Preprint of:

Heister, J., Würzner, K., & Kliegl, R. (2012). Analysing large datasets of eye movements during reading. In James S. Adelman (Ed.), *Visual word recognition, volume 2: Meaning and context, individuals and development*. (pp. 102-131). Hove, England: Psychology Press, 2012.

Analysing Large Datasets of Eye Movements during Reading

Julian Heister, Kay-Michael Würzner, and Reinhold Kliegl

University of Potsdam

Chapter to appear in**:**

James S. Adelman (ed.): Visual Word Recognition

Main body of text (w/o Summary): (8.332) words

Address for correspondence:

Reinhold Kliegl, Department of Psychology, University of Potsdam

Karl-Liebknecht-Str. 24-25, 14476 Potsdam

Tel. + 49 331 977 2868, Fax 49 331 977 2973

Email: kliegl@uni-potsdam.de

Analysing Large Datasets of Eye Movements During Reading

Most of the time visual word recognition occurs in the context of reading. Eye movements and fixations provide necessary perceptual and language-related information; prior knowledge and sentence context guide our expectations about this input. Obviously, not only the *process* of word recognition during reading, but also *research* on this topic is a highly complex undertaking. One very successful approach to deal with this complexity has been to study the recognition of isolated words, often measured in recognition times. Typically, subjects and words are assigned to each other in some form of a (counter-)balanced design. The ensemble of chapters in this volume summarizes the state of the art.

In this chapter, we take an alternative starting point: fixation durations during continuous reading of sentences. We review research based on an analysis of a large corpus of fixation durations which allows the simultaneous consideration of a large number of influences. Such corpus-based reading research departs strongly from the priniciples of orthogonal experimental design, which is the aim of research that draws inferences from measurements of one to three target words (see, e.g., Schotter and Rayner, this volume). In the following two sections, we describe conceptual and statistical frameworks for the analyses of eye movements that use fixations on (almost) all words. Within these frameworks, we build on earlier reports about the „big three" word factors influencing fixation durations (type frequency, predictability from prior sentence context, and word length). We carry out these analyses from a perspective of distributed processing, taking into consideration not only the properties of the fixated word, but also those of its left and right neighbors (Kliegl, Nuthmann, & Engbert, 2006; Kliegl, 2007). We expand the previous research by including not only the properties of the word triplet but those of the word quintet centered at the fixated word.

A Conceptual Framework for the Analyses of Fixation Durations in Reading

Visual word recognition occurs in the context of reading involving eye movements and fixations in the service of providing the necessary perceptual and language-related information. We distinguish roughly three types of effects: (1) Effects that arise from the oculomotor process and low-level processes of eye guidance leading to effects related to preferred viewing locations or launch sites. (2) Effects that arise from language-related processes. Many of the over 50 linguistic properties of words contributing to processing efficiency in isolated word recognition have also been established for fixation durations or probabilities in normal reading. In addition, during reading of sentences, fixation durations and probabilities are influenced by variables coding the context of words, such as the predictability of words from prior words of the sentence or corpus-based statistics such as transition probabilities or context diversity. (3) Finally, corpus analyses have provided reliable evidence that some indicators of language-related processes exert their effects during at least three successive fixations.

A three-dimensional taxonomy of fixations

The ocular dynamics of reading can be cast along three dimensions, comprising the *number of fixations* on a word, the *direction of saccades* bordering these fixations, and the *duration of fixations*. In Figure 1, we display fixation patterns for three successive fixations defined by the first two of these three dimensions (see Hogaboam, 1982, for an earlier taxonomy). The columns depict single, first, or second of two fixations. This covers almost all fixations, because very few words host more than two fixations during reading. The rows differentiate between eight different patterns of foward and backward movements. The figure panels inform about the absolute number and percentages of the corresponding fixations. The most frequent event are single fixations preceded and followed by a forward saccade (32%) and fixations preceding a forward skipping (12%). Regressions out of and into words (9% and 8% respectively) are much more common than regressions within words (2%).

Single fixations account for more than half (57%) of all fixation patterns, whereas the second and third column add up to just about 20%. Single fixations entering the analyses in the next section represent a subset from 125 515 single fixations in the highlighted panels of Figure 1.[1] Note, that the panels in Figure 1 do not distinguish between first and second pass reading (see below). Analyses on gaze durations (the sum of all fixations on a word prior to movement to another word) actually include more single-fixation than double-fixation cases, even if we include cases with three or more fixations (Rayner, 1998).

-- Insert Figure 1 about here --

[1] The main analysis reported in this chapter includes a total of 80,625 firstpass single-fixation durations, with durations ranging from 18 to 1.372 ms. Only 139 fixations were shorter than 50 ms and only 66 were longer than 750 ms—two typical lower and upper bounds used in other research, but an analyses of model residuals strongly suggested keeping all fixations.

The statistics in Figure 1 are based on 223.099 fixations, however 11.127 fixations (5%) fell into different fixation patterns (mostly three or more fixations on one word) leaving 211.972 fixations.[2] They were recorded from 273 subjects reading 144 sentences of the first Potsdam Sentence Corpus (PSC I, Kliegl, Grabner, Rolfs, & Engbert, 2004). Sentences with blinks or any loss of measurement were deleted, leaving 90% of the sentences for the analyses. Obviously, the presumably random loss of sentences varied across subjects, leaving 42 to 144 sentences across subjects. Conversely, sentences randomly differ in how many subjects contributed fixations, ranging from 229 to 257 subjects. Therefore, from the outset of analyses, the initial crossing of subject and sentence factors is lost; the design ends up being highly imbalanced.

Since we include (almost) all words (i.e., not only selected target words) into the analyses, we must also distinguish between variance that is due to differences between sentences and variance that is due to differences between words. Obviously, words are not crossed with sentences, but only partially crossed at best. Fortunately, statistical programs which estimate the variance components of such sparse designs in linear mixed models (LMMs) for continuous dependent variables (such as fixation durations) or generalized linear mixed models (GLMMs) for binary dependent variables (e.g., skipping vs. fixating a word) have become available during recent years. We use Bates and Maechler's (2010) *lme4* package in the R environment for statistical computing and graphics (R Development Core Team, 2010) for our analyses.

A descriptive statistical model for single-fixation durations

At the level of the eyes, reading consists of an alternating sequence of fixations and saccades with information uptake during fixations. Most saccades are forward saccades to the next word (50%), but in about 25% of the cases the next word is skipped. For the remainder of this chapter, we focus on durations of fixations that were the only fixation during firstpass reading, that is fixations which were preceded by a saccade from a previous word and followed by a saccade to a subsequent word that had not been fixated or skipped before. Such a fixation is called a single fixation duration (SFD; see Figure 1, first column; Figure 2, top part). This definition covers rougly 50% of all fixations during leisurely reading for comprehension (see Figure 1, also e.g., Kliegl et al., 2006).

[2] 70.892 fixations on first and last words as well as 67 fixations below 15 or above 1500 milliseconds were deleted beforehand. This leads to different percentages of first and second fixations in two fixation cases.

-- Insert Figure 2 about here --

Immediate lexical effects. Past research has shown that fixation durations and skipping are influenced by word frequency, word length and predictability (Rayner & McConckie, 1976; Kliegl et al., 2004; Brysbaert, Drieghe, & Vitu, 2005). Assuming *immediacy of processing*, properties of the fixated word are the primary determinants of fixation and gaze durations (Just and Carpenter, 1980). Together with the dominant role of frequency in psycholinguistic research, word length and a word's predictability from prior context occupy prominent roles in reading research (Rayner, 2009; see Rayner, 1998 for a review). Frequency, predictablity, and the inverse of length[3] correlate negatively with fixation duration and refixation probability and positively with skipping probability (Kliegl et al., 2004). Besides frequency, length, and predictability, a wide range of linguistic variables influencing lexical word identification have been distinguished by means of enhanced computational methods and large, easily accessible digital resources. Graf, Nagler and Jacobs (2005) identified effects of 57 linguistic word properties contributing to processing efficiency in isolated word recognition. Influences of lexical variables derived from surface frequency like lemma (Brysbaert & New, 2009; Beauvillain, 1996), rank (Murray and Forster, 2004), and document frequency (e.g. contextual diversity, see Adelman, Brown, & Quesada, 2006) have been shown to influence reaction times in naming and lexical decision tasks.

Fovea and parafovea. In reading we pick up information from more than just the currently fixated word. Word N-1, word N+1, and possibly also word N+2 fall into the so-called parafoveal region extending from the foveal region to about 5 degrees on either side of fixation. Starting with McConkie and Rayner (1975), numerous experiments have shown that parafoveal visual properties covering the area from 4 characters to the left to a maximum of 15 characters to the right of the current fixation location influence fixation durations. Consequently, during reading of sentences, fixation durations and probabilities depend also on variables coding the context of words, such as the predictability or plausibility of words from prior words of the sentence or corpus-based statistics such as transition probabilities and surprisal measures (cf. Boston, Hale, Kliegl, Patil, & Vasishth, 2008). In

[3] We typically use the reciprocal of word length to counteract the positive skew of the word-length distribution in German. Using the reciprocal of word length, renders the multiplicative interaction of frequency and length or predictability and length as a ratio or relative frequency and predictability measure (i.e., normalized on word length).

Figure 2, we list type frequency, predictability (given prior words of the sentence) and length for the fixated word N and for words N-2, N-1, N+1, and N+2. These independent variables are included as grand-mean centered continuous measures (covariates) in our analyses. In keeping with prior specifications, frequencies are log-transformed, predictabilities are logit-transformed, and word lengths are entered with their reciprocal values (Kliegl et al., 2004, 2006).

Lag effects. Analyses based on a large set of eye movements (e.g., Kliegl et al., 2006) have shown reliable influences of frequency, length, and predictability of word N-1 on fixations measured on word N. Word N-1 usually falls outside the foveal region comprising 2 degrees in the centre of vision (i.e., about 5 letters, presuming a distance to the monitor of 60 cm). There are (at least) two possible explanations for the influence of properties of word N-1. First, prolonged fixation durations on word N may reflect incomplete processing of word N-1. This spillover or lag effect is reflected in longer fixation durations after long, low frequent, or unpredictable words. A second explanation follows from the foveal load hypothesis (Henderson and Ferreira, 1990) to word N-1: Longer fixations occur on word N after difficult (e.g., low frequency) words N-1 because the difficulty of word N-1 causes a narrowing of the attentional span which in turn reduced parafoveal preprocessing of word N. The interaction of frequency of word N-1 and N is in agreement with this view: The frequency effect on word N is more pronounced if word N-1 is a low frequency word (Kliegl et al., 2006; Rayner & Duffy, 1986).

Influences from upcoming words. In reading we pick up information from more than just the currently fixated word. Words N+1 and, depending on the length of word N+1, possibly also word N+2 fall into the parafoveal region extending from the foveal region to about 5 degrees in reading direction. Gaze-contingent masking of parafoveal text increases fixation durations (McConkie & Rayner, 1975; Rayner, 1975). Thus, efficient reading requires parafoveal preview of the upcoming words (e.g., Balota, Pollatsek, & Rayner, 1985; Binder, Pollatsek, & Rayner, 1999; McConkie & Rayner, 1975; Rayner, 1975; Rayner & Bertera, 1979; Underwood & McConkie, 1985). Which properties of upcoming words are extracted during preview is an active area of current research. Lexical effects of upcoming words on the currently fixated words are still controversial. In their corpus analyses, Kliegl et al. (2006) found an effect of word frequency and predictability of the word

N+1 on fixations on word N. There was also an interaction of frequency of word N+1 and length of word N with a reliable N+1-frequency effect only after short words N, replicating Kennedy and Pynte (2005). Kennedy and Pynte also reported that for long words N, the parafoveal effect was limited to the initial trigram informativeness of word N+1, defined as the number of words sharing its initial three letters. Interestingly, whereas the N+1-frequency effect is always in the canonical direction with high-frequency words N+1 leading to short fixations, high-predictability words N+1 prolong fixations on word N (despite a positive correlation of frequency and predictability or words). Kliegl et al. (2006) explain this with memory retrieval of the upcoming word with prior sentence context as retrieval cue. In a corpus analyses by Pynte & Kennedy (2006) the parafoveal frequency effect also appeared, but White (2008) reported no effect of parafoveal frequency with control for orthographic familiarity. We emphasize again that the N+1-frequency effect has been questioned and only Kennedy and Pynte (2005; also Pynte & Kennedy, 2006 and Kliegl et al., 2006) found an effect (but see Inhoff, Radach, Starr, & Greenberg, 2000; Kennedy, Pynte, & Ducrot, 2002; Rayner, White, Kambe, Miller, & Liversedge, 2003; Rayner & Juhasz, 2004). In contrast, effects of sublexical effects of orthographic familiarity of word N+1 are well established (Pynte, Kennedy, & Ducrot, 2004; Starr & Inhoff, 2004; Underwood, Binns, & Walker, 2000; but see Rayner, Juhasz, & Brown, 2007; White & Liversedge, 2004).

Polynomial trends. Some of the ambiguity of corpus-analytic results, such as those just described for N+1-frequency effects, may be due to nonlinearities underlying the relation between SFD and lexical predictors in multiple regression. In Figure 3, we illustrate the non-monotonicity of the function relating SFDs onf word N to the log of the frequencies of word N (middle), word N-1 (left), and word N+1 (right). The top row is based on the fixations of 273 readers of the first PSC (144 sentences); the bottom row is based on 144 different sentences of the second PSC (159 readers). The similarities of the profiles across different readers and different sentences suggests that the different non-linearities associated with the three different frequencies are reliable. Obviously, an explanation of such statistically reliable non-linear profiles represents the most formidable challenge for theoretical accounts of eye-movement control during reading, ideally validated with simulations in computational models.

-- Insert Figure 3 about here --

Skipping. Starr and Rayner (2001) list word skipping as one of the major areas of research in reading besides regressions and the question of distributed processing and processing of upcoming words. In experimental designs word length accounts for most of the variance in skipping probabilities. As the length of a word increases, the probability that it will be skipped decreases (Brysbaert, Drieghe, & Vitu, 2005; Rayner & McConkie, 1976; Rayner, Sereno, & Raney, 1996). In their meta-analysis Brysbaert and Vitu (1998), conclude that about 1/4 of the variance in skipping probabilities is determined by word length. Thus, 3 letter words are skipped about 67% of the time, whereas 7-8 letter words are skipped only about 20% of the time; Krügel and Engbert (2010) estimate that about 90% of two-character words would be skipped if fixation locations are corrected for mislocated fixations. While there is agreement on the fact that word length influences skipping probability, it remains controversial to which extent lexical and sub-lexical properties of the skipped word influence the fixation duration prior to a skipped word. In Pynte and Kennedy (2006) skipping probability increases for high-frequency words and words with an informative beginning (see also White, 2008). In their first experiment Pynte, Kennedy, and Ducrot (2004) revealed decreased skipping probabilities before words with misspelled first letters. In their second and third experiments, misspellings induced shorter first fixation and gaze durations on word N-1 and there was no longer an effect on skipping probability. Similar to word frequency, words are also skipped more often if they are highly predictable (Ehrlich & Rayner, 1981; Brysbaert, & Vitu, 1998; Kliegl et al., 2004; McConkie, et al., 1994; Rayner, Ashby, Pollatsek, & Reichle, 2004; Rayner & Well, 1996; Vitu, O'Regan, Inhoff, & Topolski, 1995). Kliegl and Engbert (2005) also examined the influence of skippings on prior fixation durations. They find that single fixation durations before skippings vary with the length and frequency of the to be skipped word. Fixations before skipped words are shorter before short or high-frequency words and longer before long or low-frequency words in comparison with control fixations.

Oculomotor variables. SFDs also depend on oculomotor factors. We include (a) skipping of word N-1 or skipping of word N+1 (no-skip coded with „0", skipping coded with „1"), (b) launch site (i.e., the letter distance between the last fixation location and the beginning of the fixated word) and outgoing saccade amplitudes, and (c) the relative position of the fixation in the word (using a quadratic term) in our model. Of course, as reviewed for skipping in the last paragraph, these variables also reflect the effects of lexical processing (e.g., skipping of three-letter determiners is more frequent than skipping of three-letter verbs, O'Reagan, 1979). Here we examine them not as dependent, but as independent variables of SFDs. In this context, oculomotor processes may not be of primary concern for word recognition *per se*. For example, there is no reliable effect of word frequency on landing positions (Rayner, Binder Ashby & Pollatsek, 2001). Including oculomotor variables in a regression model statistically removes the influence of low-level factors on SFDs from the residual error. This should increase our chances of detecting theoeretically small effects related to lexical processing. For example, given the rapid decline of visual acuity relative to the fixation position, effects of lexical processing of parafoveal words are expected to be small. If they can be established reliably, these effects are of much relevance for constraining not only models of eye-movement control during reading, but also models of visual word recognition.

Oculomotor influences: launch site and landing site. SFDs are influenced by two oculomotor factors, namely saccadic amplitude and landing site. Several experiments have demostrated the importance of parafoveal word properties by masking words outside the fovea (parafoveal masking) increasing subsequent fixation durations (e.g., Balota, Pollatsek, & Rayner, 1985; Binder, et al., 1999; McConkie & Rayner, 1975; Rayner, 1975; Rayner & Bertera, 1979; Underwood & McConkie, 1985). The further away, that is the longer an incoming saccade, the less preview is possible due to the drop-off in visual acuity and associated lateral inhibitions. Thus, the larger the amplitude of the incoming saccade, the longer the subsequent fixations on word N (Radach and Heller, 2000; Vitu et al., 2001; see also Heller & Müller, 1983; Pollatsek, Rayner, & Balota, 1986). In the same way, long words N-1 lead to less preview and, consequently, longer fixation durations on word N (Kliegl et al., 2006). Consequently, fixations after skippings are longer. Fixation duration also depends on the landing position within a word. The effect that fixation durations in the middle of words are longer than those at the edges has been called the inverted optimal viewing position effect (IOVP) (Vitu, McConkie, Kerr, & O'Regan, 2001; McConkie, Kerr, Reddix, Zola, & Jacobs, 1989). Nuthmann, Engbert, &

Kliegl (2005) argue that this effect is due to mislocated fixations and the immediate triggering of a new saccade program (see also Nuthmann, Engbert, & Kliegl, 2007).

Interactions. There are very strong interactions between lexical and oculomotor variables. Most importantly, the effect of lexical factors on SFDs depends strongly on whether word N-1 or word N+1 was skipped (e.g., Kliegl & Engbert, 2005; Kliegl, 2007). Below we will report analyses on such interactions and discuss their implications. Past research also established a number of reliable interactions among the lexical variables (e.g., Kennedy and Pynte, 2005; Kliegl et al., 2006). From this research we keep the following previously significant interactions in the model: (a) frequency of word N with frequency of word N-1, (b) frequency of word N with length of word N, (c) length of word N with frequency of word N+1, and (d) length of word N with predicability of word N+1.

Figure 2 represents the framework for possible influences on SFDs for the present analysis. As indicated by the ellipses, the framework may be expanded with other variables as their influence is established; this is not a closed set. Clearly, we are looking at a large number of variables simultaneously and this requires a defensible perspective for statistical inferences. We propose to analyze the relation of SFDs to the oculomotor and lexical variables with LMMs, covering at least the fixated word and its two left and right neighbors respectively and a selection of interactions between them. In the next section we review the key components of LMMs and how they relate to SFDs during reading.

A Linear Mixed Model of Single Fixation Durations during Reading

We report LMM results for the statistical model of SFDs represented in Figure 2, focussing on the largest effects. The complete output of the lmer function (Bates & Maechler, 2010) is provided in Appendix A. Data and R scripts will be made available at a project website (http://www.dlexDB.de or http://read.psych.uni-potsdam.de/pmr2/) and enable not only the replication of the present model, but also the pursuit of alternatives.

Synopsis of methodological advantages of LMM

LMMs are rapidly gaining acceptance in psycholinguistic experimental research (e.g., Baayen, Davidson, & Bates, 2008; Kliegl, Masson, & Richter, 2010). Baayen (2008), Faraway (2007), and Gelman and Hill (2007) contain chapters which offer a general introduction to LMMs with an applied perspective. We suspect that LMMs are likely to replace traditional analyses of variance for inferential statistics. Three advantages of LMMs are of special relevance for the present purpose.

First, they suffer from much less loss in statistical power incurred by an imbalance in the number of observations than, for example, repeated-measures ANOVA (e.g., Quené & van den Bergh, 2008, for simulations). In eye-movement research, we have no control over whether a word is skipped or fixated once or several times; we also have to accept blinks and loss of measurement during the experiment. Thus, any experimental design, nicely counterbalanced at the outset, will end up highly unbalanced after about two seconds of measurement.

Second, traditionally experiments are specified as factorial designs; continuous lexical indicators such as printed frequency are often forced into discrete categories of, e.g., low and high frequency. Alternatively, continuous predictors (such as log printed frequency) have been used in multiple regression analysis (MRA), but the clustering of fixations by subjects requires separate MRAs for each subject (i.e., repeated-measures MRA; Kliegl et al., 2006; Lorch & Myers, 1990), limiting the number of predictors in a model by the number of fixations measured per subject. LMMs overcome this distinction between ANOVA and MRA analyses. Effects of factors and covariates can be specified along with variances of these effects associated with random factors of subjects, sentences, and words.

Third, in psycholinguistic research we typically distinguish two random factors: subjects and items (sentences or words). The (presumably) random selection of subjects and items affords generalizability for these dimensions. As we want to generalize across subjects and items, we are usually not interested in how much a certain subject's, say, average fixation duration departs from the overall mean; neither are we interested in average fixation durations of individual items. What we do need to know for test-statistics, however, are the variances of subjects' and items' average fixation durations. Therefore, psycholinguistic experiments report two ANOVAs, one using subjects (F1 ANOVA) and one using items (F2 ANOVA) as random factors. LMMs, covering both random factors, are to be preferred for obvious reasons, such as the avoidance of ambiguities relating to significant F1 and F2 effects. More importantly, simultaneous estimates of fixed effects (i.e., effects analogous to unstandardized regression coefficients in multiple regression) and of between-subject, between-sentence, and between-word variance components (and correlation parameters) yield not only appropriate inferential test statistics for fixed effects, but offer additional insights in the dynamics of reading. In the next three sections we elaborate on this distinction between model parameters of the design matrix (fixed effect estimates), random factors, and parameters specifying variance components and correlation parameters associated with (some of) the fixed effects.

Fixed effects

In the last section, we introduced the lexical, visual, and oculomotor independent variables (including also polynomial trends and interactions among them) which are known to affect SFDs. These effects are represented in the unstandardized regression coefficients of a multiple regression model. Adding also the intercept (i.e., an estimate of the mean log SFD), we count a total of 44 fixed-effects parameters. While such a large number of parameters may appear to be daunting, especially after adding 12 parameters (plus 1 for the residual) for variance components (see below), their estimation is feasible, given that the analysis will be based on 80,625 SFDs (i.e., ~ 1414 observations per parameter). Fixed effects will be presented in detail further below. We describe results for SFDs measured in the right eye.

Random factors

For a rather simple reason we cannot analyze SFDs with a standard MRA: SFDs are not independent observations, but they are clustered according to three factors: subjects, sentences, and words.

Subjects. In psychological research, subjects are considered to be drawn at random from an underlying population. In our study, SFDs were measured in 273 readers who varied widely in age, size of vocabulary, and cognitive ability. On average a reader contributed 295 fixations with a range from 78 to 444 fixations. If we proceed from the reasonable assumption that at least five observations should be available per parameter, the minimum number of observatons given our fixed-effect regression model with 44 parameters is 1.8 = 78/44. Forty subjects had less than 220 fixations, that is less than five observations per parameter of the 44 fixed-effect model. Thus, we could hardly specify a repeated-measures MRA for a model with 44 predictors. LMMs take care of this problem by „borrowing strength“ from the estimates of population values. Individual differences in SFDs are reliable. Consequently, the fixations measured in a person are more similar to each other than those between these different persons. LMM take this intra-class correlation into account.

Sentences. The sentences constructed for the original Potsdam Sentence Corpus (PSC I; Kliegl, et al., 2004) are also a presumably random sample from the population of sentences one may encounter in simple prose. The sentences did not provide particular difficulty for comprehension, but they varied widely in syntactic structure and semantic content. Consequently, the SFDs measured on a given sentence are more similar to each other than fixations measured on different sentences. LMMs also take into account the intra-class correlation associated with sentences. As each subject read the same 144 sentences of the PSC I, sentences are specified as crossed with subjects. Of course, given loss of measurement, this crossing is highly imbalanced.

Words. We also assume that the words selected to compose the sentences are a random sample from the population of words. When we restrict the analyses to SFDs as described in Figure 2, sentences are composed from 369 different words; many words occur in more than one sentence (i.e., determiners, conjunctions, and prepositions), whereas others appear only once. Of course, frequently occuring words are typically also short and, consequently, they are skipped very often. In contrast, words that occur once are typically longer and skipped less frequently. Therefore, the repetition across sentences partially compensates the high skipping rate. Aggregating across subjects and sentences, we have sufficient information for the LMM to take into account the intra-class correlation of words.

Variance components

LMMs „remove“ the dependencies between observations that are due to the clustering of SFDs „belonging“ to a subject, a sentence, or a word. They do this on the assumption of independence between the three random factors and on the assumption that the deviations of the „levels“ of these factors from the fixed effect are normally distributed.

Varying intercepts. The main difference and advantage of LMM over traditional ANOVAs relates to estimates of parameters specifying the variance components associated with random factors. For the analysis of reading eye movements, we specify three random factors: subjects, sentences, and words. We assume that subjects yield a normal distribution of mean SFDs (actually, the log-transform of them). The mean of this distribution is returned as a fixed effect (see Appendix A, block of fixed effects: (Intercept) = 5.25). The variance of the between-subject SFD distribution is estimated as a first variance component (see Appendix A, block of random effects, subjects (Intercept)). Specifically, the square root of this variance, the standard deviation for between-subject differences of mean log SFD, is 0.152 for the present data. Similarly, we assume that sentences and words yield normal distributions of SFDs distributed around the same intercept. The variances of these two between-sentence SFD distributions are estimated as a second and third variance component. The square roots of these variances are listed with a value of 0.055 for sentences (see Appendix A, top block, sentences (Intercept)) and with a value of 0.094 for words (see Appendix A, top block, words (Intercept)). Thus, the SFD variance for subjects is larger than the SFD variance for words which is larger than the SDF variance for sentences. These three estimates represent independently varying intercepts for subjects, sentences, and words: An increase of the variance between subjects (e.g., by sampling from a broad range of reading ability) may leave the SFD variances of sentences and of words unchanged; an increase of the variance between sentences (e.g., by including sentences of particularily great syntactic difficulty—e.g., multiple embedding of relative clauses) could be independent of the SFD variance between subjects and the SFD variance between words.

Varying slopes. We estimate a second group of variance components that relates to within-subject effects of three frequency effects linked to words N-1, N, and N+1. Overall, the linear trends are three slopes that are returned as three fixed effects (i.e., N-freq linear = 0.039, N-1freq linear = -0.041, N+1 = -0.019; see Appendix A, block of fixed effects). The LMM allows us to test the hypothesis that there are reliable differences between subjects in how strongly their SFDs respond to the difference between low- and high-frequency words on the assumption that these slopes are normally distributed around the overall fixed effect. The between-subject variances of these three frequency effects yield three additional variance components estimated as LMM parameters. They are given in the random-effects block of Appendix A as N-1-freq (linear), N-freq (linear), and N+1-freq(linear); the standard deviations are 0.00044, 0.00085, and 0.00020, respectively.

C*ovariances between varying intercepts and varying effects (correlation parameters)*. So far, we have introduced the concept of varying intercepts and varying slopes. These are variance component parameters. In LMM, we can also estimate the associated covariances (or correlation parameters) if there are two or more components for a random factor. In the present model, this is the case for subjects. These correlation parameters tell us, for example, whether subjects with short average fixation durations tend to have strong or weak word-N frequency effects or whether there is a correlation between N+1-frequency, N-frequency, and N-1 frequency effects across subjects (Risse, Engbert, & Kliegl, 2008). These correlation parameters are listed in the random-effects block of Appendix A. Kliegl, Risse, & Engbert (2008) proposed that the two negative correlation parameters between N-frequency effects with N-1-frequency and N+1-frequency effects (i.e., -0.33 and -0.45, respectively) together with the positive correlation parameters between the latter two effects (i.e., 0.20) are consistent with an assumption of individual differences in perceptual span. Four variance components (1 intercept + 3 slopes) yield 6 model parameters to reconstruct the subject-related correlation parameters. Thus, there is a total of 12 model variance/correlation parameters (i.e., 4 variances + 6 correlation parameters for subjects + 1 variance for sentences + 1 variance for words). Finally, the model also returns a parameter for the estimate of the residual variance, 0.27 for the standard deviation in the present data (see random-effects block of Appendix A).

Random effects. The deviations of subjects, sentences, and words from the intercept and subjects' deviations of frequency effects from the corresponding six effects are the random effects in the LMM, but, we iterate, they are *not* the model parameters. The distributions of the random effects can be described as unconditional distributions and as conditional distributions. The model parameters afford a description of the model in terms of the unconditional distribution. The actual random effects are from the conditional distribution, given the data and the values of the model parameters. In our LMM example, there are only 12 model parameters denoting variance components and associated correlation parameters, but 1605 random effects (i.e., 273 for subjects + 144 for sentences + 369 for words = 786 intercept deviations; 273 subjects x 3 different types of slopes = 819 slope deviations). The model parameters are the variances of the various regression coefficients associated with each of these factors (and possibly also their correlations). Consequently, adding data from new subjects to a data base does not change the number of model parameters to be estimated; we are still estimating the same number of fixed effects and parameters for variance components. Of course, the increase in number of observations due to SFDs of additional subjects increases the precision of the parameters for the variance components. In contrast, adding a subject in a repeated MRA requires that all regression coefficients are estimated for this person. Thus, in an LMM, adding data from additional subjects increases only the number of random effects that can be generated conditional on the data and the model parameters.

Fixed effects of type frequencies of the fixated word and its left and right neighbors

LMM framework. The model (as sketched in Figure 2) represents the baseline LMM model of the Potsdam Sentence Corpus (PSC I) in Kliegl (2007), extended with lexical and contextual as well as oculomotor variables for three successive words. We use reciprocal values of launch site to reduce the collinearity with skipping status of word N-1 (r = .40 instead of .61) and likewise for outgoing saccade amplitude to reduce the correlation with skipping status of word N+1 (r = .42 instead .48). Motivated by the ongoing debate on lexical influences in the parafovea reviewed above and using a larger sample of readers (N=273), we test additional effects of frequency of word N-2 and word N+2 as additional predictors, conditional on whether word N-1 or word N+1 were skipped or not. The complete model output is provided in the fixed-effects block of Appendix A; it serves to document consistency of the present with earlier analyses and statistical control of effects not in the current focus of analysis (e.g., oculomotor variables). Differences between previous and current results will be discussed.

Illustration of effects. We use LMM statistics to guide our theoretical arguments and use figures to illustrate the core LMM results. The figures are not in a one-to-one correspondence with the statistics. For example, we present the effects in the original, not the log-transformed metric SFDs. We also fit polynomial functions to subsets of data, ignoring intraclass correlations due to the random factors. We discuss divergences between test statistics and parts of the figure. Figure 4 summarizes the five frequency effects on a SFD measured on word N associated with words N-2 to word N+2 and their interactions with skipping status of word N-1 and word N+2; Figure 5 contains the analogous information for the five predictability effects.

-- Insert Figures 4 and 5 about here --

Canonical fixation pattern: Fixations on word N-1, N, and word N+1. The first row of Figure 4 reflects the pattern when the SFD on word N was preceded and followed by a fixation on word N-1 and word N+1, respectively. Roughly half of all SFDs (49%) are in this category. The first set of fixed effects in Appendix A lists the regression coefficients associated with this category of fixations. Replicating previous results, there are significant cubic trends for the N-frequency and significant linear N+1 and N-1-frequency effects (see Figure 3). Here, to keep matters simple, we included a cubic trend only for the N-frequency effect and linear trends for everything else. New results of the present LMM are a lack of evidence for significant N+2-frequency and N-2-frequency effects. Interestingly, inclusion of N+2-predictability in the LMM, renders the previously reported N+1-predictability effect as not significant (Kliegl et al., 2006). In a way, however, this shift corroborates the interpretation of the non-canonical direction of this effect (i.e., high predictability of word N+2 leads to long SDF on word N) as an effect of memory retrieval during the fixation on word N. In summary and in general, given three fixations on successive words, SFDs on word N relate to the frequencies of these three words in a complex and clearly non-monotonic pattern; SFDs on word N are not significantly related to the frequencies of the words outside this 3-word window (i.e., words N-2 and N+2). Does this pattern also hold for skipped words N-1 or N+1?

Interactions of frequency and predictability with skipping. The no-skip category of SFDs serves as the reference category for the following tests of interactions with skipping status of words N+1 and N-1. All coefficients „S2 x *Coeff*“ in Appendix A (third block of fixed effects) test whether the effect indicated by *Coeff* is significantly different from the corresponding coefficient estimated for no-skip reference category when word N+1 was skipped. Analogously, all coefficients „S1 x *Coeff*“ in the fourth block of fixed effects test the difference between the effect reported for the no-skip reference category and the case when word N-1 was skipped. Importantly, the signs of these regression coefficients reflect whether an effect is weaker or stronger for no-skipping or skipping category.

N+1 skipping (2nd row in Figure 4 and 3rd block in Appendix A). Overall, there are significantly longer fixation durations prior to skipped words N+1 (S2; t = 7.1). This skipping cost is at odds with the non-significant skipping benefit reported in Kliegl and Engbert (2005), but the skipping effect returned by the current LMM is estimated under the assumption that all other covariates in the model assume a value of zero. Kliegl (2007) documented interactions between skipping and length and frequency of to-be-skipped word. Also, there are considerable individual differences in the skipping behavior between subjects which are caught in the random effect of subject in the LME. Most importantly, the cubic function relating SFD to the frequency of word N is of a significantly different shape if word N+1 is skipped (S2 x N-freq; t = 3.4) than the no-skip reference function. This difference is clearly visible in Figure 4. Indeed, much of the cubic shape is replaced by an almost symmetric quadratic function. In addition, there is a significantly stronger negative linear N+1-frequency effect for skipped than fixated words N+1 (S2 x N+1-freq; t = -2.4). Together these two interactions reflect that word N+1 is processed during the fixation on word N when word N is of high frequency and word N+1 is of low frequency, as in noun phrases.

N-1 skipping (3rd row in Figure 4 and 4th block in Appendix A). Fixations following a skipped word are longer (N-1-skip; t = 23.0). Moreover, the significant effect associated with the N-frequency effect (S1 x N-freq; t = -4.9) reflects the difference in the shape of the corresponding curves shown in Figure 4. Basically, after skipping word N-1, the cubic trend of the N-frequency effect is weaker (more montonic) compared to fixations following a fixation on word N-1. Thus, assuming that skipping word N-1 indicates little additional processing need for this word, the non-monotonic shape of the N-frequency effect in the canoncial condition is at least partly due to spillover of processing from word N-1. Consistent with this interpretation is also the significantly weaker N-1-frequency effect (S1 x N-1-freq; t = 3.1).

Finally and surprisingly, after skipping of word N-1, there is a significant negative effect of word N-2 on fixation durations on word N (S1 x N-2 freq; t = -5.8). Although this lag-2 effect is much weaker than the lag-1 effect, the frequency of the word last fixated (at position N-1 or N-2) lingers on in the fixation on word N. The reverse pattern of significant effects was obtained for the predictability of word N-2: No significant N-2-predictability effect for the reference case, but a significant positve effect if word N-1 was skipped (S1 x N-2 pred; t = 2.9). Given the absence of a positive trend in the respective panel of Figure 5, it is likely related to the N-2-frequency effect through a suppressor constellation (which also explains the statistically large negative N-2 frequency effect in the presence of a weak negative trend in the respective panel of Figure 4).

Discussion

Most research on visual word recognition is carried out with narrowly defined experimental paradigms of word naming or lexical decision. Much has been learned with eye-movement measures during reading collected from a few target words in experimental designs based on the variation of some close set contraints. Only a minority of research in eye movements is based on the analyses of entire large sentence corpora (e.g., Pynte & Kennedy, 2006; Kliegl et al., 2006; Rayner & McConkie, 1976; Radach & Heller, 2000). Our results are in line with the proposal of a processing gradient that affords distributed processing across more than the fixated word, but they offer little on the precise mechanisms that generate the profiles. In the following sections, we highlight some of the methodological issues and challenges facing the corpus-analytic perspective and the theoretical issues that need to be addressed. In the end, the frequency- and predictability signatures represented in Figures 4 and 5 need to emerge from general dynamical principles linking eye guidance and lexical processing.

Methodological issues and challenges

Lack of parsimony. We consider it conceptually simpler to formulate baseline models on the assumption that, as a rule, we will need a third-order polynomial function for lexical properties such as frequency. In an ensemble of, say, 12 covariates describing all of them as third-order polynome is simpler than describing, say, 1 of them with a linear, 2 with a quadratic, and 9 with a cubic function. A „generous“ model specification, keeping possibly non-significant terms, runs counter to current practice according to which models should be specified as parsimoniously as possible. The aim for parsimony is often motivated not only by conceptual simplicity but also by statistical limitations (e.g., not enough observations), or numerical or computational limits of the software we use to estimate the parameters. The abundance of SFDs and the current generation of software render these problems as not relevant for our situation. Conceptually, we forego the attempt to interpret all model parameters, but use them for a statistically adequate description. We justify this decision with reference to the argument that the statistical description achieved with an LMM must be complemented with computational modeling for the control of eye movements during sentence reading such as SWIFT (Engbert et al., 2005; Richter, Engbert, & Kliegl, 2006) or E-Z Reader (Reichle, et al.,1998; Reichle, Rayner, & Pollatsek, 2006). Indeed, we propose that the descriptive functions may constitute better simulation targets than the observed SFDs (see below).

LMM limits. In principle, variance components of frequency, predictability, and length effects cannot only be specified for subjects, but also for sentences. In practice, however, there are serious limitations on the number of variance components and correlation parameters that can be estimated in a single model. Moreover, in our admitttely limited experience, such between-sentence variance components are stastitically not reliable; including them does not significantly improve the goodness of model fit in a LRT. Obviously, such variance components can only be estimated for within-subject, within-sentence, or within-word effects. Thus, as far as words are concerned, only predictability qualifies as a within-word effect whereas length and frequency are between-word effects and cannot be specified in such a model.

How do we cumulate knowledge across different analyses of the same corpus? One of the lures of eye-movement corpora is their almost complete representation of the peripheral behavioral process of reading. The richness of the data affords analyses from countless different perspectives. Almost all observables (e.g., SFD, skipping status) serve as dependent variable in one context and as a covariate in another. Advances in computational linguistics provide new indicators of lexical processing for eye movements collected many years back. For example, measures of syllable and lemma frequeny as well as measures of orthographic neighborhood are already available as predictors for the present SFDs; many new ones will become available soon (Heister et al., 2011). In other words, with the addition of new predictors, the same measures can be recycled for new analyses. This is all welcome, but the maintainers of such corpora will have to set up a system that keeps track of the various analyses that are being carried out and, as accurately as possible, document the overlap of the analyses in a responsible way. If such a system is not put in place, it may become very difficult to recognize the degree to which supposedly new results are basically old results presented with new labels. Another difficulty relates to the evaluation and pursuit of alternatives of complex models like the one presented in this chapter. Such models generate a large number of test statistics. Communicating such models is probably a greater challenge than fitting them. We present the summary output of the model as Appendix A and describe the results pertaining to the theoretical questions that motivated the analyses; other results serve as a statistical control. Different theoretical perspectives, equally plausible and interesting, would necessarily motivate a different foregrounding and backgrounding of information. In our opinion, it is necessary that such alternatives can be pursued without reliance on the present authors. Consequently, this type of analysis requires that data and scripts are available on a public website (in our case: http://www.dlexDB.de).

Simultaneous estimation of fixation durations and probabilities. So far, we concentrated the analyses on SFDs, but eye movements during reading represent a dynamic interplay of fixation duration and fixation probability, that is of both "when" and "where" decisions. Consequently, whether a word is skipped, fixated once, twice, or three or more times is just as valid a dependent variable as fixation duration. For purposes of statistical inference this requires the specification of generalized linear mixed models (GLMMs), using either the binomial or poisson family for the description of the error distribution along with a suitable link function (e.g., logit link for binomial family). In principle, any progress in specification of a comprehensive LMM model transfers to GLMM specification, even though some variables change their status from predictor to dependent variable and vice versa; most notably, for example, fixation duration is included as a co-variate in GLMMs. In perspective, we expect that fixation durtions and fixation probabilities will be merged in co-called rate models. At present, however, the increase in complexity is unlikely to translate into a commensurate increase in theoretical progress. Again, complex models probably need to grow rather than being legislated top down.

Compatibility between gaze-contingent display change experiments and corpus analysis. Most experiments in the field of reading rely on designs based on the variation of some close set contraints (for an overview see Rayner, 1998). The focus is on processing a few open-class words like nouns, adjectives, or verbs; comparisons typically hold the syntactic structure constant (unless it is the target of the experimental manipulation). Research on syntactical parsing, as opposed to lexical or orthographic effects, in reading necessitates analyses based on whole sentences, but only a few global measures are typically considered (e.g., total reading time, re-reading time). The *corpus-analytic approach* comprises analyses based on large sets of sentences with statistical – not experimental - control of different lexical, oculomotor factors, and possibly also syntactic factors (see Kliegl, 2007). This approach allows one to generalize across word categories and syntactical structures, complementing analyses of specific target words. At this point in time, corpus analyses as presented in this chapter are based on single fixation durations (or likewise on gaze durations) only. In the future, analyses will aim at including multiple dependent measures or categorical outcomes.

Theoretical proposals about eye guidance in reading

The reported results have consequences for theoretical proposals about eye guidance in reading. We mention three controversial topics relating to whether lexical processing occurs with sequential shifts of attention or is better characterized by processing gradients covering not only one, but several words.

Evidence for distributed processing. Figures 4 and 5 corroborate our earlier claims (e.g., Kliegl et al., 2006; Kliegl, 2007) that SFD during reading are reflective of the processing of not only the fixated word, but also of the processing of the neighboring words. Here we established that the zone of influence extends at least to two words to the right and two words to the left of the currently fixated one. The results also show that the size of effects strongly decline as we move away from the center triplet, but that the effect sizes of the neighboring words can be as large as the effect sizes of the fixated word. The cubic trend associated with the fixated word N appears to be largely a consequence of the spillover from or anticipatory processing of last and next word, respectively.

Non-monotonic effect profiles and LMM-parameters as targets for computational models. Much cognitive research may be characterized by ordinal data patterns. As the task gets more difficult or skill is reduced, reading or reaction time increases and accuracy drops. Ordinal data profiles of this kind are comparatively easy simulation targets for computational models. Consequently, such data profiles are often compatible with all competing models. We submit that the non-monotonic profiles reported in this chapter for N-1-, N-, and N+1-frequency effects during reading represent a formidable challenge for computational models. At the same time, computational models are the best hope we have to deliver explanations of non-monotonic data profiles on the basis of the theoretical principles that guided their construction. For example, we consider it plausible that cubic trends for N- and N+1-frequency effects on SFD on word N are the consequence of the nonlinear dynamical principles guiding saccade-target selection from a field of fluctuating activation levels associated with the words of a sentence, as implemented in a future version of SWIFT. Indeed, the parameters of the polynomial function may turn out to be a more suitable simulation target than individual fixation durations.

Analysing large corpora of eye movements with linear mixed models will help us understand and model individual differences in readers and further unravel lexical, sublexical, and oculomotor processes.

Summary

Visual word recognition occurs in the context of reading, involving eye movements and fixations in the service of providing the necessary perceptual and language-related information. This chapter reviews research based on analyses of a large corpus of fixation durations that allows the simultaneous consideration of a large number of influences. We distinguish roughly three sources: (1) There are contributions that arise from language-related processes. By now effects of many of the over 50 linguistic properties of words contributing to processing efficiency in isolated word recogntion have been established for fixation durations or probabilities in normal reading. In addition, during reading of sentences, fixation durations and probabilities are influenced by variables coding the context of words, such as the predictability of words from prior words of the sentence or corpus-based statistics such as transition probabilities or surprisal measures. (2) There are contributions from the oculomotor process and low-level processes of eye guidance leading to effects related to preferred viewing locations or launch sites. (3) Finally, corpus analyses have provided reliable evidence that some indicators of language-related processes exert their effects during successive fixation durations. For example, the effect of the frequency of the next word depends strongly on whether this word will be skipped or not. We also provide evidence for an N+2-effect for word predictability in cases when three subsequent words are fixated. Regarding fixations after skippings, we find an influence of the frequency of the last fixated word (which could be word N-2) on the current fixation duration. We illustrate how such distributed processing of visual word recognition can be analyzed with state-of-the-art multivariate statistical techniques such as linear mixed models. We discuss the consequences of these results for theoretical proposals about eye guidance in reading.

References

Adelman, J. S., Brown, G. D. & Quesada, J. F. (2006). Contextual diversity not word frequency determines word naming and lexical decision times. *Psychological Science, 17 (9)*, 814-823.

Baayen, R. H. (2008). Analyzing Linguistic Data. *A Practical Introduction to Statistics Using R*. Cambridge University Press. Cambridge.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59 (5)*, 390-412.

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

Bates, D. M., & Maechler, M., (2010). *lme4: Linear mixed-effect models using S4 classes*. R package version 0.999375-33.

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 (4)*, 1162–1172.

Beauvillain, C. (1996). The integration of morphological and whole-word form information during eye fixations on prefixed and suffixed words. *Journal of Memory and Language, 35 (6)*, 801-820.

Boston, M. F., Hale, J., Kliegl, R., Patil, U. & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research, 2 (1)*, 1-12.

Brysbaert, M., Drieghe, D., Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. D. M. Underwood (ed.): *Cognitive Processes in Eye Guidance*, 53-78. Oxford University Press, Oxford.

Brysbaert, M. & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41 (4)*, 977-990.

Brysbaert, M. & Vitu, F. (1998). Word skipping: Implications for theories of eye movement control in reading. In G. D. M. Underwood (ed.): *Cognitive Processes in Eye Guidance*, 125-147. Oxford University Press, Oxford.

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.

Faraway, J. J. (2007). *Changes to the mixed effects models chapter in ELM*. Available at: http://www.maths.bath.ac.uk/~jjf23/ELM

Gelman, A. & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press. Cambridge.

Graf, R., Nagler, M., & Jacobs, A. M. (2005). Factor analysis of 57 variables in visual word recognition. *Zeitschrift für Psychologie, 213(4)*, 205-218.

Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., Kliegl, R. (2011). *dlexDB – eine lexikalische Datenbank für die psychologische und linguistische Forschung. Psychologische Rundschau, 62 (2)*, 10-20.

Heller, D., & Müller, H. (1983). On the relationship of saccade size and fixation duration in reading. In R. Groner, C. Menz, D. F. Fisher, & R. A. Monty (eds.), *Eye Movements and Psychological Functions: International Views*, 287-302. Erlbaum. Hillsdale, NJ.

Henderson, J.M., & Ferreira, F. (1990). The effects of foveal difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology*: Learning, Memory, and Cognition, 16, 417-429.

Hogaboam, T.W. (1982). Reading patterns in eye movement data. In K. Rayner (ed.), *Eye movements in reading: Perceptual and language processes* (pp. 309-332). New York: Academic Press.

Inhoff, A. W., Radach, R., Starr, M., & Greenberg, S. (2000). Allocation of visuo-spatial attention and saccade programming during reading. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (eds.), *Reading as a Perceptual Process*, 221-246. Elsevier. Amsterdam.

Just, M. A. & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87 (4)*, 329-54.

Kennedy, A. & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research, 45 (2)*, 153-168.

Kennedy, A., Pynte, J., & Ducrot, S. (2002). Parafoveal-on-foveal interactions in word recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, Section A, 55 (4)*, 1307-1337.

Kliegl, R. (2007). Towards a perceptual-span theory of distributed processing in reading: A reply to Rayner, Pollatsek, Drieghe, Slattery, & Reichle (2007). *Journal of Experimental Psychology: General, 138*, 530-537.

Kliegl, R., & Engbert, R. (2005). Fixation durations before word skipping in reading. *Psychonomic Bulletin & Review, 12*, 132-138.

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

Kliegl, R., Masson, M.E.J., & Richter, E.M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition*.

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

Kliegl, R., Risse, S., & Engbert, R. (2008). Reading strategies and their implementation in the SWIFT model of eye-movement control. *International Journal of Psychology, 43,* 190. (Abstract)

Krügel, A. & Engbert, R. (2010). On the launch-site effect for skipped words during reading. *Vision Research, 50*, 1532-1539.

Lorch, R., & Myers, J. (1990). Regression analyses of repeated measures data in cognitive research: A comparison of three different methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 149-157.

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

McConkie, G. W., Kerr, P. W., & Dyre, B. P. (1994). What are “normal” eye movements during reading: Toward a mathematical description. In J. Ygge & G. Lennerstrand. (Eds.), *Eye movements in reading* (pp. 315-327). Elsevier, Oxford.

McConkie, G. W., Kerr, P. W., Reddix, M. D., Zola, D., & Jacobs, A. M. (1989). Eye movement control during reading: II. Frequency of refixating a word. *Perception & Psychophysics*, 46, 245-253.

Murray, W. S. & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review, 111(3)*, 721-756.

Nuthmann, A., Engbert, R., & Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Research, 45 (17)*, 2201-2217.

Nuthmann, A., Engbert, R., & Kliegl, R. (2007). The IOVP-effect in mindless reading: Experiment and modeling. *Vision Research, 47*, 990-1002.

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

Pollatsek, A., Rayner, K., & Balota, D. A. (1986). Inferences about eye movement control from the perceptual span in reading. *Perception & Psychophysics, 40 (2)*, 123-130.

Pynte, J. & Kennedy, A. (2006). An influence over eye movements in reading exerted from beyond the level of the word: Evidence from reading English and French. *Vision Research, 46 (22)*, 3786-3801.

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

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59 (4)*, 413-425

R Development Core Team (2010). *R: A language and environment for statistical computing*. (version 2.11.1). [Software].

Radach, R. & Heller, D. (2000). Relations between spatial and temporal aspects of eye movement control. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (eds.), *Reading as a Perceptual Process*, 165-191. Elsevier. Amsterdam.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7 (1)*, 65-81.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124 (3)*, 372-422.

Rayner, K. (2009). The Thirty Fifth Sir Frederick Bartlett Lecture: Eye movements and attention during reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62, 1457-1506.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30 (4)*, 720-732.

Rayner, K. & Bertera, J. H. (1979). Reading without a fovea. *Science, 206*, 468-469.

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(7)*, 943-954.

Rayner, K. & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition, 14 (3)*, 191-201.

Rayner, K. & Juhasz B. J. (2004). Eye movements in reading: Old questions and new directions. *European Journal of Cognitive Psychology, 16 (1&2)*, 340-352.

Rayner, K., Juhasz B. J., Brown, S. J. (2007). Do readers obtain preview benefit from word n + 2? A test of serial attention shift versus distributed lexical processing models of eye movement control in reading. *Journal of Experimental Psychology: Human Perception and Performance, 33 (1)*, 230-245.

Rayner, K. & McConkie, G. W. (1976) What guides a reader's eye movements. *Vision Research, 16 (8)*, 829-837.

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 (5), 1188-1200.

Rayner, K. & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin and Review, 3 (4)*, 504-509.

Rayner, K., White, S., Kambe, G., Miller, B., & Liversedge, S. (2003). On the processing of meaning from parafoveal vision during eye fixations in reading. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, 213-234. Elsevier, Amsterdam.

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.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2006). E-Z Reader: A cognitive-control, serial-attention model of eye-movement control during reading. *Cognitive Systems Research, 7*, 4-22.

Richter, E., Engbert, R. & Kliegl, R. (2006) Current advances in SWIFT. *Cognitive Systems Research, 7*, 23-33.

Risse, S., Engbert, R., & Kliegl, R. (2008). Eye-movement control in reading: Experimental and corpus-analytic challenges for a computational model. In K. Rayner, D. Shen, X. Bai, & G. Yan (Eds.) *Cognitive and cultural influences on eye movements.* (pp. 65-92). Tianjin: Tianjin People's Publishing House/Psychology Press.

Starr, M. S. & Inhoff, A. W. (2004). Attention allocation to the right and left of a fixated word: Use of orthographic information from multiple words during reading. *European Journal of Cognitive Psychology, 16 (1&2)*, 203-225.

Starr, M. S. & Rayner, K. (2001). Eye movements during reading: Some current controversies. *Trends in Cognitive Science, 5 (4)*, 156-163.

Underwood, G., Binns, A., & Walker, S. (2000). Attentional demands on the processing of neighbouring words. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (eds.), *Reading as a Perceptual Process*, 247-268. Elsevier, Amsterdam.

Underwood, N. R. & McConkie, G. W. (1985). Perceptual span for letter distinctions during reading. *Reading Research Quarterly, 20 (2)*, 153-162.

Vitu, F., O'Regan, J.K. Inhoff, A., & Topolski, R. (1995). Mindless reading: Eye Movement characteristics are similar in scanning strings and reading texts. *Perception and Psychophysics*, 57, 352-364.

Vitu, F., McConkie, G. W., Kerr, P., & O'Regan, J. K. (2001). Fixation location effects on fixation durations during reading: An inverted optimal viewing position effect. *Vision Research, 41 (25-26)*, 3513-3533.

White, S. J. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 34 (1)*, 205-223.

White, S. J. & Liversedge, S. P. (2004) Orthographic familiarity influences initial eye fixation positions in reading. *European Journal of Cognitive Psychology, 16 (1&2)*, 52 -78.

Appendix A: LMM Results, estimated with *lmer* (Bates & Maechler, 2010)

```
   AIC   BIC logLik deviance REMLdev
  21465 22014 -10674   20965    21347
Random effects:
 Groups   Name          Variance    Std.Dev.   Corr
 words    (Intercept)    0.00881   0.09385
 subjects (Intercept)    0.02300   0.15166
          N-1-freq (lin) 0.00044   0.02109  -0.338
          N-freq (lin)   0.00085   0.02908   0.113  -0.327
          N+1-freq (lin) 0.00020   0.01429  -0.254   0.179  -0.448
 sentences(Intercept)    0.00304   0.05509
 Residual                0.07258   0.26941
Number of obs: 80625, groups: words; 369; subjects, 273; sentences, 144

Fixed effects:
                       Estimate  Std.Error t-value
(Intercept)             5.25116   0.01301  402.21 # Mean log(SFD)
N-freq (linear)         0.03878   0.01218    3.18 # Effects of Word-N properties
N-freq (quadr)          0.00595   0.00579    1.03
N-freq (cubic)         -0.02010   0.00384   -5.23
N-pred                 -0.04033   0.00297  -13.56
1/N-length             -0.08175   0.08238   -0.99
N-1-freq               -0.04112   0.00364  -11.29 # Effects of Word-N-1 properties
N-1-pred               -0.01498   0.00298   -5.03
1/N-1-length           -0.00981   0.03803   -0.26
N+1-freq               -0.01921   0.00368   -5.22 # Effects of Word-N+1 properties
N+1-pred                0.00062   0.00312    0.20
1/N+1-length            0.29004   0.03393    8.55
N-2-freq               -0.00392   0.00329   -1.19 # Effects of Word-N-2 properties
N-2-pred               -0.00364   0.00345   -1.06
1/N-2-length           -0.05489   0.03739   -1.47
N+2-freq               -0.00142   0.00304   -0.47 # Effects of Word-N+2 properties
N+2-pred                0.00765   0.00204    3.76
1/N+2-length            0.06684   0.03359    1.99

1/launch site           0.24856   0.00562   44.23 # Oculomotor variables
1/sacc. Ampl.           0.34264   0.02854   12.01
IOVP (linear)           0.09485   0.00430   22.06
IOVP (quadr)           -0.11264   0.01086  -10.37
N-freq x N-1-freq       0.00881   0.00171    5.14 # Interactions
N-freq x 1/N-length     0.31309   0.06163    5.08 # (see Kliegl et al. 2006)
N+1-freq x 1/N-length   0.05023   0.02483    2.02
N+1-pred x 1/N-length  -0.11911   0.02403   -4.96

N+1-skip (S2)           0.02406   0.00340    7.07 # Effect of skipping word N+1
S2 x N-freq             0.00916   0.00266    3.44 # ... x Word-N properties
S2 x N-pred             0.00120   0.00247    0.48
S2 x N-1/length         0.13062   0.03204    4.08
S2 x N+1-freq          -0.00631   0.00266   -2.37 # ... x Word-N+1 properties
S2 x N+1-pred          -0.00387   0.00232   -1.67
S2 x 1/N+1-length      -0.19559   0.02961   -6.61
S2 x N+2-freq          -0.00276   0.00259   -1.07 # ... x Word-N+2 properties
S2 x N+2-pred           0.00098   0.00201    0.49
S2 x N+2-1/length       0.01641   0.02962    0.55

N-1-skip (S1)           0.07671   0.00333   23.03 # Effect of skipping word N-1
S1 x N-freq            -0.01307   0.00266   -4.92 # ... x Word-N properties
S1 x N-pred             0.02795   0.00254   11.02
S1 x 1/N-length        -0.12015   0.03314   -3.63
S1 x N-1-freq           0.00867   0.00283    3.06 # ... x Word-N-1 properties
S1 x N-1-pred          -0.00308   0.00268   -1.15
S1 x 1/N-1-length       0.33280   0.03187   10.44
S1 x N-2-freq          -0.01733   0.00301   -5.75 # ... x Word-N-2 properties
S1 x N-2-pred           0.00908   0.00314    2.89
S1 x 1/N-2-length      -0.02433   0.03140   -0.77
```

Author Note

This research was supported by DFG Grant KL 955/12-1 to Reinhold Kliegl, Alexander Geyken and Thomas Hanneforth. Data and R scripts for analyses are available at the dlexDB project website. http://www.dlexdb.de. The website provides a wide variety of German word statistics for psychological and linguistic experiments. Correspondence should be addressed to Julian Heister (heister@uni-potsdam.de) or Reinhold Kliegl (kliegl@uni-potsdam.de) , Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany.

Figure Captions

Figure 1. Different patterns of three successive fixations. N = 211.972 fixations (•) classified. Fixations are classified by forward/backward movements including skippings (rows) and single, first, and second fixations (columns). Vertical lines represent word boundaries, dashed lines indicate skippings. The LMM-model (see appendix) includes a subset (namely firstpass) fixations in the highlighted panels. N = 11.127 (5%) fixations unclassified.

Figure 2. Distributed processing framework for possible influences on single-fixation duration (•) on word N; arrows represent possible incoming (left) or outgoing (right) saccades (i.e., word N – 1 and word N + 1 are fixated or skipped). Influences of Length, Frequency (Freq), and predictability (Pred) of word N-2, N-1, N, N+1, and N+2. (…) = additional predictors. Vertical lines represent word boundaries.

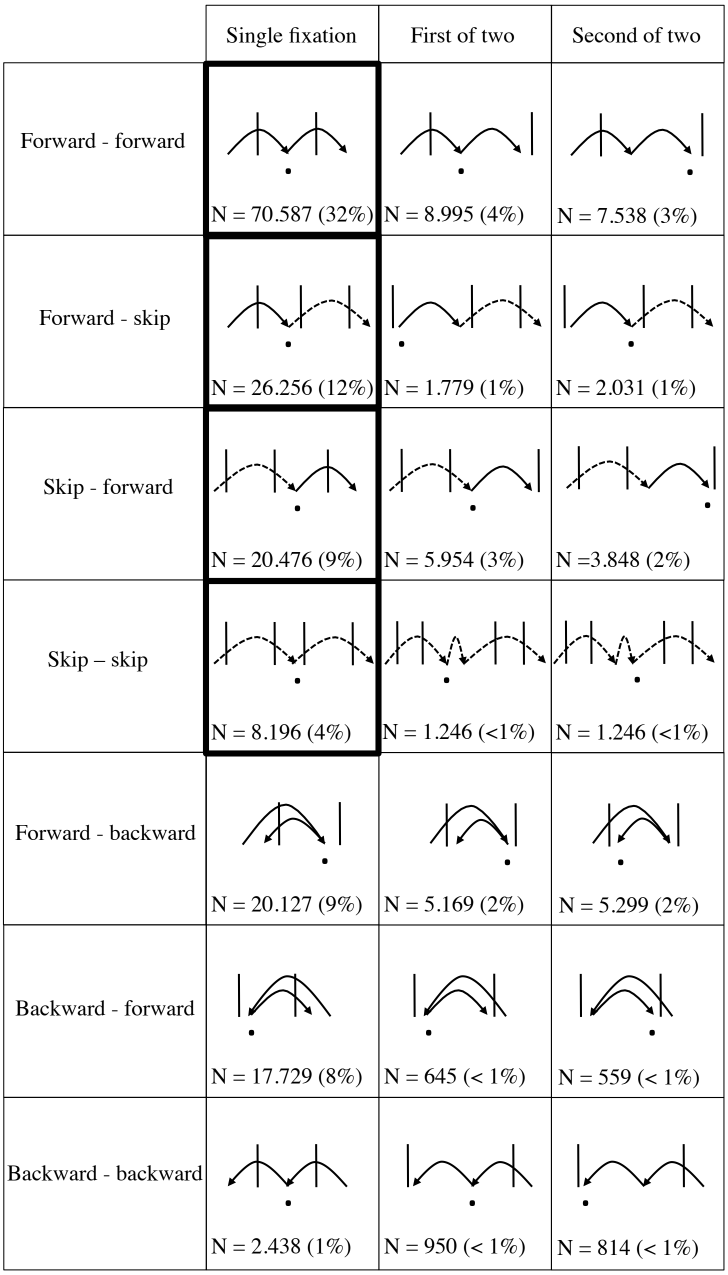
Figure 3. (top) Effects of frequency of word N-1, word N, and word N+1 on single-fixation duration on word N based on data from PSC I (first set of 144 sentences; 273 readers). (bottom) Like (top), but based on data from PSC II (second set of 144 sentences; 149 readers). Lines are third-order polynomials.

Figure 4. Dependent variable: Single-fixation durations on word N. Independent variables: Log type frequency of word N-2, word N-1, word N, word N+1, and word N+2 (columns) conditional on whether word N-1 and word N+1 were fixated (top row) or skipped (other rows). Lines are third-order polynomials. All fits are for the subset of fixations in the respective panel. Fixations after skipping word N-1 and before skipping word N+1 (9%) enter both skipping conditions. This results in a total of 110% fixations entering the plot.

Figure 5. Dependent variable: Single-fixation durations on word N. Independent variables: Logit predictability of word N-2, word N-1, word N, word N+1, and word N+2 (columns) conditional on whether word N-1 and word N+1 were fixated (top row) or skipped (other rows). All lines are linear

trends for the respective subset of fixations in the panel. Fixations after skipping word N-1 and before skipping word N+1 (9%) enter both skipping conditions. This results in a total of 110% fixations entering the plot.

Figure 1



	Single fixation	First of two	Second of two
Forward - forward	N = 70.587 (32%)	N = 8.995 (4%)	N = 7.538 (3%)
Forward - skip	N = 26.256 (12%)	N = 1.779 (1%)	N = 2.031 (1%)
Skip - forward	N = 20.476 (9%)	N = 5.954 (3%)	N =3.848 (2%)
Skip – skip	N = 8.196 (4%)	N = 1.246 (<1%)	N = 1.246 (<1%)
Forward - backward	N = 20.127 (9%)	N = 5.169 (2%)	N = 5.299 (2%)
Backward - forward	N = 17.729 (8%)	N = 645 (< 1%)	N = 559 (< 1%)
Backward - backward	N = 2.438 (1%)	N = 950 (< 1%)	N = 814 (< 1%)

Figure 2

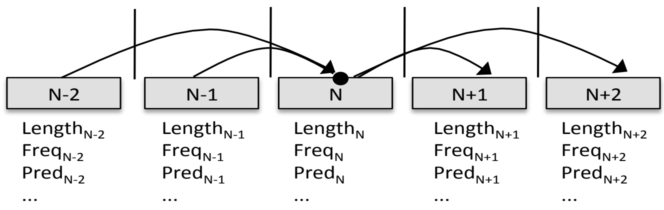

N-2
N-1
N
N+1
N+2
Length$_{N-2}$
Freq$_{N-2}$
Pred$_{N-2}$
...
Length$_{N-1}$
Freq$_{N-1}$
Pred$_{N-1}$
...
Length$_{N}$
Freq$_{N}$
Pred$_{N}$
...
Length$_{N+1}$
Freq$_{N+1}$
Pred$_{N+1}$
...
Length$_{N+2}$
Freq$_{N+2}$
Pred$_{N+2}$
...

Figure 3

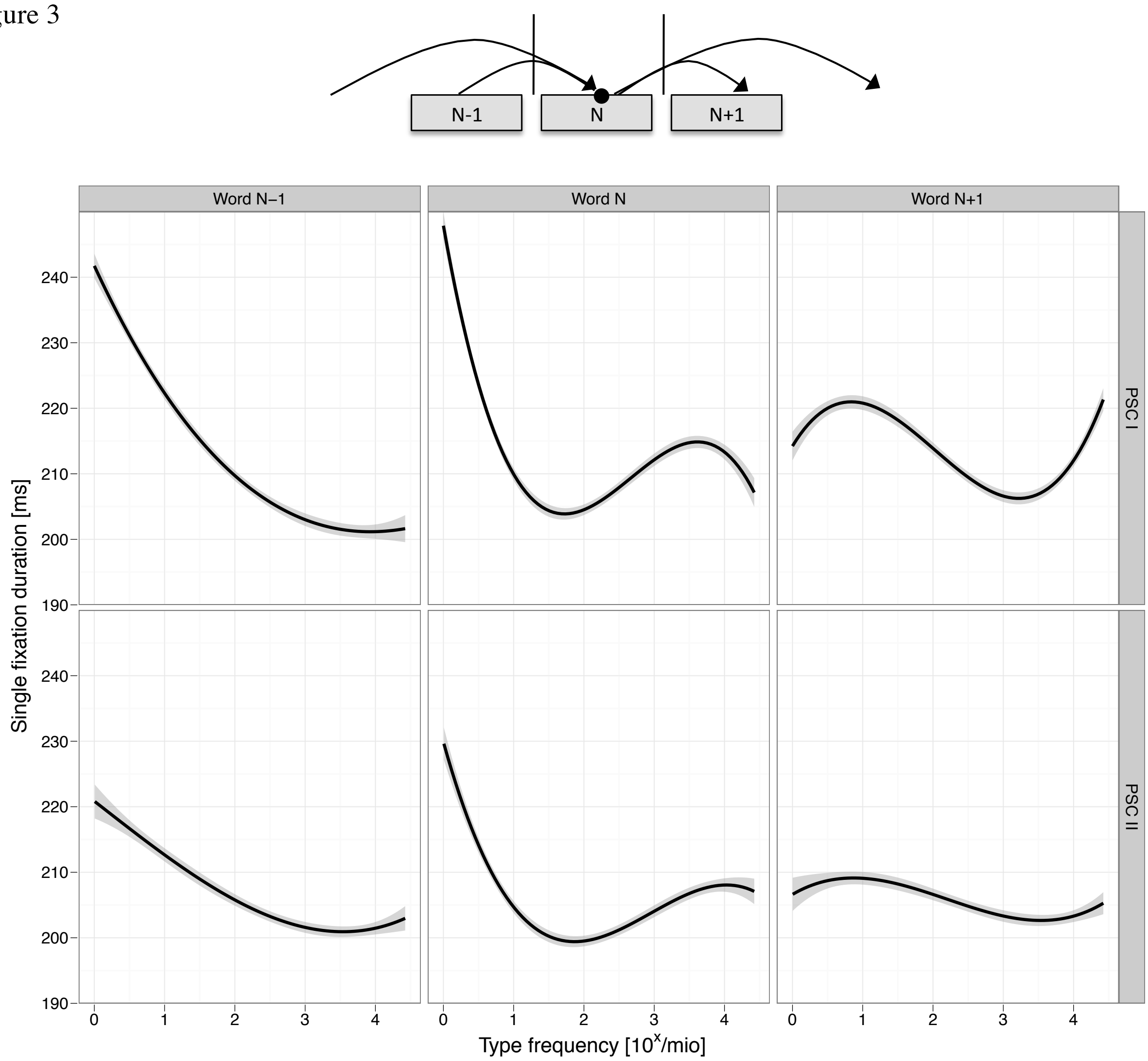

N-1
N
N+1
Word N−1
Word N
Word N+1
PSC I
PSC II
Single fixation duration [ms]
190
200
210
220
230
240
0
1
2
3
4
Type frequency [10^x/mio]

Figure 4

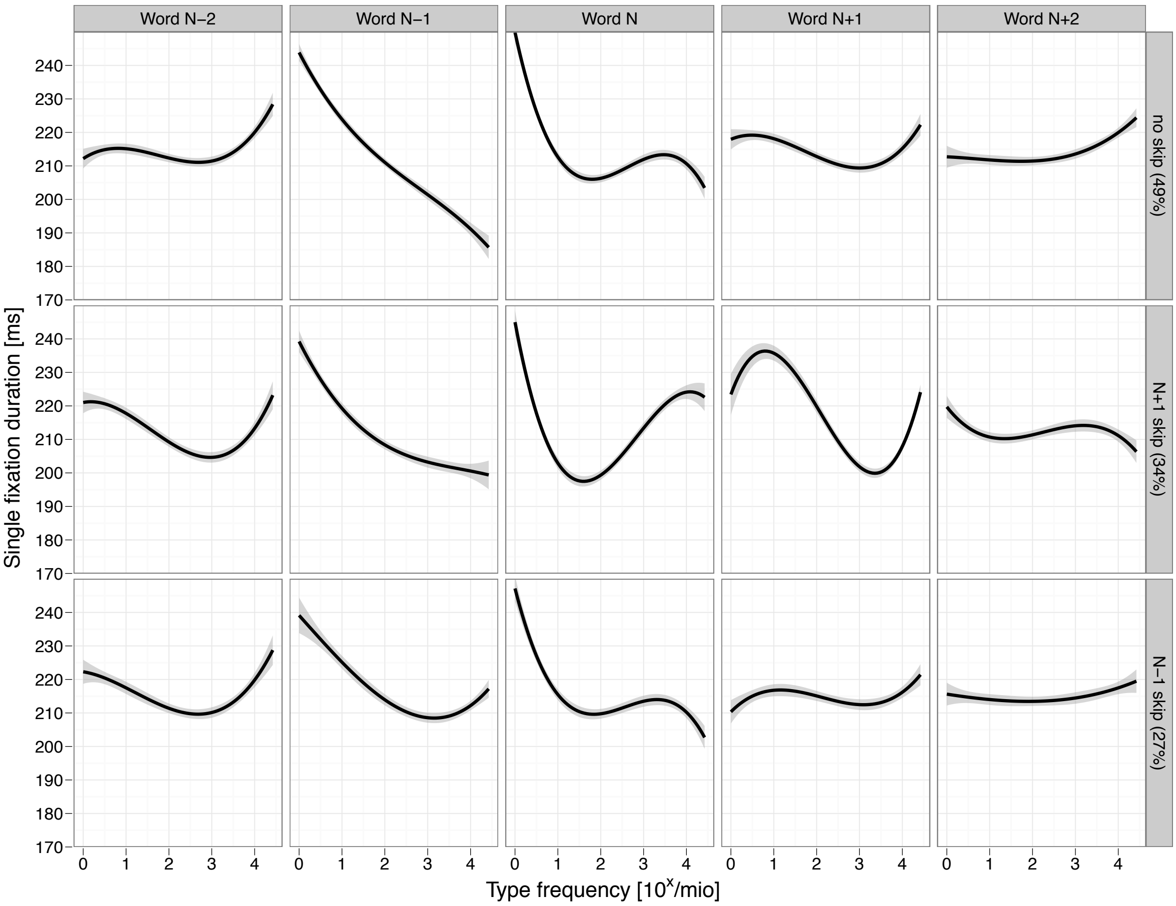
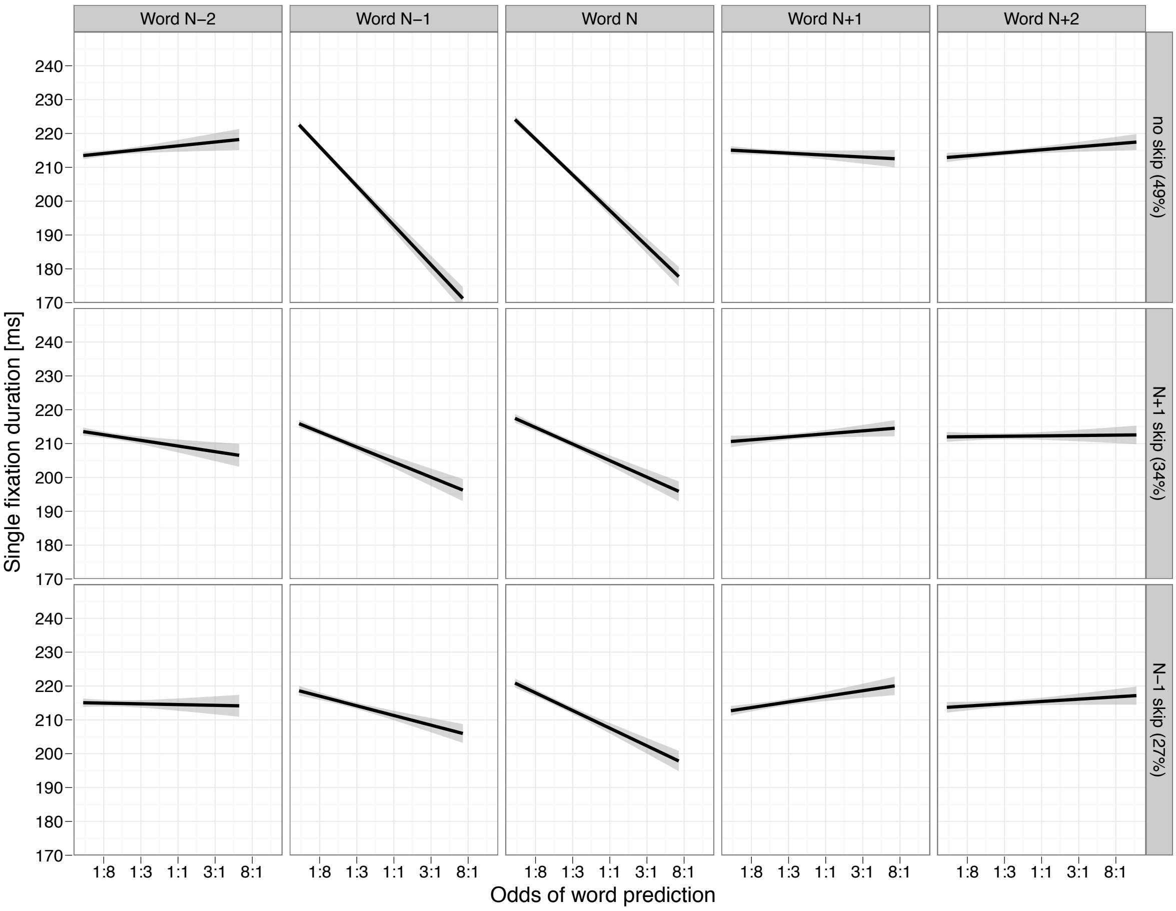

Word N−2
Word N−1
Word N
Word N+1
Word N+2
no skip (49%)
N+1 skip (34%)
N−1 skip (27%)
Single fixation duration [ms]
170
180
190
200
210
220
230
240
0
1
2
3
4
Type frequency [10^x/mio]

Figure 5


Word N−2
Word N−1
Word N
Word N+1
Word N+2
no skip (49%)
N+1 skip (34%)
N−1 skip (27%)
Single fixation duration [ms]
240
230
220
210
200
190
180
170
1:8
1:3
1:1
3:1
8:1
Odds of word prediction