, p. 80). The prediction of upcoming words using lexical statistical information is a computationally inexpensive mechanism that may contribute to proficient reading.

While this explanation of role of transitional probabilities in reading is compelling, from the current study, it is not clear that predictability as measured using corpus statistical information is *sufficient* for influencing eye movements. In other words, the correlational design we have used cannot disconfound the high-level notion of predictability from the low-level transitional probability. It will certainly be true that many cases with high-forward ransitional probabilities would also have high-Cloze values; subjects provided with the context prior to the target word may be able to supply the correct word on the basis of an integrated meaning representation of the context, not only from $P(n|n-1)$. In other words, could high-level predictive processes explain our results? We think this is unlikely, for two reasons. First, theories of contextual integration would not predict the early effects we observed for the backward transitional probability measure, which appear to depend on at least partial identification of the following word. Second, in a related study (McDonald & Shillcock, in press) we tested pairs of nouns embedded in tightly controlled sentence contexts. These noun pairs were chosen to differ in FTP value but only minimally in Cloze value. The first fixation effect we obtained could not easily be explained by differences in high-level predictability.

In summary, we have provided evidence that the reading of a given word is influenced by predictability in

[6] We thank an anonymous reviewer for suggesting this interpretation.

the form of contingency statistics involving both the previous and subsequent context. Eye movement behaviour, namely the first fixation and gaze duration of the target word, was affected by both the forward and backward transitional probability measures. Fixation durations were shorter when the word was predictable than when not, and this effect was independent of other factors such as length, launch distance and frequency. Initial fixation position was also affected by the forward measure, but to a much smaller extent; word skipping was only appreciably influenced by the backward measure. Our findings are consistent with models of eye movement control that postulate distinct when and where mechanisms: our results demonstrate that transitional probability primarily influences the when of eye movement control.

## Acknowledgements

This work was supported by project grant GR064240AIA from the Wellcome Trust. The second author was also supported by a Senior Research Fellowship from the Economic and Social Research Council.

## References


Altarriba, J., Kroll, J., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition, 24*, 477–492.

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

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

Binder, K. S., Pollatsek, A., & Rayner, K. (1999). Extraction of information to the left of the fixated word in reading. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1162–1172.

Brysbaert, M., & Vitu, F. (1998). Word skipping: implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 125–147). Oxford: Elsevier.

Burnage, G., & Dunlop, D. (1992). Encoding the British National Corpus. In J. M. Aarts, P. de Haan, & N. Oostdijk (Eds.), *English language corpora: design, analysis, exploitation* (pp. 79–95). Amsterdam: Rodopi.

Cairns, P., Shillcock, R. C., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: a bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology, 33*, 111–153.

Calvo, M. G., & Meseguer, E. (2002). Eye movements and processing stages in reading: relative contributions of visual, lexical, and contextual factors. *The Spanish Journal of Psychology, 5*, 1138–7416.

Clark, J. J., & O'Regan, J. K. (1999). Word ambiguity and the optimal viewing position in reading. *Vision Research, 39*, 843–857.

Clarkson, P., & Rosenfeld, R. (1997). Statistical language modelling using the CMU-Cambridge Toolkit. In Proceedings of Eurospeech '97.

Cox, D. R., & Snell, E. J. (1981). *Applied statistics: principles and examples*. London: Chapman and Hall.

Drieghe, D., Brysbaert, M., Desmet, T., & De Baecke, C. (in press). Word skipping in reading: on the interplay of linguistic and visual factors. *European Journal of Cognitive Psychology*.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning & Verbal Behavior, 20*, 641–665.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research, 42*, 621–636.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Hill, R. L., & Murray, W. S. (2001). Commas and spaces: effects of punctuation on eye movements and sentence parsing. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a Perceptual Process* (pp. 565–589). Oxford: Elsevier.

Hofmeister, J., Heller, D., & Radach, R. (1999). The return sweep in reading. In W. Becker, H. Deubel, & W. Mergner (Eds.), *Current oculomotor research: physiological and psychological aspects* (pp. 349–357). New York: Plenum Press.

Hyona, J. (1993). Effects of thematic and lexical priming on readers' eye movements. *Scandinavian Journal of Psychology, 34*, 293–304.

Inhoff, A. W., Radach, R., Starr, M., & Greenberg, S. (2000). Allocation of visual-spatial attention and saccade programming during reading. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 221–246). Oxford: Elsevier.

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

Jelinek, F., Mercer, R. L., & Roukos, S. (1992). Principles of lexical language modelling for speech recognition. In S. Furui, & M. M. Sondhi (Eds.), *Advances in speech signal processing*. Maral Dekku.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: evidence from reduction in lexical production. In J. Bybee, & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: John Benjamin.

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

Kennedy, A. (2000). Parafoveal processing in word recognition. *Quarterly Journal of Experimental Psychology, 53A*, 429–455.

Lavigne, F., Vitu, F., & d'Ydewalle, G. (2000). The influence of semantic context on initial landing sites in words. *Acta Psychologica, 104*, 191–214.

Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: an ideal-observer model of reading. *Psychological Review, 104*, 524–553.

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 149–157.

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

McDonald, S. A., & Shillcock, R. C. (in press). Eye movements reveal the on-line computation of lexical probabilities. *Psychological Science*.

Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 92–103.

Morton, J. (1964). The effects of context upon speed of reading, eye movements, and eye-voice span. *Quarterly Journal of Experimental Psychology, 16*, 340–354.

Rayner, K., Binder, K. S., Ashby, J., & Pollatsek, A. (2001). Eye movement control in reading: word predictability has little influence on initial landing positions in words. *Vision Research, 41*, 943–954.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: a comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1188–1200.

Radach, R., & McConkie, G. (1998). Determinants of fixation positions in reading. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 77–100). Oxford: Elsevier.

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

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye movement control in reading: comparisons to other models. *Behavioral and Brain Sciences*.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science, 274*, 1926–1928.

Schmauder, A. R., Morris, R. K., & Poynor, D. V. (2000). Lexical processing and text integration of function and content words: evidence from priming and eye fixations. *Memory & Cognition, 28*, 1098–1108.

Shillcock, R., Ellison, M. T., & Monaghan, P. (2000). Eye-fixation behaviour, lexical storage and visual word recognition in a split processing model. *Psychological Review, 107*, 824–851.

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

Taylor, W. L. (1953). Cloze Procedure: a new tool for measuring readability. *Journalism Quarterly, 30*, 415–433.

Underwood, N. R., & McConkie, G. W. (1985). Perceptual span for letter distinctions during reading. *Reading Research Quarterly, 20*, 153–162.

Vitu, F., & McConkie, G. (2000). Regressive saccades and word perception in adult reading. In R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 165–191). Oxford: Elsevier.

Vitu, F., McConkie, G., Kerr, P., & O'Regan, J. K. (2001). Fixation location effects on fixation durations during reading: an inverted optimal viewing position effect. *Vision Research, 41*, 3513–3533.

Vonk, W., Radach, R., & van Rijn, H. (2000). Eye guidance and the saliency of words beginnings in reading text. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 269–299). Oxford: Elsevier.