EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY, 2004, *16* (1/2), 262–284

Length, frequency, and predictability effects of words on eye movements in reading

Reinhold Kliegl, Ellen Grabner, Martin Rolfs, and Ralf Engbert
University of Potsdam, Germany

We tested the effects of word length, frequency, and predictability on inspection durations (first fixation, single fixation, gaze duration, and reading time) and inspection probabilities during first-pass reading (skipped, once, twice) for a corpus of 144 German sentences (1138 words) and a subset of 144 target words uncorrelated in length and frequency, read by 33 young and 32 older adults. For corpus words, length and frequency were reliably related to inspection durations and probabilities, predictability only to inspection probabilities. For first-pass reading of target words all three effects were reliable for inspection durations and probabilities. Low predictability was strongly related to second-pass reading. Older adults read slower than young adults and had a higher frequency of regressive movements. The data are to serve as a benchmark for computational models of eye movement control in reading.

An important goal of reading research is to disentangle visual, lexical, and contextual processing factors. Three word characteristics that have been reliably linked to these three factors are their length, their frequency in written texts, and their predictability from the preceding context of the sentence. These relations have been established for various measures of inspection duration (e.g., first fixation duration, single fixation duration, gaze duration; see Methods for definitions) and inspection probabilities (e.g., of skippings, single fixations, or multiple fixations) but there is still little research looking at these effects in combination within the same data set (see Rayner, 1998, for a general review, and Calvo & Meseguer, 2002, for a review and the most recent specific attempt to examine the combination of these effects). In the present experiment we examined the influence of these variables on the set of standard measures and a

Correspondence should be addressed to R. Kliegl, Dept of Psychology, University of Potsdam, PO Box 60 15 53, 14451 Potsdam, Germany. Email: Kliegl@rz.uni-potsdam.de

We thank Petra Grüttner and Sabine Kern for research assistance. Antje Nuthmann, Ralph Radach, and an anonymous reviewer provided helpful comments on an earlier version of this paper. Sentences and associated word norms are available from the authors. This research was supported by grant KL 955/3 from Deutsche Forschungsgemeinschaft.

http://www.tandf.co.uk/journals/pp/09541446.html DOI:10.1080/09541440340000213

few additional measures of processing beyond first-pass reading. Most importantly, we tested the effects (1) for all words of the sentence corpus (facing the problem of correlated effects) and (2) for a subset of target words uncorrelated in length and frequency (facing the problem of generalisability to all words of the corpus) with repeated-measures multiple regressions in both sets of analyses. This way we could compare effects of word length, frequency, and predictability in a statistical control approach with those from an experimental control approach. Finally, we also checked the stability of these effects across the adult life span with samples of young and older adults.

A VIEW FROM THE EXPERIMENTAL CONTROL APPROACH

Given the large correlation between length and frequency of words for all words of a text or sentence corpus (e.g., –.64 in the one used in this experiment), it has been a longstanding problem to determine the independent contributions of these variables to visual and lexical processes.[1] The experimental method of choice to dissociate effects such as word length and word frequency, and the one most frequently used, is to restrict the analyses to a subset of target words which then by experimental design are orthogonal or uncorrelated. Our experiment also afforded such an analysis because one target word per sentence had been specified a priori to contribute to an orthogonal design with frequency and word length as factors.[2]

There is already much information available about the effects of word length and word frequency on inspection durations and inspection probabilities. In general, they covary with processing difficulty, that is they increase with word length and decrease with word frequency (e.g., Rayner, Sereno, & Raney, 1996; Schilling, Rayner, & Chumbley, 1998). However, the effect of word frequency on inspection probabilities has not been observed consistently (e.g., Gautier, O'Regan, & Le Gargasson, 2000; Henderson & Ferreira, 1993; Kennison & Clifton, 1995) and there were also exceptions to effects of word length on early processing stages (i.e., no effect on the duration of first fixations; Hyönä & Olson, 1995).

There are quite a few studies demonstrating the influence of frequency and predictability with word length controlled. For example, Brysbaert and Vitu (1998) identified seven previously published studies for their meta-analysis of

[1] It could be argued that word length is also a linguistic variable. Specifically, one may assume that knowledge about the length of a to-be-processed word constrains the number of candidates in the lexical processing stage. However, as shown in a recent series of studies, this appears not to be the case (Inhoff & Eiter, 2003; Inhoff, Radach, Eiter, & Juhasz, 2003).

[2] Another strategy is post-hoc selection of words from the reading material for an orthogonal design (e.g., Kliegl, Olson, & Davidson, 1983; Radach & Heller, 2000; Radach & McConkie, 1998).

frequency effects on word skipping and eight studies for the analogous analysis of predictability effects (see also Rayner, Binder, Ashby, & Pollatsek, 2001, for a recent report). Low-predictability words also attract regressive movements (Rayner & Well, 1996). There are also reports about effects of predictability on inspection durations. Binder, Pollatsek, and Rayner (1999; see also Rayner et al., 2001, Rayner & Well, 1996) reported shorter first-fixation durations for predictable words. Calvo and Meseguer (2002) also examined the time course of effects: Word length was significant for early and late processing stages (i.e., first-fixation and gaze durations), frequency for early stage (i.e., first-fixation durations) and predictability primarily for late stage of processing (i.e., gaze durations). Moreover, word length accounted for more variance (25%) than frequency (4%) and predictability (10%) when averaging across a number of statistically significant eye movement measures.

A VIEW FROM THE STATISTICAL CONTROL APPROACH

Clearly, then, there is already very reliable evidence for the influence of word length, frequency, and predictability on eye movement measures from studies using the experimental control approach. There remains the question of how results based on this approach generalise to all words of a sentence corpus. One can use statistical control techniques such as multiple regression to assess the effects of correlated predictors. Lorch and Myers (1990) described an appropriate procedure following less suitable earlier proposals (Blanchard, 1985; Just & Carpenter, 1980; Kliegl, Olson, & Davidson, 1982), which nevertheless are still in use. Specifically, variables such as word length, frequency, and predictability represent continuous repeated-measures design factors. Therefore, the data need to be analysed with repeated-measures multiple regression analyses removing systematic variance between persons. This can be accomplished by (1) regressing the dependent variable of choice (i.e., various inspection durations or probabilities) on the predictor set for each participant and (2) testing whether the means of the unstandardised coefficients (averaged across participants) differ significantly from zero. Moreover, one can test whether the means differ significantly between groups such as young and older adults in the present experiment. In ANOVA terminology, a significant group difference on a predictor (such as length, frequency, or predictability) translates into a significant interaction between the group and this factor. The disadvantage associated with the multiple-regression approach (repeated-measures based or not) is that variance shared between predictors cannot be used for statistical tests of effects, leading to much more conservative tests compared to experimental designs with orthogonal factors, especially in the case of highly correlated predictors such as word length and word frequency.

The studies mentioned above examined the effects of word length and word frequency, mostly in combination with a set of other language-related factors (e.g., function vs. content words, number of syllables). In general but not always, results related to word length and frequency were consistent with those reported above but, as the statistics were not appropriate (i.e., between-subject variance was not removed), we do not present a review here. Moreover, as far as we know, there is no research examining word length and frequency together with word predictability for data of all words of sentences or texts. Indeed, apparently only Reichle, Pollatsek, Fisher, and Rayner (1998) provided length, frequency, and predictability information for all words of a corpus of sentences; they supplemented data collected by Schilling et al. (1998) with predictability information for use in the E-Z Reader model. Reichle et al. did not report any specific analyses contrasting length, frequency, and predictability effects.

Thus, there still is no information about differences in effect sizes associated with the experimental control and statistical control methods. Of course, there can be no doubt that, irrespective of generalisability, experimental effects are highly important for theories about eye movement control in reading. However, we often do not know whether these effects are of any practical relevance in normal reading. For example, if there is no reliable unique variance associated with frequency or predictability beyond that accounted for by word length in an analysis of eye movements including all words of a text, one may well ignore these variables for the prediction of individual differences in reading efficiency. More importantly even, knowledge about the generalisability of experimental effects is of high theoretical relevance for the development and evaluation of computational models of eye movement control during reading because these models claim to embody the dynamics of ''normal reading'' and, consequently, must be evaluated on a variety of eye movement measures for all words of sentences or texts, not only a select subset of them. Ideally, of course, computational models should reproduce differences between results from experimental control and statistical control techniques.

Our selection of dependent variables closely followed the scheme proposed by Reichle et al. (1998). Engbert and Kliegl (2001), Engbert, Longtin, and Kliegl (2002) and Reichle et al. (1998; Reichle, Rayner, & Pollatsek, 1999, in press) used these benchmark data as target for their computational models. Obviously, a second corpus with similar and additional information will facilitate the further development of computational models. Reichle et al. (1998) reported first fixation durations (FF), single fixation durations (SF), gaze durations (GD), as well as probabilities for skipping (P0), single fixations (P1), and multiple fixations (P2) on first-pass reading as a function of five word-frequency classes. We extended this protocol with information about the effects of word length and predictability and with measures relating to information uptake beyond first-pass reading. Specifically, we added total reading time per word (TT) and probabilities of a word serving as origin (RO) or goal (RG) of a regressive eye movement. The general expectation was that difficult words might attract reinspections to compensate for word identification problems (Vitu & McConkie, 2000). It turned out that low predictability may be the most critical feature triggering such reanalysis.

The main goal of this experiment, then, was to replicate and extend well-known effects of word length, frequency, and predictability with a standard set of eye movement measures within a coherent framework of analyses covering both data from all words as well as data from an a priori specified set of target words. Moreover, we hoped to shed new light on some less stable effects (such as the effect of word frequency on inspection probabilities and the effect of word length on first fixation duration). Finally, on another dimension of generalisability, we wanted to check the stability of all these effects across the adult life span.

METHOD

Participants

Thirty-three university students (*M* = 21.9, *SD* = 2.2, range: 19–28 years) and 32 older adults (*M* = 69.9, *SD* = 3.9, range: 65–83 years) participated in this study: Young and older adults showed the typical pattern of equivalence in Lehrl's (1977) multiple-choice measure of vocabulary (old: *M* = 33.1, *SD* = 1.2; young: *M* = 32.8, *SD* = 1.0) and large differences in Wechsler's (1964) digit symbol substitution (old: *M* = 49.3, *SD* = 10.0; young: *M* = 67.9, *SD* = 8.1). They were all native speakers of German. Sessions lasted about one hour. Participants were paid an equivalent of 5 €/ hour or received credit in partial fulfillment of study requirements. Data from an additional 5 young and 10 older adults were excluded because of calibration problems during testing or because of low data quality (see *Analyses* below); their psychometric scores did not differ from the final samples.

Apparatus

Sentences were presented on the centre line of a 21-inch EYE-Q 650 monitor (832 × 632 resolution; frame rate 75 Hz; font: regular New Courier 12) controlled by an Apple Power Macintosh G3 computer. Participants were seated 60 cm in front of the monitor with the head positioned on a chin rest. Thus, letters subtended 0.35° of visual angle. Eye movements were recorded with an SR EyeLink System (SMI) with a sampling rate of 250 Hz and an eye position resolution of 20 sec-arc. Calibrated eye position was recorded accurately at the level of letters. Data were collected in two laboratories with identical equipment and setup.

The Potsdam Sentence Corpus

Word length. The Potsdam Sentence Corpus comprises 144 German sentences (1138 words). They were constructed with the goal to represent a large variety of grammatical structures around a set of target words (one or two per sentence; see below) that are uncorrelated in length and frequency. Sentence lengths range from 5 to 11 words with a mean of 7.9 words. Excluding the first word of each sentence which was not used in the analyses, frequencies of word lengths 2–13+ are: 54, 222, 134, 147, 129, 92, 72, 66, 20, 25, 16, and 17. (The category 13+ contains seven words of length 14–20.)

Printed frequency. CELEX Frequency Norms (Baayen, Piepenbrock, & van Rijn, 1993) are available for all 1138 words. Excluding the first word of each sentence, the corpus contains at least 76 words in each of five logarithmic frequency classes: class 1 (1–10 per million): 242 words; class 2 (11–100): 207 words; class 3 (101–1000): 242; class 4 (1001–10,000): 227; class 5 (10,001–max): 76 words. The CELEX corpus is based on approximately 5.4 million words.

Predictability. Predictability of words was collected in an independent norming study from 272 native speakers of German: 116 high-school students (age range: 17–19 years), 76 university students (age range: 19–38 years), and 80 older adults (age range: 66–80 years). Participants differed in the number of sentences they worked through: Twenty older adults generated predictions for all of the sentences; the other participants generated predictions for a quarter of the sentences. Collapsing the complete and partial protocols across participants, there were 83 complete predictability protocols, specifically 29 from high-school students, 19 from university students, and 35 from older adults. Completion of a complete protocol (144 sentences) lasted about 2.5 hours. Sentences were presented in a random order for each participant. Participants guessed the first word of the unknown original sentence and entered it via the keyboard. In return, the computer presented the first word of the original sentence on the screen. Responding to this, participants entered their guess for the second word and so on, until a period indicated the end of the sentence. Correct words stayed on the screen. There were no significant differences between the predictability distributions generated by the three samples. Excluding the first word of each sentence, the corpus contains at least 88 words in each of five logit-based predictability classes: class 1 (-2.553 to -1.5): 506 words; class 2 (-1.5 to -1.0): 111 words; class 3 (-1.0 to -0.5): 114 words; class 4 (-0.5 to 0): 88 words; class 5 (0 to 2.553): 175 words. Logits are defined as 0.5*ln(pred/(1–pred)); predictabilities of zero were replaced with 1/(2*83) and those of the five perfectly predicted words with (2*83–1)/(2*83), where 83 represents the number of complete predictability protocols (Cohen & Cohen, 1975). Upper boundaries of classes correspond to predictabilities of 0.0474, 0.1192, 0.2689, 0.5, and 1.0. For a word with predictability 0.50, the odds of guessing are 1:1=1, and the log odds of guessing are ln(1)=0. Thus, words with predictability larger than 0.50 yield positive logits, those with predictabilities smaller than 0.50 negative logits. The logit transformation corrects for the dependency of mean probabilities (p) and associated standard deviations (SD) [i.e., $SD = p(1\text{-}p)$] by stretching the tail of the distribution.

Target words. Each sentence contained one target word selected from the CELEX database contributing to a 2 × 2 × 3 design with word class (noun vs. verb), printed frequency (high: > 50 occurrences/million vs. low: 1 to 4 occurrences/million), and word length (short: 3, 4 letters, medium: 5–7 letters, long: 8, 9 letters); there were 12 sentences in each cell of this design. The position of the target word ranged from being the second to the second word from the last word in the sentence; mean word position is 4.9. For a subset of 32 sentences, two directly adjacent target words (a verb–noun or noun–verb sequence) set up a four-factorial minidesign with the frequency of the second target word as a fourth orthogonal factor in addition to word class, frequency, and length of first word; the additional target word of a sentence was of the same length as the first one. There were two sentences contributing to each cell of the 2 × 2 × 2 × 2 design. This subset of 32 sentences is listed in the Appendix; target words and second target words are set in italics. Analyses of the embedded design with additional target words as well as analyses of language-specific variables beyond frequency and predictability (e.g., word category, syntactic structure of sentence, spillover effects) are not reported here.

Procedure

Participants were calibrated with a standard nine-point grid for both eyes. They were instructed to read the sentence for comprehension and fixate a dot in the lower right corner of the monitor to signal the completion of a trial. After validation of calibration accuracy, a fixation point appeared on the left side of the centre line on the monitor. If the eye tracker identified a fixation on the fixation spot, a sentence was presented so that the midpoint between the beginning and the centre of the first word was positioned at the location of the fixation spot. Therefore, each sentence-initial word was read from a word-specific optimal viewing position (e.g., O'Regan & Lévy-Schoen, 1987; Rayner, 1979). Sentences were shown until participants looked to the lower right corner of the screen. Then, the sentence was replaced by (1) an easy three-alternative multiple-choice question pertaining to the current sentence on 27% of the trials which the participant answered with a mouse click, (2) a fixation spot indicating the beginning of the next trial which participants then initiated by fixating the fixation point, or (3) a complete recalibration with the nine-point grid after 15 sentences each. In addition, the experimenter carried out an extra calibration if the tracker did not detect the eye at the initial fixation point within 2 s.

Analyses

Eye movement data from reading the 144 sentences were screened for loss of measurement and blinks. Data from participants containing less than 100 sentences after data screening were excluded from the database. The remaining participants (33 young, 32 older adults) contributed on average 133 (young: 137, older adults: 127) sentences to the database. Data of sentences without problems were reduced to a fixation format after detecting saccades as rapid binocular eye movements with amplitudes of more than 0.5°, using a velocity-based detection algorithm originally developed for the analyses of microsaccades (Engbert & Kliegl, 2003). In a second level of data screening, words with first-fixation durations shorter than 30 ms or longer than 1 s and words with gaze durations or reading times longer than 1.5 s were removed from the database. This screening amounted to a total loss of 194 words (young: 101, older adults: 93; i.e., three words per participant). Finally, we removed the first word of each sentence. Overall an average of 92% of the words (relative to total number of corpus words minus 144 first words of sentences) remained in the analyses (young: 95%, older adults: 88%).

For each word we determined whether it was fixated once (P1), fixated more than once (P2), or skipped (P0) in first-pass reading. A word contributed to first-pass reading statistics as long as the eyes had not moved past it. Note that the three probabilities sum up to 1.0. For first-pass fixations we determined the duration of first fixation on a word (if fixated at least once, FF), the duration of single fixations (if fixated exactly once, SF), and gaze durations (i.e., the sum of fixations if fixated at least once, GD). In addition to first-pass reading, we counted how often words served as the origin (RO) and the goal (RG) of a regression back to a previous word of the sentence. The durations of fixations following a regression were added to the gaze duration for this word and yielded a measure of total reading time (TT). Fixations after the first encounter of the last word (i.e., re-readings) were not included in the analyses.

Statistical analysis followed the procedure described by Lorch and Myers (1990, method 3) for multiple regressions with repeated measures. Specifically, for each participant we regressed each dependent variable mentioned above on the following independent variables: linear and quadratic effects of word length (2–13+) centred for word length 7 (i.e., range: −5–6), logarithm (base 10) of CELEX frequency, and logit-transformed predictabilities (see above under Predictability). Subsequently, we checked whether means of unstandardised regression coefficients (across persons) were significantly different from zero and whether they differed between age groups.

RESULTS

Results on four inspection durations and five inspection probabilities are presented in two sections. First, we report the results based on multiple regression analyses of repeated measures of word length, frequency, and predictability for all corpus words. These analyses were carried out for all participants. Then, we compare the effects obtained for corpus words with target words representing an orthogonal design of word length and word frequency. The significance level for ANOVA effects was set at .001 in order to protect against isolated chance effects but we report effects that exhibit a meaningful pattern across measures for the .05 level. In general, as expected, first-fixation and single-fixation durations as well as gaze durations and total reading times exhibited very similar trends.

Corpus words

Means (standard deviations) of unstandardised regression coefficients are listed in Table 1 for the entire sample of 65 participants. Values reported under Constant represent the estimate for a seven-letter word with a printed frequency of zero and a 50% predictability. We present results for first-pass inspection probabilities (P0, P1, and P2+), for inspection durations (FF, SF, GD, TT) and

TABLE 1
Means (standard deviations) of unstandardised regression coefficients based on 65 participants for corpus words

	Constant	*Length*	*L-sq*	*Freq*	*Pred*
FF	214 (36)	*0.4* (1.9)	0.3 (0.6)	−4.5 (5.0)	*1.2* (7.7)
SF	213 (37)	1.6 (2.7)	0.6 (0.7)	−4.1 (5.9)	*−0.5* (8.0)
GD	247 (50)	9.1 (5.4)	2.0 (1.0)	−8.1 (7.2)	*−3.5* (12.4)
TT	253 (51)	9.1 (5.2)	1.9 (1.0)	−9.5 (7.5)	−7.3 (13.3)
PO	7.1 (10.1)	−4.6 (1.5)	0.8 (0.4)	3.0 (3.0)	1.7 (2.8)
P1	76.2 (9.0)	*0.4* (3.1)	−1.4 (0.4)	−1.8 (2.6)	*−0.1* (2.7)
P2	16.4 (9.2)	4.2 (2.0)	0.5 (0.3)	−1.2 (1.4)	−1.7 (2.3)
RO	7.3 (5.9)	*0.0* (0.6)	*0.0* (0.2)	−0.6 (1.1)	−1.8 (1.6)
RG	*0.2* (3.0)	−1.0 (0.8)	0.1 (0.2)	*0.5* (1.4)	−2.5 (0.2)

Length = −5 to 6+ letters (i.e., centred on seven-letter word); L-sq = quadratic trend of length; Freq(uency) (in lg10-units); Pred(ictability) (in log-odds units).

FF = first fixation duration, SF = single fixation duration, GD = gaze duration, TT = total reading time; P0, P1, P2 = probability of zero (skipping), one, two+ fixations; RO, RG = probability of origin, goal of regressive eye movement (multiplied with 100 in table).

Estimates are based on a mean of 913 words per participant. Italicised coefficients were *not* significant (99.9% confidence intervals).

for reinspection probabilities (RO, RG). The median R^2 of these multiple regressions was .07.

Inspection probabilities for first-pass reading. Figure 1 displays the main effects of word length, logarithmic word frequency, and logits of predictabilities for the probability of skipping a word (P0), fixating it once (P1), or fixating it two or more times (P2+). The lines are computed from regression equations with the unstandardised regression coefficients listed in Table 1. In general, results were consistent with expectations: Word skipping (P0) decreased with word

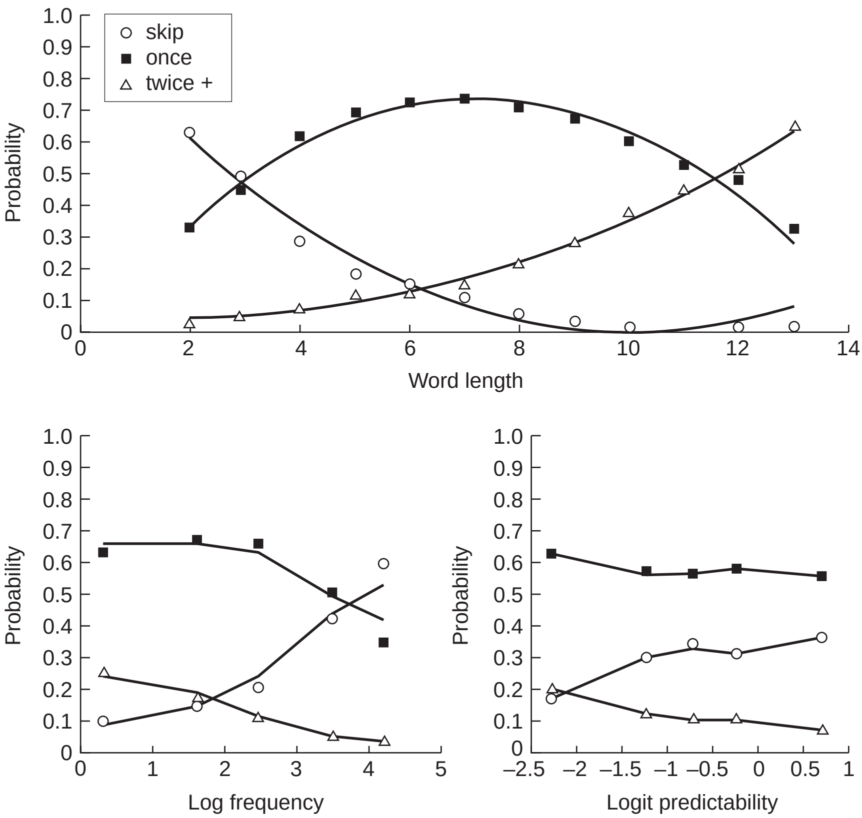

Figure 1. Observed and predicted first-pass inspection probabilities over word length (top), logarithmic word frequency (bottom left), and logit predictability (bottom right) based on all words of the corpus (except first words of sentences). Each curve was computed from the means of five regression coefficients (i.e., intercept, length, length-square, log frequency, and logit predictabilities) determined individually for 65 participants (see Table 1); they were not fitted directly to the observed means shown in symbols.

length and increased with frequency and predictability, whereas the reverse pattern was observed for multiple fixations (P2). Symbols in Figure 1 represent means across words of different lengths as well as frequency and predictability classes (see Methods for definition) and are plotted at the means of their respective classes. There is good agreement between the probabilities computed from regression equations (lines) and the observed means (symbols). The probability of a single fixation (P1) followed an inverted U-shape over word length (i.e., no significant linear trend), $F(1, 63) = 1.11$, $p = .296$, but as it is the complement of P0 and P2 (i.e., $p1 = 1 - p0 - p2$), it carries no independent information. Similarly, the effect of predictability was not significant for P1, $F(1, 63) = 0.05$, $p = .829$. There were no significant age differences in inspection probabilities. Thus, despite substantial correlations among the predictors, there were statistically independent and reliable effects of word length, frequency, and predictability on inspection probabilities.

Inspection durations. Figure 2 displays the main effects of word length, logarithmic word frequency, and logits of predictabilities for first-fixation duration (FF), single-fixation duration (SF), gaze duration (GD), and total reading time (TT). In general, durations increased with word length and decreased with frequency but predictability was not significantly associated with first-pass inspection durations after controlling for length and frequency (see Table 1); FF: $F(1, 63) = 1.64$, $p = .205$; SF: $F(1, 63) = 0.27$, $p = .605$; GD: $F(1, 63) = 5.0$, $p = .029$. Predictability was significant, however, once regressions to words were added in: TT: $F(1, 63) = 19.3$, $p < .001$. The linear effect of word length was not significant for first fixation durations, $F(1, 63) = 2.9$, $p = .093$. All other regression coefficients associated with length and frequency were significantly different from zero in the expected direction (all $F > 12.8$). Again, there was a very good agreement between estimates based on multiple regressions for individual participants (lines) and the observed mean durations computed at the level of words (symbols).

Age differences in inspection durations. Nominally, there were no age differences at the .001 level of significance for any of the regression coefficients but there was a very consistent pattern at the .01 level across the four inspection measures: Older adults read somewhat more slowly than young adults as reflected in larger constants: FF: 228 vs. 201 ms, SF: 227 vs. 200 ms, GD: 265 vs. 230 ms, TT: 271 vs. 234 ms, $F(1, 63) = 10.4$, 10.4, 9.0, and 9.5 for the age differences, respectively. Also, they were more sensitive to word frequency: FF: -6.4 vs. -2.8, SF: -6.0 vs. -2.3, GD: -10.3 vs. -5.9, TT: -12.0 vs. -7.2, $F(1, 63) = 9.6$, 7.0, 6.5, and 7.1 for the corresponding age differences. However, the frequency effects were significant within both groups.

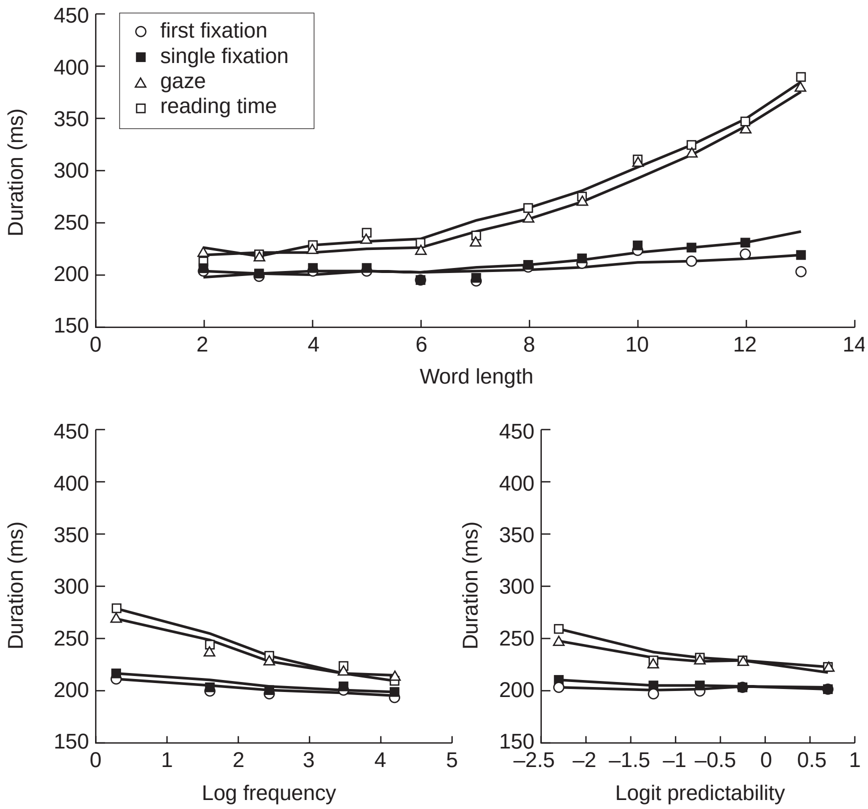

Figure 2. Observed and predicted inspection durations over word length (top), logarithmic word frequency (bottom left), and logit predictability (bottom right) based on all words of the corpus (except first words of sentences). Each curve was computed from the means of five regression coefficients (i.e., intercept, length, length-square, log frequency, and logit predictabilities) determined individually for 65 participants (see Table 1); they were not fitted directly to the observed means shown in symbols.

Reinspection probabilities. In addition to first-pass reading measures we also examined regressive movements (see RG and RO in Table 1). Regressions originated at words of low frequency and low predictability and landed on short words of low predictability (RO and RG in Table 1) (all $F > 40$). The absence of a frequency effect did not quite match our expectations given Vitu and McConkie's (2000) results of high regression probabilities to low-frequency skipped words. However, the fact that short words of low predictability were probable regression goals makes sense if these words were initially skipped due to their shortness but subsequently inspected due to their low predictability.

Sources of reinspection probabilities. For a better understanding of reinspection probabilities, we searched for regressive movements from word $n+1$ to a previously fixated word n (i.e., an average of 16 refixations per person contributed by 64 participants) and for regressive movements to a previously skipped word n (i.e., an average of 31 first fixations after regression per person contributed by 62 participants). Comparing lengths, frequencies, and predictabilities of word n with those of control words (i.e., averages of 528 and 196 cases per person without regressive movement to word n, respectively), we found refixated words to be shorter (5.5 vs. 5.8 letters), paired $t(63) = -3.0$, $p = .004$, $N = 64$, of lower log frequency (1.8 vs. 2.0), $t(63) = -2.0$, $p = .047$, and of lower logit predictability (−2.0 vs. −1.3), $t(63) = -14.2$, $p < .001$, $N = 64$. Previously skipped words targeted by a regressive movement from the next word differed only in logit predictability (−1.3 vs. −0.8), paired $t(61) = -14.7$, $p < .001$, $N = 62$; they were of similar length (3.7 vs. 3.8 letters), $t(61) = -1.1$, $p = .281$, and log frequency (2.92 vs. 2.90), $t(61) = 0.293$, $p = .770$.

Age differences in reinspection probabilities. Consistent with results reported for inspection durations, higher reinspection probabilities also contributed to a slower reading of older adults (means of 11% vs. 7% of all fixations), $t(63) = 3.5$, $p = .018$. However, old and young adults did not differ in the length, frequency, and predictability effects reported in the last paragraph (all $t < 1.5$). Thus, the strong predictability and relatively weak frequency and lengths effects replicated across age groups.

Target words

One well-known problem associated with analyses of corpus words is the correlation among predictors. Word length and log frequency of 994 corpus words (i.e., excluding the first word of each sentence) correlated −.64, length and logit predictability −.29, and log frequency and logit predictability .38. For a subset of 144 target words (one word per sentence), the corresponding correlations were −.01, −.19, and .31. Therefore, for this subset we could assess the independent contributions of length and frequency. Means (standard deviations) of unstandardised regression coefficients (averaged across 65 participants) for these analyses are summarised in Table 2. The median R^2 of these multiple regressions was .07.

Inspection probabilities. There were no effects of length (linear trends), frequency, or predictability on single-fixation probabilities (all Fs < 1.9). All other coefficients were in the expected direction and were highly significant (all Fs > 7.3). A comparison between corpus and target word analyses (i.e., Table 1 vs. Table 2) revealed mainly a reliable effect of frequency on single-fixation

TABLE 2
Means (standard deviations) of unstandardised regression coefficients based on 65 participants for target words

	Constant	*Length*	*L-sq*	*Freq*	*Pred*
FF	207 (36)	*1.5* (5.0)	*0.3* (1.9)	−5.4 (7.4)	−*3.2* (10.0)
SF	210 (38)	3.3 (6.0)	*0.5* (2.1)	−6.3 (8.1)	−5.3 (10.9)
GD	241 (49)	8.5 (9.7)	*0.5* (3.1)	−11.8 (11.4)	−10.3 (15.0)
TT	245 (48)	7.4 (10.4)	−*0.1* (3.4)	−14.5 (13.5)	−17.5 (17.4)
P0	9.1 (13.5)	−3.4 (2.6)	0.8 (0.9)	2.3 (2.8)	*1.8* (5.5)
P1	74.1 (17.2)	−*0.6* (5.6)	−1.0 (1.3)	*0.7* (3.8)	*0.9* (6.3)
P2	17.0 (12.2)	4.1 (4.2)	*0.2* (1.1)	−2.5 (2.9)	−2.6 (4.4)
RO	12.5 (14.7)	−*0.6* (1.8)	−*0.3* (0.8)	−*0.8* (2.9)	−*0.0* (3.9)
RG	0.4 (7.7)	−1.0 (1.9)	−*0.1* (0.6)	−*0.7* (2.5)	−3.7 (3.5)

Length = −5 to 6+ letters (i.e., centred on seven-letter word); L-sq = quadratic trend of length; Freq(uency) in lg10-units); Pred(ictability) (in log-odds units).

FF = first fixation duration, SF = single fixation duration, GD = gaze duration, TT = total reading time; P0, P1, P2 = probability of zero (skipping), one, two+ fixations; RO ($N = 63$), RG ($N = 63$) = probability of origin, goal of regressive eye movement (multiplied with 100 in table).

Estimates are based on a mean of 135 words per participant. Italicised coefficients were *not* significant (99.9% confidence intervals).

probability of corpus words and a stronger effect of frequency on multiple-fixation probability of target words. The overall pattern is quite similar.

Age differences in inspection probabilities. Young and old adults differed in the measure that showed susceptibility to the predictability of words: High predictability reduced young adults' single-fixation probability but increased that of old adults (young: −.023, old: .042), $F(1, 63) = 23.2$, $p < .001$; high predictability increased only young adults' probability of word skipping (young: .036, old: −.001), $F(1, 63) = 7.9$, $p = .007$; finally, high predictability decreased old adults' probability of multiple fixations (young: −.012, old: −.042, $F(1, 63) = 9.0$, $p = .004$. Thus, for both groups predictability allowed faster reading but they differed in how they realised the speed-up.

Inspection durations. With the exception of (1) the quadratic trends of word length, (2) the linear trend of word length for FF, and (3) the predictability for FF, regression coefficients for the prediction of inspection durations were significant at the .001 level (all Fs > 15.0). (The two nonsignificant effects associated with FF were in the expected direction at the .05 level.) A comparison between the corpus and target word analyses (i.e., Table 1 vs. Table 2) reveals reliable effects of predictability on SF and GD for target words only. The restriction in range of word length for target words (i.e., 3–9 letters) eliminated quadratic trends of word length observed for corpus words. Finally, there were no age differences associated with inspection durations of target words.

Reinspection probabilities. There were reliable effects of linear word length and predictability on the probability of a word to be selected as the goal of a regressive movement (all Fs > 19.9). (See Table 2 for corresponding unstandardised regression coefficients in lines RG and RO.) As in the corpus analyses, short words and low predictable words had a higher probability of being selected as a goal; different from the corpus analyses, however, low frequency was also marginally associated with a high selection probability for a regressive movement (RG in Table 2), $F(1, 62) = 4.3$, $p = .042$. Thus, for target words there is a weak effect of low word frequency and predictability on reinspection probability compatible with word identification problems as their source (Vitu & McConkie, 2000). There were no reliable frequency or predictability effects associated with the launch site of regressive movements (RO in Table 2).

DISCUSSION

We replicated and extended well-known effects of word length, frequency, and predictability for a standard set of eye movement measures within a coherent framework of repeated-measures multiple regression analyses covering data from corpus words as well as data from an a priori specified set of target words that were uncorrelated in length and frequency. We summarise our results in relation to first-pass and second-pass reading and the generalisability of effects from target to corpus words. Then we address the quite remarkable stability of these results across the adult life span, compare effect sizes associated with three predictors, and ''defend'' the selection of predictors for the present set of multiple regressions. Finally, we argue that the set of results presented here, extended with other features of the text material and related key findings of eye movement research, will serve a useful purpose in constraining computational models of eye movement control in reading.

First-pass reading

Inspection probabilities. First-pass reading was influenced by word length, frequency, and predictability of words. The reliability of effects could be established for corpus and target words. Word length decreased the word-skipping probability and increased multiple-fixation probability as in all other previous eye movement research. Also, consistent with previous research (e.g., meta-analysis by Brysbaert & Vitu, 1998), word frequency reliably increased skipping probability. Finally, predictability increased word skipping and decreased the associated multiple-fixation probability for corpus and target words.

Inspection durations. The three measures of first-pass inspection duration (FF, SF, GD) increased with word length in corpus and target words, in agreement with most previous research (see introduction). Quadratic trends of word length were reliable only for corpus words; probably this difference reflects the restriction in word length for target words. There were statistically reliable effects of word frequency on these measures in both sets of analyses. Finally, first-pass reading effects of predictability were observed only for single-fixation and gaze durations of target words.

Comparison of effects for corpus and target words. Can we generalise results from a select set of target words (i.e., the experimental control approach) to an unselected set of corpus words (i.e., the statistical control approach)? Given the similarity of coefficients in Tables 1 and 2, our answer is an almost unqualified ''yes''. For both corpus and target words, we obtained reliable effects of word length and word frequency on inspection probabilities (P0, P2+) as well as inspection durations. Also the effects of word predictability on inspection probabilities were very similar for corpus and target words, even in terms of the size of regression coefficients. Aside from the quadratic effect of word length, the main difference between analyses of corpus and target words related to effects of predictability on inspection durations that were reliable only for single-fixation durations and gaze durations of target words. Thus, counter to expectations, the main benefit of an a priori selection of target words was not the dissociation of word length and word frequency but the reliability of word predictability in first-pass reading. Regression coefficients for frequency, however, were consistently larger in the analyses of target compared to corpus words.

Second-pass reading and predictability effects

Second-pass reading effects were primarily linked to the predictability of words. This is quite plausible if we assume that words of low predictability set up small-scale garden-path effects that are known to trigger reanalysis and regressive movements (e.g., Rayner, 1998, for a review). We observed a much larger effect of predictability on total reading time than on gaze duration (-7.3 vs. -3.5 for corpus words and -17.5 vs. -10.3 for target words, respectively). These results are consistent with the expectation that low-predictable words were a primary target of regressive movements because differences between gaze duration and total reading time reflect the time spent in refixations. Vitu and McConkie (2000) reported their results for low-frequency words. In our data, after statistical control of word length and predictability, the frequency effect was not significant but it showed a trend in the expected direction for target words.

We attempted to extend the set of dependent measures with reinspection probabilities after first-pass reading and distinguished between the probabilities of a word for serving as the origin or the goal of a regressive eye movement. For the goal probability we had expected a tendency towards low-frequency and low-predictable words (Vitu & McConkie, 2000). The strongest effect was associated with low predictability, which showed the expected effect both for corpus and target words. Moreover, short words were more likely to be selected as a goal for a regressive movement in both analyses. Obviously, one may easily and inadvertently skip a short and low-predictable word. As a first attempt to look at the dynamics of this situation, we checked predictability, length, and frequency effects for regressions to the immediately preceding word, given that this word was fixated or skipped prior to the regression. Refixated words were less predictable, shorter, and of lower frequency than a set of control words; effect sizes were largest for predictability and weakest for frequency. In contrast, previously skipped words differed only in (low) predictability from their control words. Finally and in general, regressions originated from longer, less frequent, and less predictable words relative to the means of all corpus words.

Adult age differences

Easy reading yielded only minor effects of age. Older adults read somewhat slower than young adults as reflected in higher intercepts for inspection durations compatible with earlier reports of age-related slowing of saccadic latency (Abel, Troost, & Dell'Osso, 1983) and age-related slower reading (Solan, Feldman, & Tujak, 1995). Older adults were also more likely to reinspect an earlier word than young adults, again in agreement with results reported by Solan et al. (1995).

Aside from this general age-related speed effect, there were a few specific age differences related to effects of word frequency and predictability: Older adults responded earlier (i.e., in first and single fixation durations) and more consistently to word frequency. This second age difference related to predictability effects on inspection probabilities of target words: Young adults increased skipping probability (P0) with word predictability, whereas old adults reduced the probability of multiple fixations (P2+) for such words more than young adults. Obviously, both effects reflect a utilisation of predictability for an increase of reading speed.

Overall the similarities in the reading profiles, especially the similarity of first-pass inspection probabilities, of the two samples of adult readers, which differed an average of 47 years in age, are much more impressive than the differences between them. Obviously, it remains to be seen how age affects eye movement control when difficult sentences must be read, when difficult questions need to be answered, or when working memory load is manipulated simultaneously. The similarity in profile of young and older adults under the easy-reading conditions of the present experiment could serve as a very useful baseline against which specific effects of adult age (e.g., deficits suspected in executive control processes; Mayr, Spieler, & Kliegl, 2001) could manifest themselves.

Effect sizes

Figures 1 and 2 provide direct information about absolute effect sizes associated with word length, frequency, and predictability in units of the dependent variables relative to the range of predictor values. Obviously, word length affects inspection durations and probabilities much more than frequency and predictability. For example, for gaze durations the difference between a 13-letter and a 2-letter word was 373 ms − 252 ms = 121 ms (for log frequency = 0 and logit predictability = 0), whereas the difference between words of the lowest and highest frequency was 36 ms (for word length = 7 and logit predictability = 0) and the difference between words of zero and perfect predictability was 18 ms (for word length = 7 and log frequency = 0). In addition, unstandardised regression coefficients can be compared directly to this end. For example, from Tables 1 and 2 it is quite clear that the effects of frequency and predictability were about twice as large for aggregated measures (i.e., gaze duration, total reading time) than for first-fixation and single-fixation durations.[3]

We want to point out a limitation of the present set of analyses. We carried out separate multiple regression analyses for several measures of inspection probabilities and inspection durations (see Figures 1 and 2). Evidently, these measures are highly redundant. Most blatantly, the three inspection probabilities sum to one. There is also an obvious dependency between the probability of multiple fixations (P2+) and gaze duration. Finally, all inspection measures are correlated given their cumulative operational definitions—single fixations are a subset of first fixation durations, both are part of gaze duration, and gaze durations are part of total reading time. Strictly speaking, such technical dependencies prevent statistical tests of differences between the measures. We opted for the present analysis framework because it represents most of the measures used in the research community. In perspective, a coherent framework of nonredundant, independently defined dependent measures would be highly desirable.

Higher-order (multiplicative) predictor terms

Aside from linear terms of word length, log word frequency, and logit predictability, we included a quadratic term for word length as an additional higher order predictor in the regression equation. The need for this predictor was

[3] We restricted our comparison of effect sizes to unstandardised regression coefficients; for arguments against the interpretation of unique amounts of variance see Duncan (1975, pp. 63–66).

motivated by the visual inspection of word-length functions and clearly needed for an adequate account of the observed means (see Figures 1 and 2). In additional analyses we also included quadratic terms for the other two predictors as well as terms coding multiplicative interactions between them. With an average of 913 words (i.e., cases) in the multiple regressions, all of these terms were significant for at least one of the dependent variables but they did not lead to a qualitative improvement of the reproduction of the observed means as shown in Figures 1 and 2; mean incremental R^2s were always smaller than 1%. In addition, the collinearity between linear and higher order predictor terms greatly reduced the interpretability of associated unstandardised regression coefficients given in Tables 1 and 2. Thus, although based on a qualitative rather than a quantitative judgement of goodness of fit, we restricted the report of results to the four-predictor model.

There is a second argument for a small set of predictors. We did not expect major theoretical advances in our understanding of eye movement control during reading from fitting higher order polynomial regression models because, at this point in time, we simply have no good theoretical reasons to postulate specific quadratic or multiplicative effects. If these higher order predictors are significant, they mainly refine the description of our particular set of corpus words. Rather, we expect theoretical advances from the further development of inherently nonlinear computational models (see next section).

Finally, although the regression model comprised only linear and quadratic effects of word length as well as linear effects of word frequency and predictability, the regression lines in Figures 1 and 2 exhibit trends that are suggestive of higher order terms. For example, word-predictability curves for skipping probabilities are suggestive of a cubic term (i.e., they contain nonmonotonic segments), which actually traces the observed pattern of means. These effects were due to nonlinear relations between the predictors (e.g., words of low and high predictability were of lower frequency than words of medium predictability). Such nonlinear relations among predictors can generate disordinal trends.

Implications for computational models of eye movement control

Descriptive accounts of eye movement data from reading are needed as simulation targets for computational models of eye movement control. The data reported here are to serve as a complement to the corpus of Schilling et al. (1998, supplemented by Reichle et al., 1998). The present corpus represents a more comprehensive benchmark for the evaluation of computational models of eye movement control during reading because we report eye movements not only as a function of word frequency but also as a function of word length and as a function of word predictability. Moreover, there is information about independent effects of word length and word frequency for a subset of target words. In principle, the data pattern reported for this experiment is within reach of current theoretical proposals (see Reichle et al., in press, for a comparison of models).

There are a few challenges for future computational models. For example, it might be very difficult to reproduce the different pattern of word-frequency and word-predictability effects on inspection probabilities and inspection durations during first-pass reading of corpus and target words (see above). At the same time, to avoid overfitting a specific set of corpus words, it may be advantageous to fit models to length, frequency, and predictability functions based on a limited number of unstandardised regression coefficients rather than to means derived from post-hoc categorisations of frequency and predictability classes. Moreover, these functions should be based on a less redundant and technically dependent set of measures than the suite of inspection probabilities and durations presented here.

The present set of data represents only a small segment of eye movement measures distilled from the reading protocols for a select set of word characteristics. There are other features that can be determined for the present corpus of words, among them neighbourhood effects (Grainger, O'Regan, Jacobs, & Sequi, 1992) as well as initial trigram frequency and informativeness (Kennedy, Pynte, & Ducrot, 2002). Some of them appear to be relevant for a better understanding of dynamics of eye movement control and hypotheses about distributed lexical processing (e.g., parafoveal preview and spillover effects). To this end we also need to integrate knowledge about the distributions of landing sites within words, that is the observations that landing sites are normally distributed with a dependency of the mean on the launch distance of the last saccade (McConkie, Kerr, Reddix, & Zola, 1988; Radach & McConkie, 1998) and the associated effects on fixation durations (Vitu, McConkie, Kerr, & O'Regan, 2001). According to Rayner et al. (2001) these eye-position effects are dissociated from lexical processing, that is they are not influenced by word frequency or predictability. In summary, eye movement research has produced a rich and reliable set of results since the 1980s. It is time to document that harvest in a single database because we will need such a corpus for constraining the parameter space of extant and future computational models of eye guidance during reading.

PrEview proof published online September 2003

REFERENCES

Abel, L. A., Troost, B. T., & Dell'Osso, L. F. (1983). The effects of age on normal saccade characteristics and their variability. *Vision Research*, *23*, 33–37.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (Release 1)* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

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

Blanchard, H. E. (1985). A comparison of some processing measures based on eye movements. *Acta Psychologica*, *58*, 1–15.

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

Calvo, M. G., & Meseguer, E. (2002). Eye movements and processing stages in reading: Relative contribution of visual, lexical, and contextual factors. *Spanish Journal of Psychology*, *5*, 66–77.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Duncan, O. D. (1975). *Introduction to structural equation models.* New York: Academic Press.

Engbert, R., & Kliegl, R. (2001). Mathematical models of eye movements in reading: A possible role for autonomous saccades. *Biological Cybernetics*, *85*, 77–87.

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, *43*, 1035–1045.

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

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*, 2517–2531.

Grainger, J., O'Regan, K., Jacobs, A. M., & Sequi, J. (1992). Neighborhood frequency effects and letter visibility in visual word recognition. *Perception and Psychophysics*, *51*, 49–56.

Henderson, J. M., & Ferreira, F. (1993). Eye movement control during reading: Fixation measures reflect foveal but not parafoveal processing difficulty. *Canadian Journal of Experimental Psychology*, *47*, 201–222.

Hyönä, J., & Olson, R. K. (1995). Eye movement patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1430–1440.

Inhoff, A. W., Radach, R., Eiter, B. M., & Juhasz, B. (2003). Distinct subsystems for the parafoveal processing of spatial and linguistic information during eye fixations in reading. *Quarterly Journal of Experimental Psychology*, *56A*, 803–827.

Inhoff, A. W., & Eiter, B. (2003). Knowledge of word length does not constrain word identification. *Psychological Research*, *67*, 1–9.

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

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

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

Kliegl, R., Olson, R. K., & Davidson, B. J. (1982). Regression analyses as a tool for studying reading processes: Comments on Just and Carpenter's eye fixation theory. *Memory and Cognition*, *10*, 287–296.

Kliegl, R., Olson, R. K., & Davidson, B. J. (1983). On problems of unconfounding perceptual and language processes. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 333–343). New York: Academic Press.

Lehrl, S. (1977). *Mehrfach-Wortschatz-Test Form B* [Multiple-choice vocabulary test]. Erlangen, Germany: Perimed.

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

Mayr, U., Spieler, D., & Kliegl, R. (Eds.). (2001). Ageing and executive control. *European Journal of Cognitive Psychology, 13* (Whole no. 1/2).

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., & Lévy-Schoen, A. (1987). Eye movement strategy and tactics in word recognition and reading. *Attention and Performance*, *12*, 363–383.

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* (pp. 165–192). Oxford, UK: Elsevier.

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

Rayner, K. (1979). Eye guidance in reading: Fixation location 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., Binder, K. S., Ashby, J., & Pollatsek, A. (2001). Eye movement control in reading: Word predictability has little influence on initial landing positions in words. *Vision Research*, *41*, 943–954.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 1188–1200.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraints on eye movements in reading: A further examination. *Psychonomic Bulletin and 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.

Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision Research*, *39*, 4403–4411.

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

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory and Cognition*, *26*, 1270–1281.

Solan, H. A., Feldman, J., & Tujak, L. (1995). Developing visual and reading efficiency in older adults. *Optometry and Vision Science*, *72*, 139–145.

Vitu, F., & McConkie, G. W. (2000). Regressive saccades and word perception in adult reading. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process (*pp. 301–326). Amsterdam: Elsevier.

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

Wechsler, D. (1964). *Der Hamburg Wechsler Intelligenztest für Erwachsene (HAWIE).* Bern, Switzerland: Huber.

APPENDIX

1 Den *Ton gab* der Künstler seinem Gehilfen.
2 Der *Hof lag* weit außerhalb des eigentlichen Dorfes.
3 Die Wanderer sahen Rehe auf einer Lichtung im *Wald äsen*.
4 Den *Kopf hieb* man früher nur Mördern und Verrätern ab.
5 Vorne am *Bug sah* man eine prächtige Galionsfigur.
6 Sogar aus *Raps läßt* sich Kraftstoff herstellen.
7 Torsten beobachtete gestern eine Maus, die *Efeu fraß*.
8 Der schüchterne kleine *Gnom mied* die Nähe der Elfen.
9 Claudia hatte ihr Fahrrad auf der *Straße stehen* lassen.
10 Wir hätten schon vor einer *Stunde wissen* sollen, ob ihr kommt.
11 Die Eltern konnten ihre Kinder im *Garten raufen* hören.
12 Er hätte nicht auch noch am *Telefon nörgeln* sollen.
13 Wegen ihrer Diät hatte die Gräfin leider keine *Auster nehmen* dürfen.
14 Die meisten *Hamster bleiben* bei Tag in ihrem Häuschen.
15 Man sollte nie Geschirr mit einem dreckigen *Lappen spülen* müssen.
16 Man kann *Spargel dämpfen* oder in viel Wasser kochen.
17 Manchmal *sagen Opfer* vor Gericht nicht die volle Wahrheit.
18 Die meisten Befragten *hören Musik* zur Entspannung.
19 Kinder *essen Quark* am Liebsten mit Früchten.
20 Bei Wölfen *leben Rudel* nicht verwandter Tiere in getrennten Revieren.
21 Die Frauen in den Andendörfern *weben Stoff* noch auf traditionellen Webstühlen.
22 Die Platzwarte *ebnen Stück* für Stück den Rasen nach dem Spiel.
23 In den Fässern *gären Beize* und Lauge.
24 Die Förster *küren Ahorn* zum Baum des Jahres.
25 Wolfgangs Töchter *studieren Literatur* und Maschinenbau.
26 In der Klosterschule *herrschen Schwester* Agathe und Schwester Maria.
27 Hier *scheinen Klempner* am Werk zu sein.
28 Im Aussehen *gleichen Bratsche* und Geige sich sehr.
29 Angeblich *flunkern Künstler* oft bezüglich ihrer Einnahmen.
30 Manchmal *krakeelen Politiker* genauso wie Demonstranten.
31 Die Armen *plündern Speicher* und Vorratskeller der reichen Bauern.
32 Die Richter der Landwirtschaftsschau *prämieren Rhabarber* und Mangold.