The word grouping hypothesis and eye movements during reading.

Denis Drieghe[1], Alexander Pollatsek[2], Adrian Staub[2], and Keith Rayner[3]

[1]Ghent University, Belgium

[2]University of Massachusetts at Amherst

[3]University of California, San Diego

Address: Denis Drieghe
Department of Experimental Psychology
Ghent University
Henri Dunantlaan 2
B-9000 Ghent
Belgium
Tel. 32 – 9 264 64 05
Fax. 32 – 9 264 64 96
e-mail: denis.drieghe@UGent.be

Running Head: Word grouping hypothesis

Word Count: 4996 words

Abstract

The distribution of landing positions and durations of first fixations in a region containing a noun preceded by either an article (e.g. *the soldiers*) or a high-frequency three-letter word (e.g. *all soldiers*) were compared. Although there were fewer first fixations on the blank space between the high-frequency three-letter word and the noun than on the surrounding letters (and the fixations on the blank space were shorter), this pattern did not occur when the noun was preceded by an article. Radach (1996) inferred from a similar experiment that did not manipulate the type of short word that two words could be processed as a perceptual unit during reading when the first word is a short word. As this different pattern of fixations is restricted to article-noun pairs, it indicates that word grouping does not occur purely on the basis of word length during reading; moreover, as we demonstrate, one can explain the observed patterns in both conditions more parsimoniously, without adopting a word grouping mechanism in eye movement control during reading.

Based on an elegant analysis of a large corpus of eye movement data, Radach (1996) reported a rather surprising finding concerning the distribution of the eyes' landing sites. The typical finding for landing sites on a word is that they are normally distributed with the mean at the *preferred viewing location* (Rayner, 1979; Vitu, O'Regan, & Mittau, 1990), which is between the beginning and the middle of a word. However, Radach observed that when the first of two adjacent words was a three-letter word, the distribution of initial fixation locations on the two-word region seemed to constitute a single normal distribution. Based on these findings, Radach (1996) hypothesized that during reading, short word groups could serve as a perceptual unit (i.e., when the first word to the right of fixation is a short word, the reader may group this word and the next one as a perceptual unit and target fixations on this region as if it were a single word). In Radach's account, this mechanism is activated exclusively on the basis of parafoveal word length. This hypothesis is of interest because it would seem to undermine the conclusion that in normal reading words are processed and identified in a serial, left-to-right manner. A range of processing evidence has been taken to support this conclusion (e.g., Frazier & Rayner, 1982; Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007), which is formalized in some models of eye movement control, such as the E-Z Reader model (Pollatsek, Reichle, & Rayner, 2006; Reichle, Pollatsek, Fisher, & Rayner, 1998). Other models posit parallel processing of words within the perceptual span (e.g. the SWIFT model; Engbert, Nuthmann, Richter & Kliegl, 2005, and the Glenmore model; Reilly & Radach, 2006). In recent years, the question whether readers process information from more than one word at a time, has been at the core of many debates in the literature (e.g. Kliegl, 2007; Pollatsek et al., 2006), necessitating a closer examination of Radach's observations.

The main focus of the current experiment was an attempt to replicate Radach's (1996) finding in an experiment in which there was tighter control of the experimental materials. Radach's analysis did not distinguish between classes of three-letter words, and we wanted to determine whether his finding applied to all three-letter words followed by a longer word or was restricted to articles. There are two potentially distinct reasons why the processing of word groups as a perceptual unit during reading might be restricted to article-noun pairs: (a) the article might be treated by the processing system as a "default" determiner, thereby requiring little independent processing, and (b) because an article is so high in frequency, it can be easily processed even from a fixation location that is not optimal. In either case, the chances of the system being able to process the group in one - or in the case of longer nouns two - well-placed fixation(s) should be relatively high and thus a strategy for placing that first fixation on the group may be optimal. It is less likely that this would be the case when the three-letter word is not a "default" from the point of view of the language processing system, or is of lower frequency. Alternatively, if the finding of a single normal distribution of first fixation locations is restricted to article-noun pairs, it might be that the pattern of results can be explained on the basis of an already known mechanism (a higher skipping rate for an article than for a high-frequency three-letter adjective, see below), and doesn't require the introduction of a new mechanism in saccade planning during reading.

To test whether Radach's observation would be limited to article-noun pairs, we compared the distribution of initial landing sites on an article-noun pair (*the soldiers*) with a pair in which the article was replaced by a high-frequency three-letter word (*two soldiers*). We chose high-frequency three-letter words to make the test as severe as possible; that is, although the three-letter words we used were less frequent than the word *the*, they were still very high-frequency words. Thus, if Radach's finding is replicated only for the article-noun condition, we would have a strong case that such a grouping mechanism is not being triggered by the length of the word to the right of fixation.

The optimal region to focus on in comparing the two distributions of initial fixations is the blank space between the two words. As mentioned earlier, the typical finding in reading is that each word has a normal distribution of landing sites, with the peak of the distribution a bit to the left of the center of the word. As a consequence, the blank space between two words is typically fixated relatively infrequently (Rayner & McConkie, 1976; McConkie, Kerr, Reddix & Zola, 1988). This makes perfect sense from an information-processing perspective: due to visual limitations, the blank space is a poor location for encoding a word. As a likely consequence, when the eyes land on a blank space between two words, there is often a relatively short fixation on this location, followed either by a regression to the previous word or a progressive saccade to the next word (whichever was the intended target of the saccade). In the current analysis, we focus on the eye movement behavior on the blank space compared to the last letter of the three-letter word and the first letter of the following word. Should Radach's (1996) finding be restricted to article-noun pairs, we expect to observe both of the typical findings (i.e. fewer and shorter fixations on the blank space than for the other two positions) when the noun is preceded by a high-frequency three-letter word that is not an article, but not when the noun is preceded by an article.

Finally, there was another reason why we wanted to compare the eye movement behavior on an article versus a high-frequency three-letter word. Another indication of the special status of articles comes from research on the phenomenon of word skipping by Gautier, O'Regan, and LeGargasson (2000; see also O'Regan, 1979). They observed that in French the plural article '*les*' was skipped more often than a three-letter verb. This finding was interpreted as an indication that the probability of skipping a word can be influenced by the processing ease of that word. Robust influences of both a low-level visual nature and a high-level linguistic nature have been shown to affect skipping behavior (Brysbaert, Drieghe & Vitu, 2005; Rayner, 1998). Examples of these influences are the effect of word length as a low-level visual influence (readers tend to skip short words more often than long words, Rayner & McConkie, 1976) and predictability as a high-level linguistic influence (a word that is predictable from the preceding context is skipped more often than a word that is not predictable, Drieghe, Rayner, & Pollatsek, 2005; Ehrlich & Rayner, 1981; Rayner & Well, 1996).

Unfortunately, Gautier et al. (2000) did not report the actual skipping rates; their claim that the article *les* was skipped more than a three-letter verb was based on the observation that the landing distribution was shifted about 1.5 letters further to the right in the case of an article. The present experiment allows us to examine the size of skipping effects when an article is directly compared with a word that is also three letters long and is also easy to process. Moreover, we created a third condition in which the noun was preceded by a high frequency five-letter word (*other soldiers*). The addition of this latter condition allowed us to observe the effects of both syntactic category and word length on word skipping in the same experiment.

METHOD

Participants. Thirty members of the University of Massachusetts community participated; all were native speakers of English and had 20/20 vision or soft contact lenses. They were either given extra course credit or paid $10 for their participation.

Apparatus. Participants were seated 61 cm from a 15-inch NEC MultiSync FGE color monitor. Sentences were displayed on a single line with a maximum length of 80 characters. At this distance, 3.8 characters equaled 1 degree of visual angle. Eye movements were recorded using a Fourward Technologies Dual Purkinje Eyetracker (Generation V) interfaced with a Pentium computer. Although reading took place binocularly, eye movements were recorded only from the right eye (sampling every millisecond).

Materials. Sixty sentence frames were created so that the prototypical sentence featured a succession of the following three words: a verb[1], the article *the,* and a noun which was at least five letters long. In the two other conditions, the article was replaced either by a high-frequency three-letter word or a high-frequency five-letter word (see Appendix). The three-letter words were *two*, *one*, *old*, *his*, *her*, *new* or *all*. The mean frequency of these words was 2611 counts/million (Francis & Kuĉera, 1982). The five-letter words were *three, every, great, small, other,* or *dirty*, with a mean frequency of 868 counts/million. Even though the frequency of the three- and five-letter words was less than *the* (69971 counts/million), all of these words are very high frequency words. For the sake of convenience, we will refer to the word preceding the noun (the article, the three-letter high-frequency word, or the five-letter high-frequency word) as $word_n$ and the noun as $word_{n+1}$. Examples of the three conditions are shown in Table 1. A counterbalanced design was employed in which each of the 60 sentence frames was read only once by each participant, resulting in 20 sentences per condition per participant. The 60 experimental sentences were embedded in a pseudo-random order within a set containing 100 filler sentences and were preceded by 10 practice sentences.

INSERT TABLE 1 ABOUT HERE

Procedure. When a participant arrived for the experiment, a bite bar was prepared, which served to eliminate head movements. Participants were asked to read sentences on the monitor as their eye movements were monitored. They were told that they would be asked questions about the sentences and were instructed to read for comprehension. The initial calibration of the eye-tracking system required about five minutes. Prior to the presentation of each sentence, a series of five boxes appeared on the monitor, extending from the first to the last character position of an 80-character sentence. During this calibration check, participants looked at each box so that the experimenter could verify that the eye position was accurately recorded. If the calibration was not accurate, the participant was recalibrated. If the calibration was accurate, the participant looked at the first box and the experimenter displayed the sentence. Questions about the meaning of the sentence were asked after 25% of the trials and participants had little difficulty answering the questions (overall accuracy was 96%). The experiment lasted about 40 minutes.

RESULTS

Fixation durations of less than 80 ms or more than 1200 ms were removed from the analyses as were trials on which the eye-tracker lost track of the eye position. Thus, 2.2% of the trials were excluded; these trials were approximately equally distributed across conditions. A series of repeated measures analyses of variance (ANOVAs) were undertaken with participants (F1) and items (F2) as random variables.

Eye movement behavior preceding the critical region. To make sure our manipulation of $word_n$ did not affect the eye movement behavior prior to the saccade entering the region consisting of $word_n$ and $word_{n+1}$, we first examined the launch site for these saccades and their latency. Because 95.7% of these saccades originated from the verb preceding the region, we also report the gaze duration on the verb. As can be clearly seen from the values reported in Table 2, there were no significant differences between conditions for any of these measurements (all Fs < 1).

INSERT TABLE 2 ABOUT HERE

Skipping probability of $word_n$. The skipping probabilities are shown in Table 3. There was a significant effect due to word type [$F1(2,58)=72.11$, $p<.001$; $F2(2,118)=153.32$, $p<.001$]. The article was skipped 16% more often than the three-letter word, which in turn was skipped 28% more often than the five-letter word. All three differences were significant [article vs. three-letter word: $t1(29)=4.79$, $p< .001$; $t2(59)=6.25$, $p<.001$; article vs. five-letter word: $t1(29)=11.03$, $p<.001$; $t2(59)=15.46$, $p< .001$; three-letter word vs. five-letter word: $t1(29)=7.31$, $p<.001$; $t2(59)=12.17$, p<.001; all the reported *p* values for these contrasts and those reported below were Bonferroni adjusted][2].

INSERT TABLE 3 ABOUT HERE

Fixation times on word$_n$. Regardless of whether word$_n$ was an article, a high-frequency three-letter word, or a high-frequency five-letter word, it usually received only one fixation when it was fixated (respectively 99%, 97% and 94% of the time). Therefore we only report the *single fixation durations*[3] for word$_n$ (see Table 3). Mean single fixation duration showed a significant effect of the manipulation of word$_n$ [$F1(2,54)=5.20$, $p<.01$; $F2(2,116)=4.42$, $p<.05$]. Contrasts showed that although the pattern of results is what one would expect (fixation times on the article shortest and the 5-letter adjective longest), only the 38 ms difference between the article and the five-letter word was reliable [article vs. three-letter word: $t1(27)=-1.46$, $p>.20$; $t2(58)=-1.35$, $p>.20$; article vs. five-letter word : $t1(27)=-2.85$, $p<.05$; $t2(58)=-2.80$, $p<.05$; three-letter word vs. five-letter word : $t1(29)=-2.43$, $p>.05$; $t2(59)=-2.02$, $p>.10$].

Fixation times on word$_{n+1}$. The fixation times on word$_{n+1}$ are also shown in Table 3. There were no significant effects due to the manipulation of word$_n$ on first fixation duration [$F1(2,58)=2.37$, $p>.10$; $F2\ (2,118)=2.00$, $p>.10$]. However, there were significant effects on gaze duration [$F1(2,58)=10.54$, $p<.001$; $F2(2,118)=7.86$, $p<.001$]. This was mostly due to the mean gaze duration on word$_{n+1}$ being 28 ms shorter when word$_n$ was the five-letter word than when word$_n$ was the article or the three-letter word [article vs. three-letter word: all *t*s<1, n.s.; article vs. five-letter word: $t1(29)=3.22$, $p< .01$; $t2(59)=3.40$, $p<.01$; three-letter word vs. five-letter word: $t1(29)=4.90$, $p<.001$; $t2(59)=3.63$, $p<.01$].

Although this latter finding seems anomalous, a similar inverted word length "spillover" effect was reported by Pollatsek, Juhasz, Reichle, Machacek and Rayner (2008). They also observed longer gaze durations on a noun when preceded by a short adjective than when preceded by a long adjective, and modeled this finding by a combination of (a) less than optimal fixations on the noun after short adjectives due to targeting errors, (b) the shorter adjectives being more difficult to process because they have more orthographic neighbors, and (c) fixations prior to skipping being at less than optimal positions for extracting parafoveal information. The longer gaze durations on the noun after the high-frequency three-letter word and the article than after the five-letter word in the current study indicates that this effect is not likely to be restricted to adjective-noun sequences, as the word prior to the noun in the present study was frequently a determiner in the high-frequency three-letter word condition and was an article in the remaining condition.

First-pass time on the region consisting of $word_n$ *and* $word_{n+1}$. As seen in Table 3, there was a clear effect in the expected direction of the manipulation of $word_n$ on the first-pass times on the region consisting of both $word_n$ and $word_{n+1}$ [$F1(2,58)=64.26$, $p<.001$; $F2(2,118)=53.98$, $p<.001$]. The region containing the article was read 31 ms faster than the region containing the three-letter word, which in turn was read 70 ms faster than the region containing the five-letter word [article vs. three-letter word: $t1(29)=-4.35$, $p<.001$; $t2(59)=-2.89$, $p<.05$; article vs. five-letter word: $t1(29)=-9.38$, $p<.001$; $t2(59)=-11.07$, $p<.001$; three-letter word vs. five-letter word: $t1(29)=-7.62$, $p< .001$; $t2(59)=-6.73$, $p< .001$].

Distribution of first fixation locations on the region consisting of $word_n$ *and* $word_{n+1}$. In these distribution analyses, we focus on a comparison of the article with the other three-letter word condition (see Figure 1). Consistent with the skipping data, there was a shift to the right in the article condition versus the three-letter word condition. The shift to the right even occurred when analyzing only those instances when $word_n$ was skipped: the average landing position on $word_{n+1}$ after skipping $word_n$ was 1.7 (in character positions of $word_{n+1}$) in the article condition, and 1.1 in the three-letter word condition [$F1(1,28)=22.12$, $p<.001$; $F2(1,59)=26.80$, $p<.001$]). Although the distribution for the article condition does appear, at first glance, to be a single distribution with the mean located at the beginning of $word_{n+1}$, we think it can be explained otherwise. (We defer further discussion of this finding to the Discussion section.) Although the distribution of first fixation locations for the high-frequency three-letter word condition also appears to be a single normal distribution, the blank space between $word_n$ and $word_{n+1}$ is fixated less often than either the last letter of $word_n$ or the first letter of $word_{n+1}$. A comparison of the graphs in Figure 1 for the blank space and the surrounding letter positions indicates that the main difference is situated in the probability of landing on the last letter of $word_n$. A repeated measures ANOVA on the probabilities of landing on the last letter of $word_n$ and landing on the blank space preceding $word_{n+1}$ showed a significant main effect of word type; the eyes landed 6% more often on either of these two positions in the high-frequency three-letter word condition than in the article condition [$F1(1,29)=14.36$, $p<.001$; $F2(1,59)=12.78$, $p<.001$]. There was no main effect of letter position on the landing probabilities (all *F*s<1), but the crucial interaction between word type and letter position was significant, although only marginally in the participant analysis [$F1(1,29)=3.91$, $p=.058$; $F2(1,59)=4.72$, $p<.05$). We take these data as indicating separate landing-site distributions: one for saccades targeted at the high-frequency three-letter word and the other at the noun, although these two distributions overlap to a considerable extent – close to the point where this overlap obscures the bi-modality. Fortunately, a clearer picture of this bi-modality arises in the first fixation duration data.

INSERT FIGURE 1 ABOUT HERE

First fixation times on the blank space between word$_n$ and word$_{n+1}$. The mean first fixation times for the region consisting of $word_n$ and $word_{n+1}$, as a function of landing position, are shown in Figure 2. An analysis with items as a random variable carried out for each condition separately showed that in the condition in which $word_n$ was an article, the first fixation duration on the empty blank space was not significantly different from the first fixation duration on the last letter of $word_n$ and the first fixation duration on the first letter of $word_{n+1}$ [$F(2,42)<1$]. However, for the condition where $word_n$ was a three-letter word, there was a landing position effect on first fixation duration [$F(2,68)=4.94$, $p<.01$]. The first fixation duration in the region was 18 ms shorter when it was on the blank space than when it was on the last letter of $word_n$ [$t(42)=2.84$, $p<.01$] and 12 ms shorter than when it was on the first letter of $word_{n+1}$ [$t(41)=2.18$, $p<.05$]. The difference between the last letter of $word_n$ and on the first letter of $word_{n+1}$ was not significant [$t(39)<1$].

INSERT FIGURE 2 ABOUT HERE

DISCUSSION

Radach (1996) observed that the landing site distribution of a three-letter word followed by a five to seven letter word appeared to form a single normal distribution[4]. He hypothesized that short word groups could form a perceptual unit during reading. However, from his analysis, it was not clear whether this effect was largely or exclusively due to the fact that a large percentage of his three-letter words were articles. To examine this hypothesis we compared the landing distribution and the duration of the first fixations on a region consisting of either an article-noun pair or a high-frequency three-letter word followed by a noun. When the noun was preceded by a high-frequency three-letter word, we observed that the blank space between the three-letter word and the noun received fewer and shorter first fixations than did the letters immediately preceding and following the blank space (see also McConkie et al., 1988). However, we did not observe these patterns when the noun was preceded by the article *the*: The landing distributions seemed to make up a single distribution with its peak near the beginning of the noun.

At this point, it would be tempting to assume a word grouping phenomenon which is restricted to article-noun pairs. While we cannot rule out this possibility, we consider a strategy of targeting the first fixation on the group in order to be able to process it as a whole to be quite implausible. That is, if the identity of the article has already been established up to the point where it can be sufficiently distinguished from other high-frequency three-letter words, why would one still need to process the article-noun pair as a group? Among other things, this would leave the reader in a worse position to process the noun than if the article were skipped.

As previously mentioned, we thought it was plausible that a unimodal distribution of landing positions on the two word region could result from a mixture of two distributions: (a) saccades targeted to $word_n$ and (b) saccades targeted to $word_{n+1}$ (i.e., intended skips of $word_n$; we use the word "intended" because not all saccades land on their intended target, e.g. Drieghe, Rayner, & Pollatsek, 2008). Moreover, we suggested that if we found a somewhat different pattern for articles and the high-frequency three-letter words, the difference in the distributions may be explained by the percentage of intended skips in the two cases. To assess this possibility, we attempted to model the distributions for the article and the high-frequency three-letter word condition as mixtures of the two distributions indicated above, with the only difference in the two conditions being the percentage of time the initial fixation on the region was an intended skip. Assuming that fixations targeted to $word_n$, on average, landed on the middle of it and that fixations targeted to $word_{n+1}$, on average, landed a bit towards the beginning of it[5], we could model quite successfully both distributions assuming the only difference between the conditions was the probability of an intended skip of $word_n$ (see Table 4). The assumption of a single distribution of landing sites on the region, resulted in an inferior fit (for the article condition Root Mean Squared Deviation=0.018 for the mixture model and 0.037 for the single distribution model; for the high-frequency three-letter word condition, RMSD=0.017 for the mixture model and 0.027 for the single distribution model).

INSERT TABLE 4 ABOUT HERE

A few comments are in order. First, we should make clear that we do not view this as a serious modeling effort in the sense that we were not modeling the saccade process and, in particular, were not taking into account the varying launch sites from which a saccade into this area could be launched. Nonetheless, the parameters for the assumed distributions are not unreasonable, especially if one considers that attempted skips are generally fairly long programmed saccades and, as such, would tend to fall short of their intended target which is presumably the center of the noun (McConkie et al., 1988). Second, one object lesson is clear from the modeling effort. The most difficult thing to achieve is something other than a resulting unimodal distribution; getting bimodality requires fairly small standard deviations and/or quite widely spaced distributions. Thus, the conclusion that a unimodal distribution over a two-word region implies that the region is being treated as a perceptual unit should be treated with great caution. Third, we did not attempt to model the fixation duration data. However, the difference between the conditions in the fixation duration on the space between the words seems plausible. That is, one way to characterize the pattern of fixation durations in Figure 2 is that the difference between the two curves gets smaller, the further to the right one goes (with the exception of the pair of points at the −3 location). This makes sense if one assumes: (a) that fixations on which $word_n$ was the target are appreciably shorter for the article than for the high-frequency 3-letter word, and (b) that the further to the right one is in the figure, the higher the probability that the word targeted (and hence being processed) is $word_{n+1}$. If so, then fixation duration differences should get smaller, the further to the right the initial fixation in the region is. (The exception to this pattern, at location −3, may be because some of these fixations were intended for the word prior to $word_n$ and thus would dilute the difference between the conditions.)

The current research also offered a good opportunity to examine the phenomenon of word skipping during reading, and more specifically the skipping of articles. Whereas previous research (Gautier et al., 2000) had confirmed that an article is skipped more frequently than a verb, it was of interest to compare the skipping rate of an article with another relatively easy to process three-letter modifier. Also, Gautier et al. based their claim that the article *les* was skipped more often than a three-letter verb on the observation that the landing distribution was shifted about 1.5 letters further to the right in the case of an article; the actual skipping rates were not reported. We observed that the article was skipped 16% more often than the high-frequency three-letter word and that the high-frequency three-letter word was skipped 28% more often than the high-frequency five-letter word. The fact that word length is a strong influence on word skipping is in line with previous findings (Brysbaert et al., 2005). Whereas we undoubtedly also tapped into some existing differences between the article and the high-frequency three-letter word on the level of frequency and predictability, the fact that this 16% difference is the largest reported effect of a linguistic nature when controlling for word length very strongly points toward an extra effect of syntactic category on the observed skipping rates (for a discussion on effect sizes in skipping see, Drieghe, Desmet & Brysbaert, 2007).

In summary, in an experiment with tightly controlled stimuli, we were able to replicate Radach's (1996) finding of a single normal distribution of initial landing sites on two adjacent words when the first word was a three-letter word. However, we were able to show that this finding is restricted to those instances when the first word was an article, and it appears that the patterns of data in both the article and the high-frequency three-letter word condition can be explained by assuming that they each result from mixtures of two distributions (saccades intended for $word_n$ and saccades intended for $word_{n+1}$). The results of the current study appear to be compatible with the assumption that lexical processing in normal reading proceeds in a serial, word-by-word manner, and thus with models such as E-Z Reader (Pollatsek et al., 2006; Reichle et al., 1998).

References

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.

Drieghe, D., Desmet, T., & Brysbaert, M. (2007). How important are linguistic factors in word skipping during reading. British Journal of Psychology, 98, 157-171.

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Word skipping during reading revisited. Journal of Experimental Psychology: Human Perception and Performance, 31, 954-969.

Drieghe, D., Rayner, K., & Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. Quarterly Journal of Experimental Psychology, in press.

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

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. Psychological Review, 112, 777-813.

Francis, W., & Kuĉera, H. (1982). Frequency Analysis of English Usage: Lexicon and Grammar. Boston: Houghton Mifflin.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178-210.

Gautier, V., O'Regan, J. K., & Le Gargasson, J. F. (2000). 'The-skipping' revisited in French: programming saccades to skip the article 'les'. Vision Research, 40, 2571-2531.

Kliegl, R. (2007). Toward a perceptual-span theory of distributed processing in reading: A reply to Rayner, Pollatsek, Drieghe, Slattery and Reichle (2007). Journal of Experimental Psychology: General, 136, 530-537.

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.

O'Regan, J.K. (1979). Eye guidance in reading: Evidence for the linguistic control hypothesis. Perception & Psychophysics, 25, 501-509.

Pollatsek A., Juhasz, B. J., Reichle, E. D., Machacek, D., & Rayner, K. (2008). Immediate and delayed effects of word frequency and word length on eye movements in reading: A reversed delayed effect of word length. Journal of Experimental Psychology: Human Perception and Performance.

Pollatsek, A., Reichle, E.D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. Cognitive Psychology, 52, 1-56.

Radach, R. (1996). Blickbewegungen beim Lesen: Psychologische Aspekte der Determination von Fixationspositionen (Eye Movements in Reading). Waxmann: Münster/New York.

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. Perception, 8, 21–30.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124, 372–422.

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? Vision Research, 16, 829-837.

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., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. Psychological Review, 105, 125-157.

Reilly, R., & Radach, R. (2006). Some empirical tests of an interactive model of eye movement control in reading. Cognitive Systems Research, 7, 34-55.

Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun-noun compounds. Journal of Experimental Psychology: Learning, Memory, and Cognition, 33, 1162-1169.

Vitu, F., O'Regan, J. K., & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous texts. Perception & Psychophysics, 47, 583-600.

Author Note

This research was supported by National Institute of Health Grant HD26765 and was carried out when the first author was a visitor at the Eye Movement Lab at the University of Massachusetts, Amherst. Denis Drieghe is a postdoctoral fellow of the Fund for Scientific Research (Flanders, Belgium). His stay was funded by an extra scholarship provided by the Fund for Scientific Research (Flanders, Belgium). We thank Reinhold Kliegl, Wayne Murray, and an anonymous reviewer for helpful comments on earlier drafts.

Correspondence should be sent to Denis Drieghe (denis.drieghe@UGent.be).

Footnotes

1. To examine an unrelated research question, half of the verbs used in the experiment were obligatorily transitive verbs and the other half were optionally transitive verbs that are most likely to occur in an intransitive frame. Since this manipulation did not interact significantly in any way with the analyses reported in the current paper, the data we present are collapsed over both verb groups.

2. Because E-Z Reader states that a word is skipped by cancelling the saccade to the skipped word, it predicts that an inflated fixation duration precedes the skip. However, this issue is quite controversial (for a discussion see: Drieghe et al., 2005; Kliegl, 2007). Controlling for launch position by restricting the analysis to single fixations on the preceding verb, an inflated fixation duration was observed in the current experiment prior to skipping the article [277ms prior to skipping vs. 249ms prior to landing, $t1(25) = 2.42$, $p < .05$; $t2(55) = 3.26$, $p < .01$]. However, this effect was not observed in the three-letter word condition (283ms vs. 278s, all *t*'s < 1) and appeared to be going in the opposite direction in the five-letter word condition although this latter analysis lacks power due to relatively rare skipping of the five-letter word [265 ms vs. 285ms, $t1(18) = -1.60$, $p > .10$; $t2(49) = -1.84$, $p > .05$].

3. Single fixation duration is the duration of readers' fixations on the word on those trials on which the word was fixated exactly once. The results for first fixation duration and gaze duration were virtually identical to the single fixation data.

4. Of the 60 nouns used in the current experiment, 40 fell within the 5 to 7 letter-word range used by Radach (1996). Removing the data of the other words, which were 8 to 11 1etters long, did not alter any of the observed patterns.

5. The mean of the landing sites for the saccades intended for $word_{n+1}$ (i.e., intended skips) was +1.5, which is the average first fixation location on a 6 letter-word coming from a launch site located 7 character positions away (i.e. in the current experiment 3 character positions preceding the blank space in front of the

3 letter-word modifier), as observed by McConkie et al. (1988) and implemented in the E-Z Reader model (Pollatsek et al., 2006).

Table 1. Example sentences from the experiment illustrating each of the 3 conditions.

1. The article “the”
The general placed the soldiers on the battlefield before the enemy arrived.
2. High frequency three-letter word
The general placed all soldiers on the battlefield before the enemy arrived.
3. High frequency five-letter word
The general placed other soldiers on the battlefield before the enemy arrived.

Table 2: Launchsite (LS) and Saccade Latency (SL) of the saccade going into the region consisting of both word$_n$ and word$_{n+1}$ (in characters spaces to the blank space preceding word$_n$), and gaze duration (GD) on the verb preceding the word group as a function of the manipulation of word$_n$.

Word$_n$	**Start of Saccade**		**Word$_{n-1}$**
	LS	SL (ms)	GD (ms)
the	4.4	265	305
two	4.4	269	303
other	4.3	269	309

Table 3: Skipping probabilities (SP) and single fixation duration (SF) of word$_n$, first fixation duration (FF) and gaze duration (GD) of word$_{n+1}$ and first-pass time (FP) of the region consisting of both word$_n$ and word$_{n+1}$ as a function of the manipulation of word$_n$.

Word$_n$ Condition	Region of Interest				
	Word$_n$		Word$_{n+1}$		Word$_n$ + Word$_{n+1}$
	SP	SF (ms)	FF (ms)	GD (ms)	FP (ms)
the	.66	248	273	317	414
two	.50	264	267	322	445
other	.22	286	261	292	515

Table 4: Observed and predicted probabilities of landing on letter locations -3 to +3 for the article and the high-frequency three-letter word condition. Letter position 0 is the blank space between word$_n$ and word$_{n+1}$.

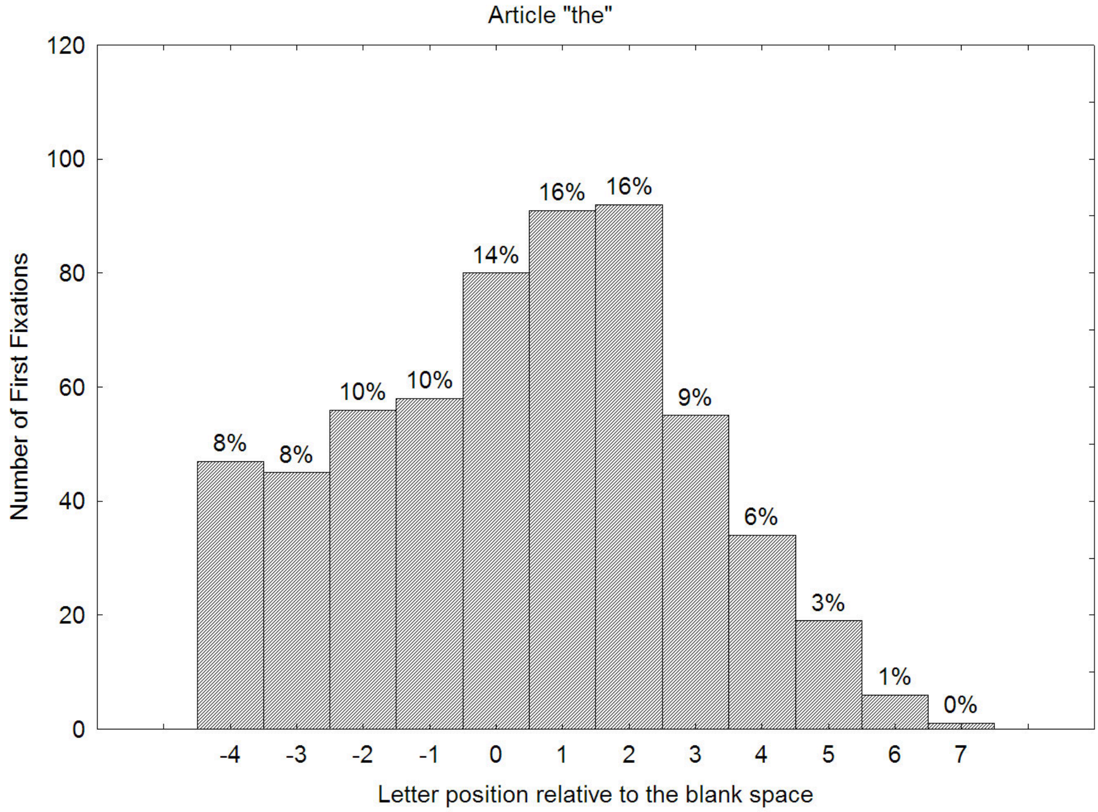
	Landing Position						
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**
				Article			
Observed Probability	**0.09**	**0.12**	**0.13**	**0.17**	**0.19**	**0.19**	**0.12**
Predicted Probability by Mixture Model	**0.07**	**0.13**	**0.15**	**0.16**	**0.21**	**0.19**	**0.10**
Predicted Probability by Single Distribution	**0.10**	**0.14**	**0.18**	**0.19**	**0.17**	**0.13**	**0.09**
				High-frequency three-letter word			
Observed Probability	**0.10**	**0.16**	**0.21**	**0.18**	**0.19**	**0.11**	**0.05**
Predicted Probability by Mixture Model	**0.11**	**0.17**	**0.22**	**0.19**	**0.20**	**0.12**	**0.05**
Predicted Probability by Single Distribution	**0.13**	**0.19**	**0.21**	**0.19**	**0.14**	**0.09**	**0.04**

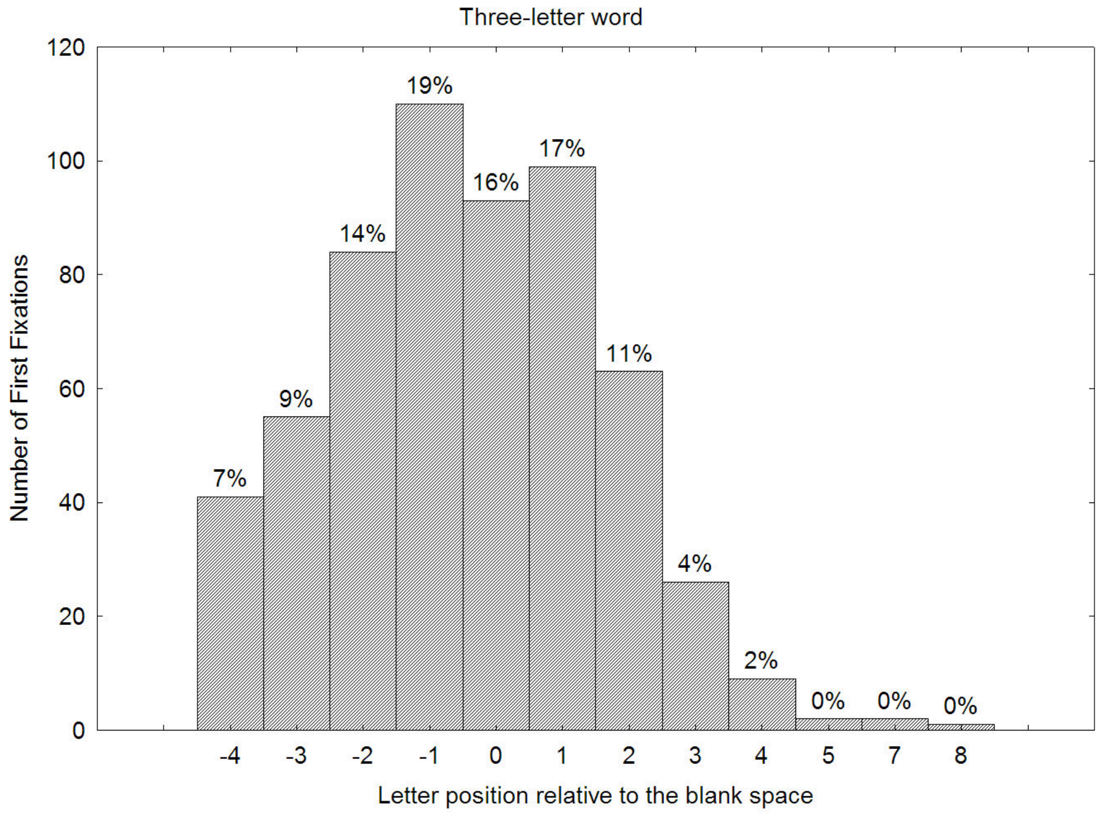
Note: The distributions assumed for the mixture model in both the article and the high-frequency three-letter word conditions were normal, with standard deviations equal to 1.3 character spaces. The mean assumed (for both conditions) for the distribution of fixations intended for word$_n$ was -1.5 and the mean for the saccades?fixations intended for word$_{n+1}$ (i.e., intended skips) was +1.5. The probability of an intended skip was assumed to be 0.5 for the high-frequency three-letter word condition and 0.6 for the article condition. The single distribution histograms were computed by fitting a normal distribution for each condition over the entire region containing the three-word modifier and the noun.

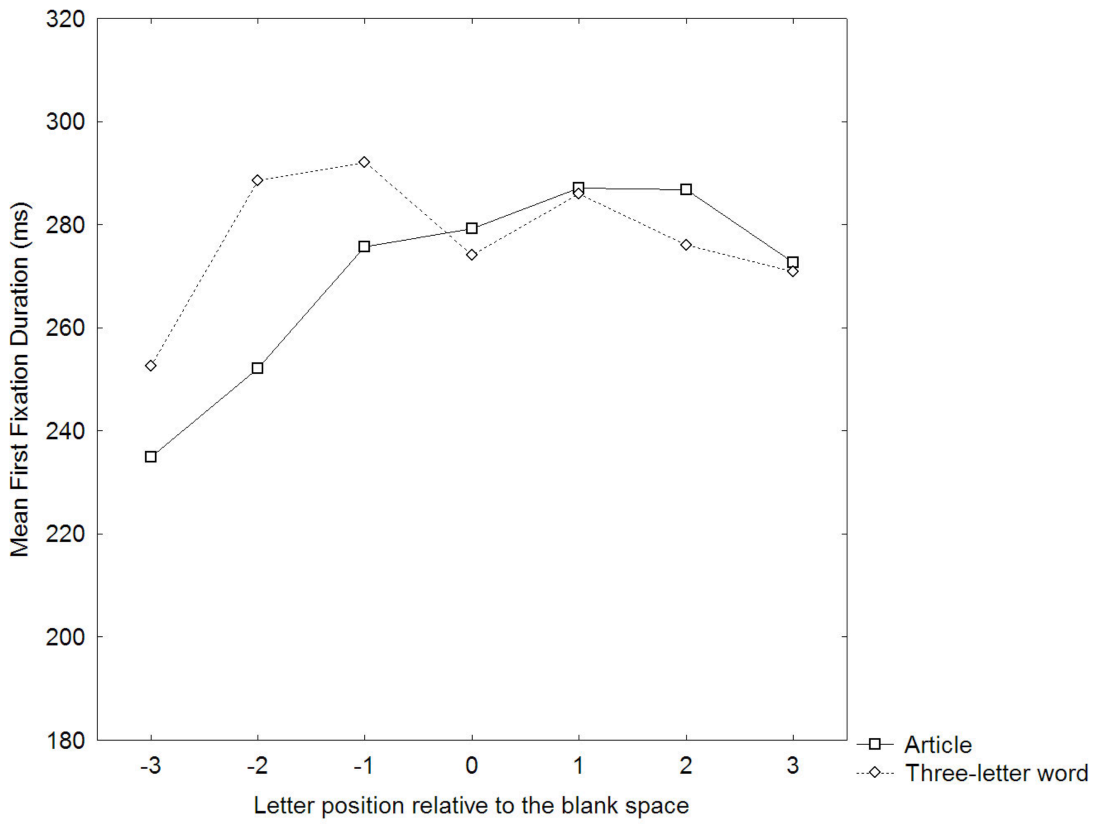
Figure captions

Figure 1. Distribution of the landing position of the first fixation in the region consisting of $word_n$ and $word_{n+1}$. Position 0 stands for the blank space between $word_n$ and $word_{n+1}$. Position -4 stands for the blank space preceding $word_n$. In the top graph $word_n$ = article (total N = 584), in the bottom graph $word_{n+1}$ = high-frequency three-letter word (total N = 585).

Figure 2. Mean first fixation durations in the region consisting of $word_n$ and $word_{n+1}$ as a function of landing position. Position 0 stands for the blank space between $word_n$ and $word_{n+1}$.


Article "the"
Number of First Fixations
120
100
80
60
40
20
0
8%
8%
10%
10%
14%
16%
16%
9%
6%
3%
1%
0%
-4
-3
-2
-1
0
1
2
3
4
5
6
7
Letter position relative to the blank space


Three-letter word
Number of First Fixations
120
100
80
60
40
20
0
7%
9%
14%
19%
16%
17%
11%
4%
2%
0%
0%
0%
-4
-3
-2
-1
0
1
2
3
4
5
7
8
Letter position relative to the blank space


Mean First Fixation Duration (ms)
320
300
280
260
240
220
200
180
-3
-2
-1
0
1
2
3
Letter position relative to the blank space
Article
Three-letter word

Appendix.

Materials used in the experiment. Each sentence contains either the article *the*, a high-frequency three-letter word or a high-frequency five-letter word. Participants read each sentence containing one of the three words.

1. After lunch the officer arrested the|two|other terrorists responsible for the bombing.
2. The politician annoyed the|two|three liberals with his extremist right-wing views.
3. The sound terrified the|one|every inhabitant of the house.
4. The attack prompted the|all|other diplomats to withdraw from the peace conference.
5. The architect designed the|two|other houses in the town but not the museum.
6. The divorce settlement included the|all|other furniture as well.
7. The sailors persuaded the|two|other pirates that their ship was not worth robbing.
8. Billy persuaded the|two|three teenagers not to use a forged license to get into a bar.
9. The popular king included the|two|three peasants in the wedding celebration.
10. The math teacher praised the|two|three students for their cooperation during class.
11. She imitated the|two|three tricks she saw at the circus with little difficulty.
12. Joe carried the|old|dirty garbage to the dumpsters while wishing he found another job.
13. George coaxed the|his|every puppy to come inside.
14. The crowd destroyed the|one|every statue that they thought was blasphemous.
15. The farmer bought the|one|every donkey for use on his farm.
16. The conqueror guarded the|one|every native who lived on the island.
17. The janitors blocked the|all|other spots that were slippery in the hallway.
18. The nervous child carefully carried the|two|three rabbits into their pen.
19. The officials regretfully blocked the|her|every attempt to approve the budget.
20. The pianist carelessly bothered the|his|every conductor with his silly demands.
21. The trainer bought the|two|three greyhounds for training them to run in the races.
22. The delivery boy lifted the|two|three pizzas onto the table and went home.
23. The businessman hired the|one|every trainer from the downtown gym.
24. The professor invited the|two|great lecturers from Mexico to give a talk.
25. The teacher always corrected the|all|small errors the students made during class.
26. The teacher occasionally included the|all|other students at the back of the class.
27. The actor described the|his|every monologue from the play without much enthusiasm.

28. The captain accurately mimicked the|two|other excuses of the team members.
29. Joey sold the|his|every carpet which had this rather annoying red color.
30. The general placed the|all|other soldiers on the battlefield before the enemy arrived.
31. Linda talked the|two|other visitors into buying some souvenirs.
32. The board decided the|two|three issues without any difficulty.
33. John believed the|two|other stories the old man told in the bar.
34. He shrunk the|new|dirty pieces of clothing by washing them in too hot water.
35. James danced the|one|every tango which he knew the judges would like.
36. The boy rushed the|all|other documents to the director for his approval.
37. The cold hardened the|two|three pastries much quicker than intended.
38. The painter stood the|two|three ladders against the wall.
39. The bad mixture shrunk the|two|three loaves to a distasteful looking puddle.
40. The airline flew the|one|every charter flight because it was cheaper.
41. The cowboys raced the|two|three horses back to the ranch.
42. He leaned the|all|three boxes against the wall so they wouldn't collapse.
43. The ice skaters attempted the|new|other routines that nobody else could do.
44. The admiral sailed the|two|three boats to India.
45. The driver crashed the|all|three automobiles which were assigned to him for testing.
46. The writer continued the|his|three books he was working on after his holiday.
47. The chairman sat the|all|other committee members around the table.
48. The trainer rested the|one|every boxer before the great fight.
49. The artist refused the|all|other prizes since they meant nothing to him.
50. The horse jumped the|two|three hurdles but tripped when trying to jump over the ditch.
51. The frightened sheep leaped the|two|three fences to get out of the barnyard.
52. The senators argued the|one|every point but the president had his way.
53. The mother hurried the|her|every child in order to get out of the store.
54. The virus mutated the|all|other cells causing cancer in many of the test animals.
55. The mother bird perched the|her|three chicks on the branch.
56. The current floated the|two|three sailors toward the bank.
57. The old man tired the|two|three nurses with his constant nagging.

58. The delivery boy rested the|all|other groceries on the table while he got out the bill.

59. The captain marched the|one|everysoldier toward the encampment.

60. The boy grew the|two|three plants for his biology assignment.