

available at [www.sciencedirect.com](http://www.sciencedirect.com)[www.elsevier.com/locate/brainres](http://www.elsevier.com/locate/brainres)


---



---

**BRAIN  
RESEARCH**


---



---



---

**Research Report**
**Frequency and predictability effects on event-related potentials during reading**
**Michael Dambacher<sup>a,\*</sup>, Reinhold Kliegl<sup>a</sup>, Markus Hofmann<sup>b</sup>, Arthur M. Jacobs<sup>b</sup>**
<sup>a</sup>Helmholtz Center for the Study of Mind and Brain Dynamics, Department of Psychology, University of Potsdam, P.O. Box 60 15 53, 14451 Potsdam, Germany

<sup>b</sup>Department of Psychology, Freie Universität Berlin, Germany

---

**ARTICLE INFO**
**Article history:**

Accepted 7 February 2006

Available online 20 March 2006

---

**Keywords:**

Event-related potentials

Word frequency

Word predictability

Reading

Lexical access

Repeated measures multiple regression analysis

**Abbreviations:**

rmMRA, repeated measures multiple regression analysis

rmANOVA, repeated measures analysis of variance

---

**ABSTRACT**

Effects of frequency, predictability, and position of words on event-related potentials were assessed during word-by-word sentence reading in 48 subjects in an early and in a late time window corresponding to P200 and N400. Repeated measures multiple regression analyses revealed a P200 effect in the high-frequency range; also the P200 was larger on words at the beginning and end of sentences than on words in the middle of sentences (i.e., a quadratic effect of word position). Predictability strongly affected the N400 component; the effect was stronger for low than for high-frequency words. The P200 frequency effect indicates that high-frequency words are lexically accessed very fast, independent of context information. Effects on the N400 suggest that predictability strongly moderates the late access especially of low-frequency words. Thus, contextual facilitation on the N400 appears to reflect both lexical and post-lexical stages of word recognition, questioning a strict classification into lexical and post-lexical processes.

© 2006 Published by Elsevier B.V.

---

**1. Introduction**

The frequency of words and their predictability in the context of a given sentence are two of the strongest determinants influencing reading. Despite much research, the role of word frequency as an indicator of ease of lexical access and of word predictability as an indicator of ease of semantic processing or of post-lexical integration, as well as the interaction of these two variables, are not yet well understood. Here, we report timelines of these effects as revealed in early (P200) and late

(N400) event-related potentials (ERPs) which were measured on open-class words in a sentence-reading experiment.

Word frequency (i.e., the printed frequency of a word in a text corpus) is well known to affect the speed of word identification. Readers take longer to recognize low than high-frequency words (e.g., Forster and Chambers, 1973; Rubenstein et al., 1970). Eye movement research corroborated this finding, revealing longer fixations on low than on high-frequency words (e.g., Inhoff and Rayner, 1986; Kliegl et al., 2004, 2006; Rayner and Duffy, 1986; Schilling et al., 1998).

---

\* Corresponding author. Fax: +49 331 977 2793.

E-mail address: [dambach@uni-potsdam.de](mailto:dambach@uni-potsdam.de) (M. Dambacher).

Also, word predictability or cloze probability (i.e., the proportion of subjects that fill in a particular word as the most probable next word in a sentence) influences word recognition. Reaction times (e.g., Fischler and Bloom, 1979; Kleiman, 1980) as well as fixation or gaze durations during natural reading (e.g., Kliegl et al., 2004, 2006; Rayner and Well, 1996; Rayner et al., 2001) are shorter for high than for low predictable words.

Despite an agreement on independent contributions of frequency and predictability to word recognition, there are conflicting theoretical perspectives on the exact time course and interaction of the two variables. In general, lexical access (i.e., the moment, when an orthographic word form uniquely activates the corresponding representation in the mental lexicon and therefore is identified) is assumed to be fast and automatic, whereas post-lexical integration is presumably a much slower process. Word frequency has served as one of the prime indicators of difficulty in lexical access (e.g., Hudson and Bergman, 1985; Monsell et al., 1989) and is one of the key factors constraining models of word recognition (Grainger and Jacobs, 1996; Jacobs and Grainger, 1994). In contrast, there is some controversy about whether predictability affects word recognition at an early stage, at the moment of lexical access, or whether it only influences post-lexical levels, like semantic integration. These perspectives are reflected in different implementations of lexical and contextual information in models of language comprehension:

In modular approaches (e.g., Fodor, 1983; Forster, 1979), functionally independent lexical subsystems are assumed to activate word representations by bottom-up processing, whereas context merely affects post-lexical integration processes. Consequently, these approaches do not predict interactions between frequency and context. In contrast, interactive activation models (e.g., McClelland, 1987; Morton, 1969) allow interactions between these two variables: both frequency and context may affect early stages in word recognition.

Experimental evidence relating to this theoretical distinction has not been consistent. Context was shown to facilitate recognition of low-frequency words stronger than recognition of high-frequency words (e.g., Becker, 1979), but purely additive effects have been reported as well (e.g., Schuberth et al., 1981). In eye movement measures, frequency and predictability generally did not interact reliably although there were some deviations from additivity (for review, see Rayner et al., 2004). In summary, while there is strong evidence for the relevance of frequency and predictability on language comprehension, it has not been resolved whether they link specifically to temporally distinct processes of lexical access and post-lexical integration.

### 1.1. Frequency and predictability in ERPs

ERPs can be used to delineate the time course of word recognition because they provide an online measure of neural activity with a high temporal resolution (Kutas and Van Petten, 1994). The first occurrence of a frequency effect in ERPs serves as an upper time limit for lexical access (Hauk and Pulvermüller, 2004). ERP differences after this point are often interpreted as post-lexical. Several researchers reported frequency effects in the time range of approximately 400 ms after stimulus onset (N400, see below; e.g., Rugg, 1990; Van

Petten and Kutas, 1990). However, the eyes of a skilled reader usually rest for less than 250 ms on a word before they move on to the next word; therefore, some part of lexical access is likely to occur during this typical fixation duration (Serenio et al., 1998). Indeed, Serenio et al. obtained a word frequency effect as early as 132 ms post-stimulus in an ERP study. Similarly, results of a single-case MEG study revealed a frequency effect for short words in a window from 120 to 160 ms and for all word lengths between 240 and 290 ms (Assadollahi and Pulvermüller, 2001). Hauk and Pulvermüller (2004) reported smaller amplitudes for high-frequency than for low-frequency words in an epoch from 150 to 190 ms. In summary, lexical access as indicated by word frequency effects appears to occur within the first 200 ms after stimulus presentation, but there is also evidence for temporally later influence of word frequency.

Context effects in ERPs were predominantly found on the N400 component, a negative deflection occurring in a time range between 200 and 500 ms after stimulus presentation. It is largest over centro-parietal sites, although it can be observed across the whole scalp (Coulson and Federmeier, *in press*; for review, see Kutas and Federmeier, 2000; Kutas and Van Petten, 1994). The N400 was described first by Kutas and Hillyard (1980). They presented sentences with final words that were semantically congruent or incongruent with the preceding context. Semantically incongruent words elicited a large N400. The sensitivity of the N400, however, is not constrained to anomalous words within a context; its amplitude correlates negatively with predictability (Kutas and Hillyard, 1984; Kutas and Van Petten, 1994). Moreover, Kutas and Hillyard (1983) reported N400s for positions other than final ones with larger amplitudes for earlier than later word positions.

Serenio et al. (2003) investigated effects of word frequency and context effects on an early ERP component. Ambiguous words with a low- and a high-frequency meaning were used as final words in sentences. The context of the preceding sentence fragment was either neutral or biased the low-frequency meaning. The neutral context should activate the dominant high-frequency meaning of the final word. In contrast, the subordinate low-frequency meaning should only play a role in the biasing context. In a time window from 132 to 192 ms post-stimulus, ambiguous words in a biasing context elicited amplitudes similar to those of low-frequency words, whereas in a neutral context, amplitudes resembled those of high-frequency words. Thus, a biasing context selectively activated the subordinate meaning of an ambiguous word and marginally facilitated low-frequency but not high-frequency words. The authors concluded that this pattern of results provides evidence for an early influence of context on lexical stages in word recognition.

The relation between word frequency and context was also addressed by Van Petten and Kutas (1990; see also Van Petten and Kutas, 1991; Van Petten, 1993). They categorized open-class words (nouns, verbs, adjectives, and “ly” adverbs) according to their frequency. Cloze-probability values were available for the terminal words in each sentence. For the remaining words, the position in a sentence was taken as a proxy of contextual support. The authors reported three main results on the N400. First, amplitudes were larger for low-frequency than for high-frequency words. Second, N400 amplitudes decreased with increasing position, presumably

reflecting the build-up of context “online”. Third, low-frequency words elicited a larger N400 than high-frequency words only if they occurred early in the sentence, not at later positions. The authors considered this finding as “evidence that frequency does not play a mandatory role in word recognition but can be superseded by the contextual constraint provided by a sentence” (Van Petten and Kutas, 1990, p. 380). Premise for this argument is that the N400 reflects lexical processes. However, there is disagreement concerning the temporal nature of N400 effects: some experimental results indicated that the amplitude is modulated by lexical processes (e.g., Besson et al., 1992; Deacon et al., 2000); other studies argued that the N400 is sensitive to post-lexical integration (e.g., Brown and Hagoort, 1993; Holcomb, 1993).

In summary, the question of timelines associated with lexical access and post-lexical integration during reading still requires further investigation. Frequency plays an important role in lexical access but apparently also modulates temporally later ERP components like the N400. Predictability (or, alternatively, position of word in sentence) correlates with the N400 amplitude but also with word recognition processes on early components. Interactions of these variables have also been described early and late in the ERP time course. However, these effects have been assembled across several experiments. To our knowledge, there is no study yet which examined lexical access and post-lexical integration during reading with independent measures of frequency, predictability, and word position in early and late ERP components.

## 1.2. Present study

In the present study, a corpus of 144 sentences (1138 words) was used as stimulus set. Values for frequency and predictability were available for all corpus words, along with other independent variables such as word length and ordinal position of the word in the sentence. To our knowledge, there exist only two sets of sentences with predictability norms for all words (i.e., Kliegl et al., 2004; Schilling et al., 1998, augmented by Reichle et al., 1998).

We tested effects of word frequency, predictability, and position in sentence, as well as the interactions between these variables, in early and late stages of word recognition using single-trial EEG amplitudes as dependent variables. This design allows us to go beyond previous research in at least two respects. First, we assume that predictability is a more direct measure of the contribution of sentence context than word position. Therefore, we hypothesized that, irrespective of the position of the word in the sentence, frequency, and predictability would interact on the N400 as previously was shown for frequency and position. Second, we expected that the decrease of N400 amplitudes across word position would be attributable to the build-up of contextual information as proposed by Van Petten and Kutas (1990, 1991). If predictability completely accounts for context-related variance in ERPs, there should be no unique variance associated with word position after statistical control for the effects of predictability. In other words, predictability should absorb all N400 effects associated with word position but not vice versa.

We examined the data using repeated measures multiple regression analyses (rmMRAs; Lorch and Myers, 1990, method

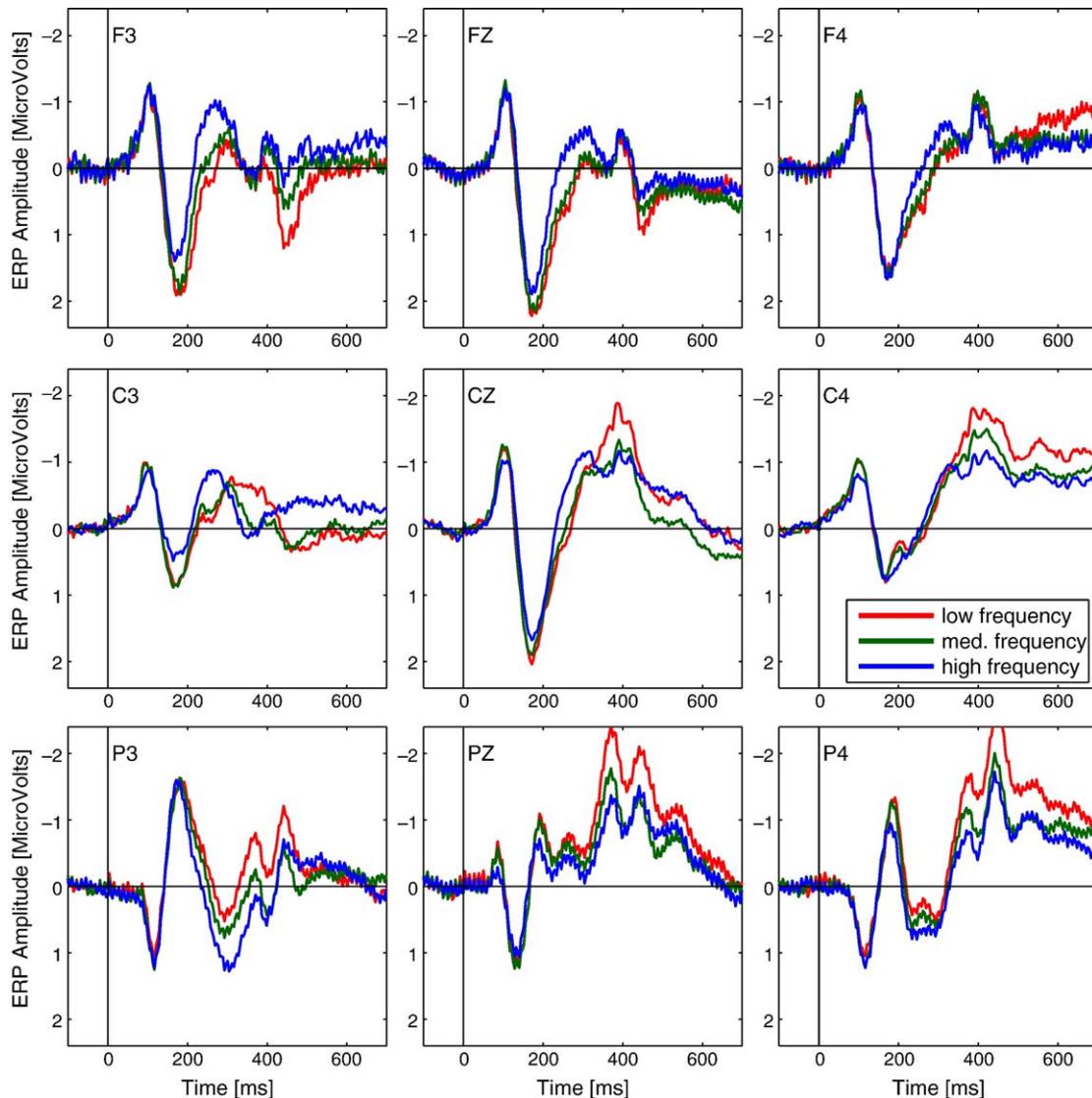
3; see Kliegl et al., 2006, for a recent application to the analyses of eye movements in reading) in an early (P200) and a late (N400) time window. Mean EEG amplitudes were computed within these time windows (collapsed across sampling points and selected electrodes for the components) for each word within each subject. These single-trial EEG amplitudes served as criterion in the rmMRAs. An advantage of this procedure is that rmMRAs statistically control for differences between participants. Then, after between-subject variance has been removed, effects of variables such as frequency, predictability, and word position as well as their interactions can be estimated within one single model statistically controlling for correlations between the predictors. Since predictors need not be divided into discrete categories but can be submitted to the models as continuous values, the whole variability of word properties mapping on the dependent variable is used. Using EEG amplitudes on a single-trial level instead of values collapsed across many items provides information of electrophysiological correlates as a function of different properties of single words. Furthermore, the large amount of data points yields high statistical power. However, waiving data averaging results in a loss of noise reduction. Thus, necessarily the variance accounted for by rmMRA models on single-trial EEG amplitudes is very small.

We limited our analyses to open-class words, i.e., nouns, verbs, adjectives, and most of the adverbs. Closed-class words, like auxiliary verbs, pronouns, conjunctions, and determiners, were excluded. This restriction was motivated by findings suggesting that words of different classes are processed by distinct neural systems, because open-class and closed-class words evoke different ERP components. For instance, an N280 component was elicited only by closed-class words, whereas open-class words evoked an N400 (Neville et al., 1992). However, this issue is discussed controversially. Results of other studies revealed that differences between word classes do not reflect qualitatively separate processing mechanisms but rather are a function of word frequency or of frequency and length (e.g., King and Kutas, 1998; Münte et al., 2001; Osterhout et al., 1997).

Another restriction was the exclusion of sentence-final words. Previous studies revealed that ERPs for sentence-final words differ from those of words occurring earlier in a sentence. They often appear to evoke more positive-going ERPs than sentence-intermediate words (e.g., Friedman et al., 1975; Osterhout and Holcomb, 1995; Osterhout, 1997; see also Kutas et al., 1988; Van Petten, 1993). This effect can most probably be attributed to sentence wrap-up, decision, and/or response and reduces the comparability of ERPs of sentence-intermediate and sentence-final words (Hagoort, 2003; Osterhout and Nicol, 1999).

## 2. Results

Grand-average plots for open-class words are presented in Figs. 1 and 2 illustrating the effects for three frequency classes and three predictability classes, respectively. A small negativity, peaking at 100 ms, was followed by a large positive deflection reaching its maximum amplitude 170 ms after stimulus onset (P200). At this latency, differences in ERPs for



**Fig. 1 – Frequency grand averages.** Grand average plots of effects of three frequency classes for open-class words in sentences comprising seven to nine words; sentence-final words are excluded. The three classes are based on categories of Table 3. Amplitude differences are visible on the P200 predominantly over fronto-central electrodes on the left hemisphere.

word frequency are visible on fronto-central electrodes predominantly on the left hemisphere. After about 260 ms, a negative deflection occurred mainly over centro-occipital electrode sites peaking at a latency of approximately 400 ms (N400). During this epoch, grand average curves of predictability classes are gradually arranged with larger amplitudes for words of low than of high predictability classes.

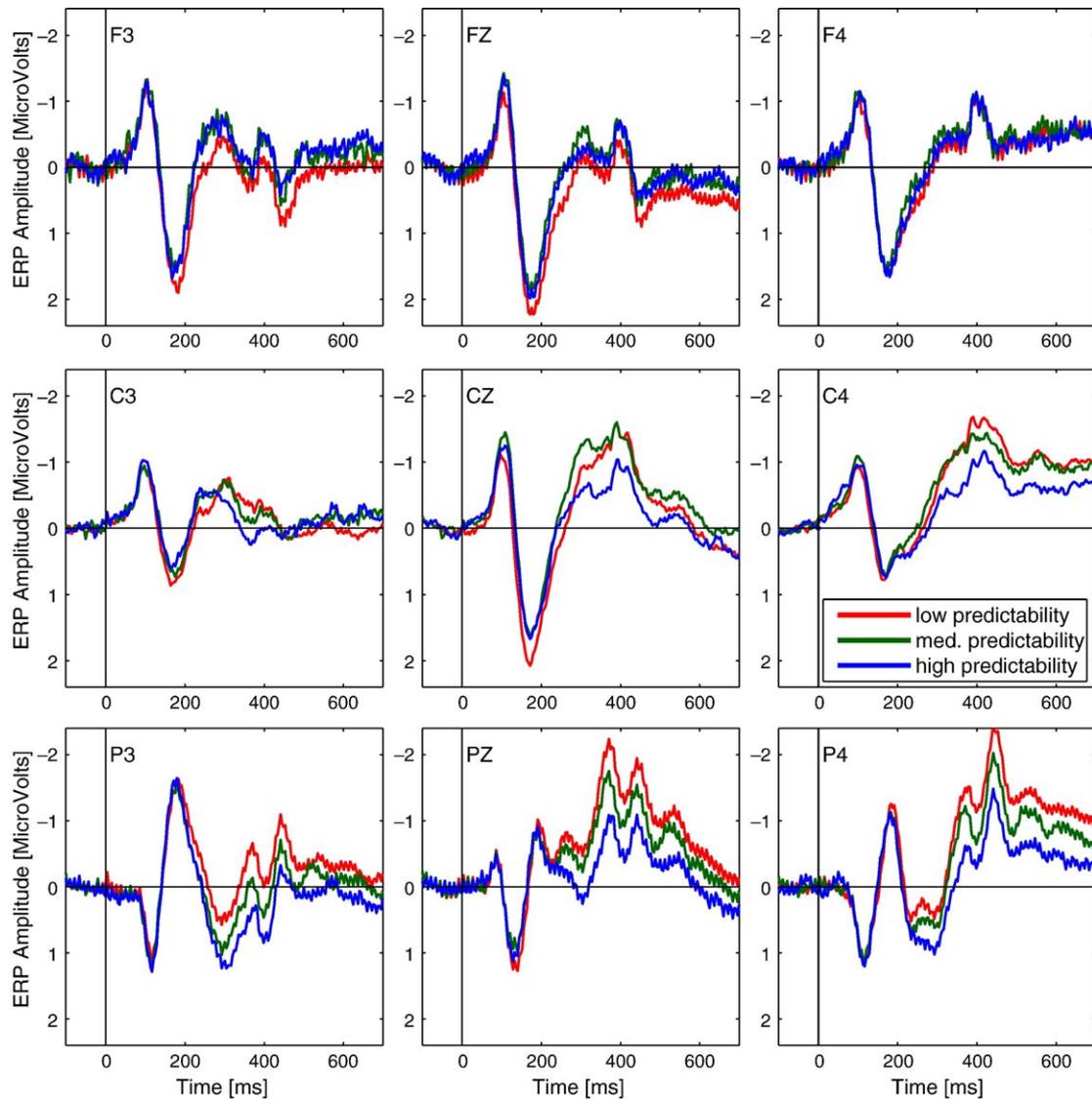
### 2.1. P200

Effects of frequency, predictability, and position on P200 amplitudes were examined in two separate  $3 \times 3$  repeated measures analyses of variance (rmANOVAs). The first rmANOVA with frequency and predictability as within-subject factors revealed significant results for frequency [ $F(2,94) = 6.52, P < 0.01$ , partial  $\eta^2 = 0.12$ ] and predictability [ $F(1,92) = 4.33, P = 0.02$ , partial  $\eta^2 = 0.08$ ]. The interaction between predictability and frequency was not reliable [ $F(3,177) = 0.63, P = 0.63$ , partial  $\eta^2 = 0.01$ ].

The second rmANOVA comprised frequency and position as factors. Again, the main effects were significant [Frequency:  $F(1,79) = 11.79, P < 0.01$ , partial  $\eta^2 = 0.20$ ; Position  $F(1,75) = 13.03, P < 0.01$ , partial  $\eta^2 = 0.22$ ], whereas the interaction was not [Frequency  $\times$  Position:  $F(3,166) = 0.94, P = 0.43$ , partial  $\eta^2 = 0.02$ ].

The effects of frequency, predictability, and position were scrutinized within a single rmMRA model. The regression coefficients of the rmMRA for open-class words on the P200 are listed in Table 1. They are the mean of the unstandardized regression coefficients calculated separately for each subject (Lorch and Myers, 1990, method 3, individual regression equations). Moreover, Table 1 lists standard errors of regression coefficients, the drop of  $R^2$  for removing the predictor from the complete model, as well as probabilities of significance tests for the regression coefficients and the  $R^2$  decrement.

The effects of predictors are visualized in Fig. 3. Open symbols reflect the mean of empirical ERP amplitudes in the time range from 140 to 200 ms post-stimulus. Bins in the plots



**Fig. 2 – Predictability grand averages.** Grand average plots of effects of three predictability classes for open-class words in sentences comprising seven to nine words; sentence final words are excluded. The three classes are based on categories of Table 3. Amplitudes are graded on the N400 over centro-occipital electrodes.

for frequency and predictability (panels 1 and 2) were computed on the basis of predictor quantiles ensuring a similar number of data points for each category. Categories for frequency and predictability in the interaction plots (panels 4 and 5) correspond to classes in Table 3. Error bars reflect 99% within-subject confidence intervals (Loftus and Masson, 1994). Raw correlations between predictor and the criterion are given in parentheses as supplementary information along with the description of the results.

P200 amplitudes were smaller for high than for low-frequency words (panel 1). The quadratic frequency term ( $r = -0.045$ ) was significant, whereas the linear ( $r = -0.040$ ) was not. Amplitude differences were larger among three high-frequency bins than among those of low-frequency words. That means the size of the frequency effect increased with augmenting frequency. Consequently, the quadratic trend accounted for a larger amount of unique variance than the linear trend.

The predictors accounting for most of the unique variance in P200 amplitudes were linear and quadratic terms of word position ( $r = -0.044$  and  $r = -0.034$ , respectively). Amplitudes decreased during early positions in a sentence, reached a minimum around the middle position (5th word), and started to increase again towards the end of the sentence (panel 3). This is an unexpected and, as far as we know, novel result.

Neither predictability (panel 2;  $r = -0.029$ ) nor the interaction of predictability and frequency (panel 4;  $r = 0.007$ ), nor the interaction of position and frequency (panel 5;  $r = -0.053$ ) were significant in the rmMRA model for the P200.

## 2.2. N400

Like on the P200, two rmANOVAs were carried out to examine effects on N400 amplitudes. In the first rmANOVA with frequency and predictability as within-subject factors, predictability [ $F(1,89) = 24.21$ ,  $P < 0.01$ , partial  $\eta^2 = 0.34$ ] and the

**Table 1 – Mean and standard errors (SE) of regression coefficients of the rmMRA for ERP amplitudes of open-class words in the time window 140–200 ms at fronto-central electrode sites**

P200 (7 predictors)	Mean	SE	t	p <sub>t</sub>	–ΔR <sup>2</sup>	p <sub>–ΔR<sup>2</sup></sub>
Constant	1.134	0.128	8.83	<0.01		
Frequency	0.074	0.090	0.82	0.21	<0.0001	0.46
Frequency <sup>2</sup>	–0.030	0.015	–1.96	0.03	0.0002	0.05
Predictability	0.004	0.051	0.09	0.47	<0.0001	0.92
Position	–0.293	0.058	–5.08	<0.01	0.0018	<0.01
Position <sup>2</sup>	0.029	0.006	4.74	<0.01	0.0014	<0.01
Predictability × Frequency	0.003	0.021	0.16	0.44	<0.0001	0.84
Position × Frequency	–0.007	0.009	–0.84	0.20	<0.0001	0.47
[R <sup>2</sup> <sub>Predictors</sub> = 0.005; R <sup>2</sup> <sub>Subjects</sub> = 0.110; R <sup>2</sup> <sub>Model</sub> = 0.115]						

Note. Means, SE, t values, and associated P values for predictors. –ΔR<sup>2</sup> is the drop of variance of the full model due to removal of the predictor; p<sub>–ΔR<sup>2</sup></sub> gives P values for the significance of the variance decrement. R<sup>2</sup><sub>Predictors</sub>, R<sup>2</sup><sub>Subjects</sub>, and R<sup>2</sup><sub>Model</sub> show variance accounted for by predictors alone, by subjects alone, and by the full model, respectively. Statistics are based on 48 subjects, i.e., 47 degrees of freedom for t statistics.

interaction between predictability and frequency [ $F(2,125) = 2.94$ ,  $P = 0.04$ , partial  $\eta^2 = 0.06$ ] were reliable. Frequency was marginally significant [ $F(1,71) = 3.39$ ,  $P = 0.05$ , partial  $\eta^2 = 0.07$ ].

The second rmANOVA with the factors frequency and position revealed significant effects for frequency [ $F(1,68) = 24.86$ ,  $P < 0.01$ , partial  $\eta^2 = 0.35$ ] and the interaction between position and frequency [ $F(3,168) = 2.60$ ,  $P = 0.04$ , partial  $\eta^2 = 0.05$ ]. Word position yielded a trend [ $F(1,78) = 2.89$ ,  $P = 0.07$ , partial  $\eta^2 = 0.06$ ].

The effects in the rmANOVAs on the N400 were scrutinized in rmMRAs. The results are listed in Table 2 showing unstandardized regression coefficients, along with associated standard errors, the drop of R<sup>2</sup> for removing the predictor from the model, and probabilities of significance tests for the regression coefficients and the R<sup>2</sup> decrement. Fig. 4 presents a visualization of the effects; error bars reflect 99% within-subject confidence intervals (Loftus and Masson, 1994).

In the first rmMRA [Table 2; 1: N400 (6 predictors)], the strongest predictor for the N400 was predictability ( $r = 0.077$ ). Panel 2 shows that amplitudes decreased substantially with increasing predictability.

The interaction of predictability and frequency ( $r = -0.006$ ) was also reliable. Panel 4 reveals a larger predictability effect for words of low than of high frequency.

The interaction of position and frequency (panel 5;  $r = 0.057$ ) was not significant in the rmMRA. However, the pattern of means corresponded to previous reports: The frequency effect was strong at early positions and became weaker across the sentence.

Neither the linear ( $r = 0.066$ ) nor the quadratic terms ( $r = 0.056$ ) of frequency (panel 1), nor word position ( $r = 0.027$ , panel 3) reached significance.

To test whether the interaction of predictability and frequency absorbed variance of other predictors, we carried out a second rmMRA without this interaction term. The results of this five-predictor model are listed in Table 2 [2: N400 (5 predictors)]. In this model, predictability still accounted for the largest amount of variance and was highly reliable.

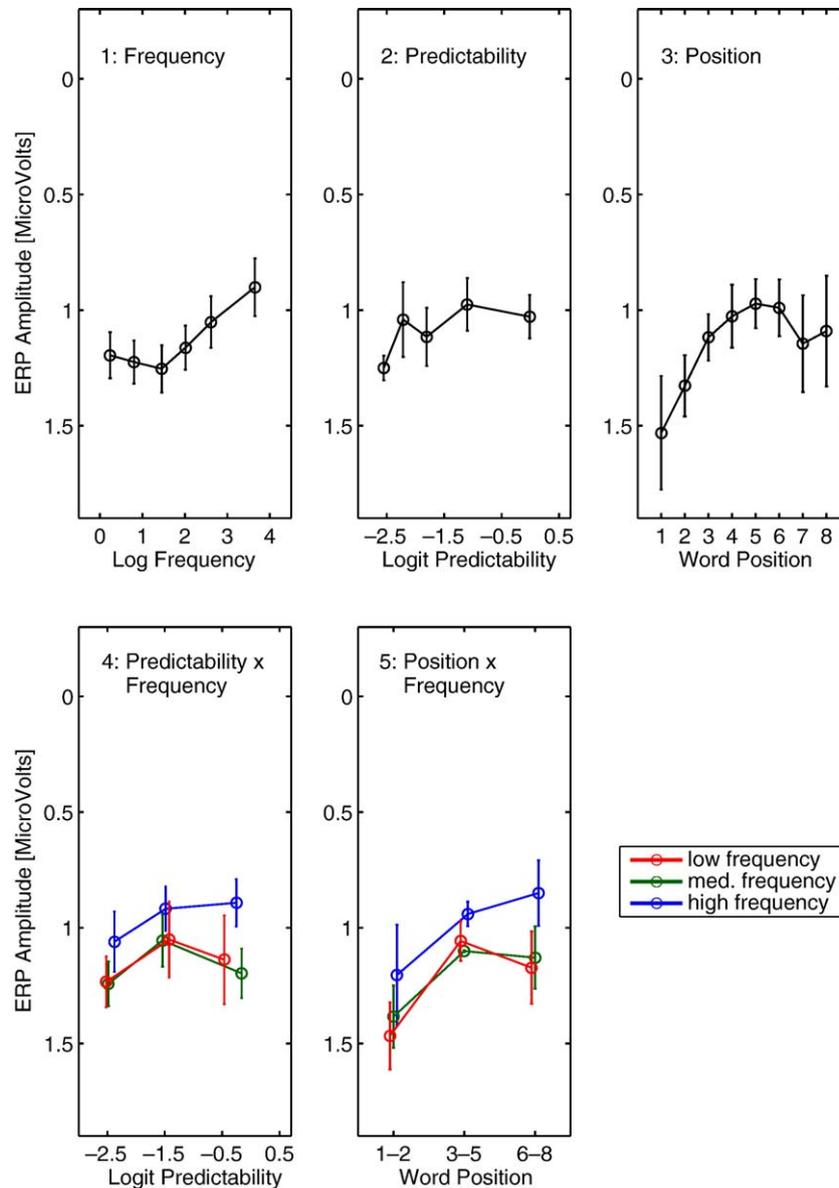
Different from the first rmMRA, linear and the quadratic frequency terms, as well as the interaction of position and frequency, were significant. This indicates that variance related to frequency and word position was absorbed by the interaction of predictability and frequency in the rmMRA with

six predictors. The visualization of the frequency effect on the N400 (Fig. 4, panel 1) reveals a striking contrast to the one on the P200 (Fig. 3, panel 1): Amplitude differences are now largest between the three low-frequency bins. Modulations among the bins of high-frequency words are much smaller. Word position was not significant in the second model.

Finally, in order to examine whether the effect of position was superseded by predictability, the latter was also excluded. We carried out an rmMRA on the four remaining predictors of linear frequency, quadratic frequency, position, and the interaction of position and frequency. In this model, the coefficient for position was significant ( $t = 2.99$ ,  $P < 0.01$ ), indicating that N400 amplitudes decreased with increasing word position (panel 3). The result provides evidence that predictability had absorbed variance of word position. Concerning significance, the other predictors did not change when compared to the rmMRA on five predictors. All coefficients revealed significant results (Frequency:  $t = 5.26$ ,  $P < 0.01$ ; Frequency<sup>2</sup>:  $t = -3.35$ ,  $P < 0.01$ ; Position × Frequency:  $t = -2.07$ ,  $P = 0.02$ ).

### 2.3. Supplementary analyses

For a further validation of the above results, we carried out additional analyses. First, the predictor of word length was added to the rmMRA models. In previous studies, length was found to affect ERP amplitudes particularly around the P200 time window (e.g., Hauk and Pulvermüller, 2004; Van Petten and Kutas, 1990). Furthermore, frequency and length are not independent of each other but are negatively correlated ( $r = -0.56$ ). Thus, we tested whether the pattern of results would change by including both predictors at the same time. When added to the primary rmMRA models, word length was neither reliable on the P200 ( $t = -0.46$ ,  $P = 0.32$ ) nor on the N400 ( $t = -0.29$ ,  $P = 0.39$ ). The basic patterns of significance concerning the other predictors did not change. Additionally, we included the interaction between word length and frequency. This predictor also failed to reach significance for P200 ( $t = 1.13$ ,  $P = 0.13$ ) and N400 ( $t = 0.28$ ,  $P = 0.39$ ) amplitudes. However, on the P200, it absorbed variance accounted for by the quadratic term of frequency, which was no longer significant ( $t = -0.63$ ,  $P = 0.26$ ). The other predictors on the N400 did not change with respect to significance.



**Fig. 3 – mmMRA on P200 Amplitudes.** Illustrations of the predictor effects of the mmMRA in the interval from 140 to 200 ms over fronto-central electrodes. Bins of frequency and predictability in panels 1 and 2 are based on quantiles of the predictors. Categories of frequency and predictability in panels 4 and 5 reflect predictor classes of Table 3. Open symbols show empirical mean amplitude values. Error bars represent 99% within-subject confidence intervals.

#### 2.4. Goodness of fit

The total variance accounted for by each of the mmMRA models described above was small. For example, from the 11.5% in the model for the P200, 11.0% can be attributed to between-subjects variance, whereas the predictors explained 0.5%. The model for the N400 accounted for a total of 7.3% of variance; 6.5% were due to differences between subjects and 0.9% could be traced to the influence of the predictors. At first glance, this seems to be a very poor fit in all cases. Remember, however, that we predicted single-trial EEG amplitudes. As mentioned earlier, this results in substantial loss of noise reduction in the data. The power of variance reduction due to data aggregation can be seen in the values of partial  $\eta^2$  in the rmANOVAs. This measure of effect-size has roughly the same dimension as in

analyses on averaged data from studies using experimental designs. In contrast to that, a small  $R^2$  in analyses on un-averaged data is the rule rather than the exception. Consequently, the amount of variance accounted for should not be unconditionally considered as an adequate measure for the evaluation of model fit, at least not for analyses on unaggregated data.

### 3. Discussion

The present ERP study addressed four issues. The first issue related to the timeline of word recognition during reading. The first appearance of a word frequency effect was considered as an upper limit for lexical access. The second issue addressed

**Table 2 – Mean and standard errors (SE) of regression coefficients of the rmMRA for ERP amplitudes of open-class words in the time window 300–500 ms at centro-occipital electrode sites**

	Mean	SE	t	p <sub>t</sub>	–ΔR <sup>2</sup>	p <sub>–ΔR<sup>2</sup></sub>
<b>1: N400 (6 predictors)</b>						
Constant	–0.283	0.122	–2.31	0.01		
Frequency	0.093	0.131	0.70	0.24	<0.0001	0.39
Frequency <sup>2</sup>	–0.022	0.018	–1.26	0.10	0.0001	0.20
Predictability	0.396	0.068	5.84	<0.01	0.0019	<0.01
Position	0.011	0.027	0.40	0.34	<0.0001	0.76
Predictability × Frequency	–0.090	0.028	–3.22	<0.01	0.0005	<0.01
Position × Frequency	–0.010	0.012	–0.77	0.22	<0.0001	0.37
[R <sup>2</sup> <sub>Predictors</sub> = 0.009; R <sup>2</sup> <sub>Subjects</sub> = 0.065; R <sup>2</sup> <sub>Model</sub> = 0.073]						
<b>2: N400 (5 predictors)</b>						
Constant	–0.283	0.122	–2.31	0.01		
Frequency	0.439	0.097	4.54	<0.01	0.0015	<0.01
Frequency <sup>2</sup>	–0.057	0.017	–3.44	<0.01	0.0006	<0.01
Predictability	0.203	0.036	5.56	<0.01	0.0026	<0.01
Position	0.042	0.028	1.47	0.07	0.0001	0.10
Position × Frequency	–0.022	0.012	–1.84	0.03	0.0002	0.03
[R <sup>2</sup> <sub>Predictors</sub> = 0.008; R <sup>2</sup> <sub>Subjects</sub> = 0.065; R <sup>2</sup> <sub>Model</sub> = 0.073]						

Note. Means, SE, t values, and associated P values for predictors. –ΔR<sup>2</sup> is the drop of variance of the full model due to removal of the predictor; p<sub>–ΔR<sup>2</sup></sub> gives P values for the significance of the variance decrement. R<sup>2</sup><sub>Predictors</sub>, R<sup>2</sup><sub>Subjects</sub>, and R<sup>2</sup><sub>Model</sub> show variance accounted for by predictors alone, by subjects alone, and by the full model, respectively. Statistics are based on 48 subjects, i.e., 47 degrees of freedom for t statistics.

the role of context in word recognition. Unlike any previous study, we used predictability norms for each word in the sentences as an independent measure of prior sentence context. Third, with this information, we could also test the interaction of predictability and frequency and study the question of whether they map onto temporally distinct stages of word recognition. Finally, we could assess the contribution of word position, independent of context effects reflected in predictability. In the following two sections, we discuss results on the P200 and on the N400. Thereafter, we attempt to present an integrative account of our results.

### 3.1. P200

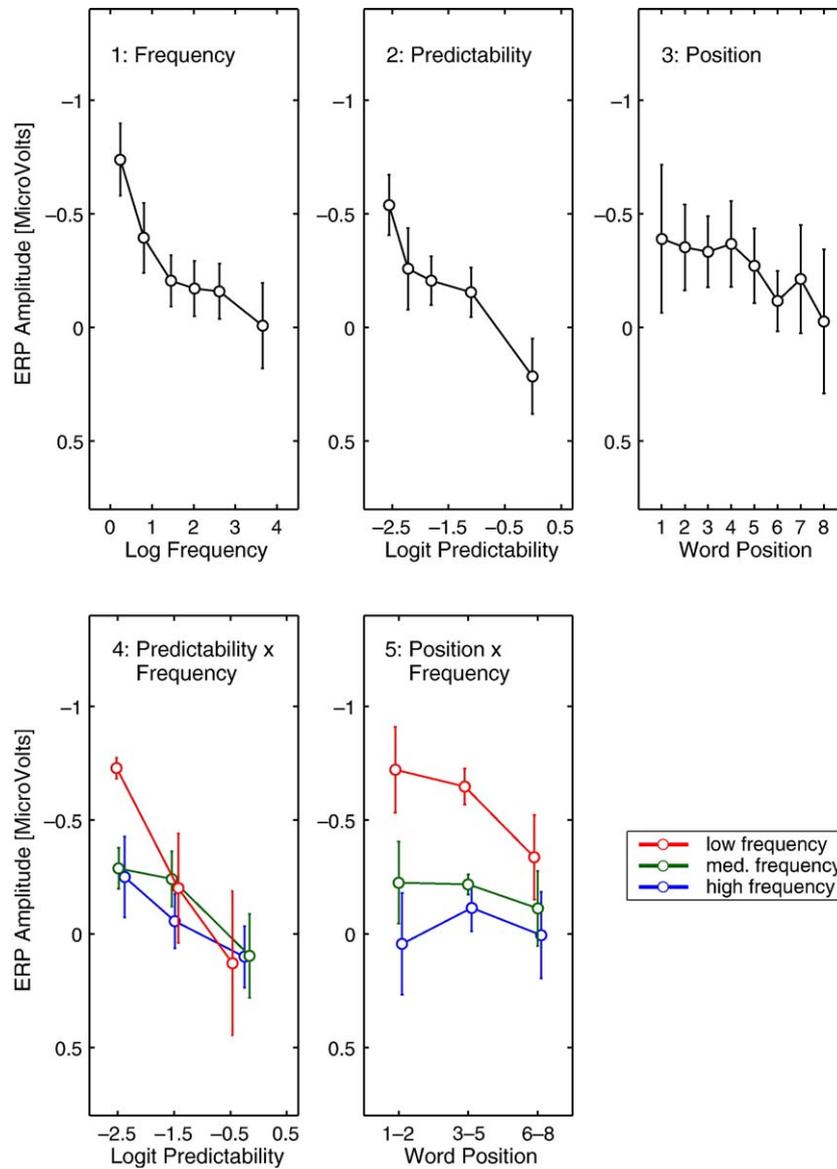
In both rmANOVAs in the latency range from 140 to 200 ms post-stimulus, we found a strong frequency effect over fronto-central electrodes. Amplitudes were smaller for high-frequency than for low-frequency words. With respect to frequency as an index for lexical access, this provides evidence that words

are identified within the first 200 ms after stimulus presentation during sentence reading. This result is in line with previous studies. Early frequency effects were reported by Sereno et al. (1998) at 132 ms, by Sereno et al. (2003) between 132 and 192 ms, by Assadollahi and Pulvermüller (2001) between 120 and 170 ms, and by Hauk and Pulvermüller (2004) between 150 and 190 ms. The rmMRA on single-trial EEG amplitudes confirmed this finding. The quadratic trend of frequency illustrated in Fig. 3 (panel 1) revealed larger amplitude differences among high-frequency than among low-frequency words. Thus, lexical access was presumably completed for high-frequency words while low-frequency words were still being processed. Results from behavioral and eye movement studies corroborate this hypothesis revealing longer reaction times (e.g., Forster and Chambers, 1973; Rubenstein et al., 1970) and fixation durations on low-frequency words (e.g., Inhoff and Rayner, 1986; Kliegl et al., 2004, 2006; Rayner and Duffy, 1986; Schilling et al., 1998). In supplementary analyses, we tested whether the result was caused by words of different lengths rather than by frequency. This was necessary because frequency and length are negatively correlated, i.e., on average, high-frequency words are shorter than low-frequency words. Previous studies also revealed effects of word length on ERP amplitudes in early time windows (e.g., Hauk and Pulvermüller, 2004; Van Petten and Kutas, 1990). However, word length did not affect P200 amplitudes. Also the interaction between length and frequency was not reliable, but it should be noted that, as a consequence of the additional predictor, word frequency lost significance in the P200 time window. This can be attributed to the fact that both variables account for variance of the very same effect: The interaction plot of length and frequency (not illustrated in this paper) revealed that especially short words (i.e., high-frequency words) show a frequency effect in the P200 time window. The quadratic frequency effect demonstrated

**Table 3 – Number of words, mean values, and standard deviations (SD) in three categories of logarithmic frequency and logit-transformed predictability for open-class words in sentences containing seven to nine words in the Potsdam Sentence Corpus**

Class	Frequency			Predictability		
	Number of words	Mean	SD	Number of words	Mean	SD
1	153	0.47	0.29	278	–2.49	0.13
2	218	1.82	0.42	108	–1.50	0.27
3	126	3.36	0.48	111	–0.24	0.58

Sentence-final words are excluded.



**Fig. 4 – mmMRA on N400 Amplitudes. Illustrations of the predictor effects of the mmMRA in the interval from 300 to 500 ms over centro-occipital electrodes. Bins of frequency and predictability in panels 1 and 2 are based on quantiles of the predictors. Categories of frequency and predictability in panels 4 and 5 reflect predictor classes of Table 3. Open symbols show empirical mean amplitude values. Error bars represent 99% within-subject confidence intervals.**

that predominantly high-frequency words (i.e., short words) are lexically accessed. Although the correlation between the variables complicates an ascription of the amplitude modulations to either length or frequency, we attribute the effect on the P200 primarily to the contribution of frequency rather than of length, because the latter predictor did not significantly account for unique variance in the mmMRAs.

Word predictability revealed a significant effect in the rmANOVA on P200 amplitudes, suggesting an early influence of context information on word recognition. However, it is important to note that word position strongly modulating the P200 was not included as factor in this analysis. Since predictability and position are highly correlated ( $r = 0.41$ ), it is conceivable that the effect was related to position rather than to predictability. This possibility was examined in the mmMRA

where effects of predictability and position were estimated within one model. Neither word predictability nor the interaction of predictability and frequency affected P200 amplitudes in the mmMRA. The variance was absorbed by word position better accounting for this effect. Thus, on the basis of the present results, we cannot conclude that predictability influenced word recognition at this latency. This is at odds with the results of Sereno et al. (2003) reporting that context affected lexical access of ambiguous words and marginally facilitated processing of low-frequency words. The conflicting results can be attributed to differences between the studies. Sereno et al. experimentally manipulated the context, in which selected ambiguous words appeared. In the present study, neither predictability nor sentences realized extreme conditions for context effects and, consequently, were not significant.

Surprisingly, the strongest influence on the P200 was provided by word position. The rmMRA made clear that amplitude modulations were not linear but a quadratic function of position. Words occurring early or late in sentences elicited larger amplitudes than words in middle positions. This effect was independent of word length. One might wonder whether the frequency effect on the P200 was an artifact of the influence of position. However, this is very unlikely because both position and frequency were included as predictors in the regression models at the same time. If the frequency effect was an artifact of word position, the latter would have absorbed the variance accounted for by frequency, which was not the case. Furthermore, the correlation between position and frequency is small ( $r = 0.12$ ). A systematic effect of position would have caused a rather unsystematic effect of frequency.

Reasons for the decreasing P200 towards the center of a sentence and for subsequent increasing amplitudes remain unclear on the basis of the present data. There were no a priori theoretical considerations predicting a quadratic word position effect. We included the quadratic term in the rmMRA only after visual inspection of the data, so suggestions for a solution are speculative. One possibility is that increasing working memory load in the middle of a sentence caused a negative shift, tantamount to decreasing P200 amplitudes. At the beginning of a sentence, only very few words must be kept in mind; towards the end of a sentence, high predictability facilitates recognition and semantic integration of new words and contextual information eases remembering the content of the sentence. Compared to this, the effort of recognizing and integrating upcoming words while keeping the previous sentence fragment in mind might be largest in the middle of a sentence. Another possibility is that different parts of a sentence vary in importance of semantic content. In the German language, it is very likely that the words carrying the most important meaning for a fast and correct understanding occur in the middle of a sentence (e.g., a verb). Expectancy or “alertness” could have caused a long-term negative variation whenever a sentence proceeded towards its major contents. Finally, it is also possible that the position effect was specific for the stimulus material of the present study. In any case, further investigation is necessary to clarify the nature of the word position effect.

### 3.2. N400

Both the rmANOVA and the rmMRA showed a strong effect of predictability on the N400. This is in line with findings of previous experiments (e.g., Kutas and Hillyard, 1984; Kutas and Van Petten, 1994). N400 amplitudes are inversely correlated with predictability. Obviously, this measure is an appropriate predictor for modulations of N400 amplitudes. Considering that none of the sentences contained any semantic violation and that no artificially strong variation of predictability was intended during the construction of the stimulus material, this result corroborates once more the robustness of the N400 effect.

In the rmANOVA with the factors of frequency and position, we found a strong main effect of frequency; N400 amplitudes decreased with augmenting frequency, which corresponds to previous reports (Rugg, 1990; Van Petten and Kutas, 1990). The

size of this effect was attenuated in the rmANOVA with the factors frequency and predictability indicating that either predictability or the interaction term absorbed variance of frequency. The rmMRAs supported this hypothesis: Linear and quadratic frequency terms were strongly reliable only when the interaction of predictability and frequency was excluded from the model. Obviously, the interaction term was enough to explain frequency-related variance.

The interactions of predictability and frequency as well as of position and frequency were significant in the rmANOVAs, pointing to an interplay of frequency and context information on the N400. Given the argument that predictability and position capture similar concepts, the two interactions may account for the same effect: The frequency effect degraded as context information increased. The results of the rmMRA confirmed this view showing a strong interaction of predictability and frequency while the interaction of position and frequency was not significant. Although the interaction plot clearly reveals that the frequency effect was decreasing with increasing word position (Fig. 4, panel 5), this pattern could be completely due to the interaction of predictability and frequency (Fig. 4, panel 4). Thus, the joint effect of predictability and frequency is sufficient to account for the decrease of the frequency effect across words; there may be no independent contribution of word position. The finding is in line with our hypothesis that predictability as a more direct measure of context information accounts better for N400 effects than word position.

Further support is provided regarding the main effect of position. Amplitudes were smaller for words occurring late in a sentence as reported in previous studies (Van Petten and Kutas, 1990, 1991; Van Petten, 1993). While the rmANOVA showed a statistical trend, the rmMRAs made clear that the effect of word position was absorbed by predictability. The position effect was significant only when predictability as well as the interaction of predictability and frequency were removed from the rmMRA model.

The results can be compared directly with Van Petten and Kutas' (1990, 1991) reports of a word position effect and a significant interaction of position and frequency. Except for the final words, they used word position as a metric for the strength of contextual information. They proposed that the decline of N400 amplitudes and the decrease of the frequency effect across the sentence reflect the influence of contextual constraint rather than word position. Given that predictability better accounted for the N400 effects than position and therefore absorbed the variance of position and frequency, our data strengthen Van Petten and Kutas' (1990) view that position “can serve as metric of the semantic and structural links that differentiate a sentence from a string of unconnected words” (p. 388).

### 3.3. Frequency and predictability: lexical and post-lexical processes?

The decreasing N400 amplitude with increasing predictability demonstrates that context facilitates word processing and language comprehension, independent of the position of the word in the sentence. Additionally, we showed that the interaction of predictability and frequency absorbed the

variance accounted for by the interaction of word position and frequency. Thus, word position in a sentence reflects primarily the build-up of contextual constraint (Van Petten and Kutas, 1990, 1991, see also Van Petten, 1993). Do the results also confirm the proposal that context supersedes the role of word frequency concerning lexical access while we read through a sentence? It was concluded that “word frequency plays a role in these processes only when meaningful semantic context is weak, as at the beginning of a congruent sentence [...]” (Van Petten, 1993, p. 498). This interpretation implies a unidirectional influence of contextual constraint on the impact of word frequency in a sense that context can affect the relevance of frequency but not the other way round. On the basis of the present results, we propose that word frequency and context interact in a bidirectional way.

Concerning the N400 amplitudes, there are two crucial illustrations in Fig. 4. First, panel 5 reveals that the frequency effect decreases with increasing word position. In principle, this replicates Van Petten and Kutas’ (1990, 1991) results. One might conclude that, on the N400, frequency does not play a role as context increases. Remember, however, that the term was only significant when the interaction between predictability and frequency was left out of the rmMRA.

The second relevant illustration relates to the interaction of predictability and frequency (panel 4). This plot allows an alternative interpretation: the effect of contextual information (indicated by predictability) is larger for low-frequency than for high-frequency words. In other words, frequency modulates the strength of the predictability effect on the N400.

This conclusion is also in line with the frequency effects in the rmMRAs. Our results suggest that high-frequency words are lexically accessed before 200 ms indicated by the quadratic trend of frequency in the P200 epoch (Fig. 3, panel 1). Predictability did not influence this fast process. As was shown in the analysis using a reduced rmMRA model, frequency affected the N400 amplitude following a quadratic trend: amplitudes of low-frequency words differed from high-frequency words, whereas differences were smaller among the latter. This indicates that, at this later time, especially low-frequency words were accessed. The variance accounted for by frequency on the N400 was absorbed by the interaction of predictability and frequency. Thus, both lexical access of low-frequency words and the effect of predictability affected ERPs at the same latency. The interaction suggests that both variables act on the same stage of word recognition. Lexical access of low-frequency words benefits from contextual information and this benefit is strongly reduced in the case of high-frequency words having been recognized earlier (see also Becker, 1979; Sereno et al., 2003).

Interactive models of word recognition (e.g., Grainger and Jacobs, 1996; McClelland, 1987; Morton, 1969) can explain the present results, because they allow feedback from higher to lower levels of processing. However, the findings present a problem for modular approaches (e.g., Fodor, 1983; Forster, 1979) assuming distinct and sequential lexical and post-lexical stages, at least using word frequency and predictability as primary indicators. Alternatively, one would need to establish post-lexical sources in word-frequency and lexical sources in word-predictability norms. After all, there is a substantial correlation ( $r = 0.41$ ) between them.

In sum, word recognition seems to be a gradual process rather than a strict sequence of distinct stages (see also Coulson and Federmeier, *in press*). Also, Van Petten (1995) pointed out that “although word frequency is a lexical variable, the human language-processing system does not always respect the boundary between lexical and sentential processing” (p. 520). The brain seems to use all sources of information as soon as they become available in order to provide a fast and correct understanding.

### 3.4. Conclusions

The purpose of this study was to investigate joint effects of frequency and predictability on early and late ERP components, taking into account also effects of word position. In the present experiment, we reconciled several isolated findings of previous studies and contributed a few novel results: High-frequency words triggered a differential ERP response in the first 200 ms after stimulus onset; there was no evidence for an effect of predictability on this early P200 component. In contrast, predictability correlated strongly and linearly with the N400 amplitude. In addition, the N400 amplitude exhibited a larger predictability effect for low-frequency than for high-frequency words, compatible with a late-access interpretation of low-frequency words. Finally, P200 amplitudes decreased across sentence-initial words and increased towards the end of a sentence. Apparently, this effect does not relate to the recognition of the currently presented word, at least not exclusively. In general, the results suggest different time constraints but also overlapping processes for frequency-related lexical access and predictability-related post-lexical integration during reading.

## 4. Experimental procedures

### 4.1. Participants

Fifty students (19 to 35 years; 19 males) of the Catholic University of Eichstätt-Ingolstadt were paid for their participation. All were native German speakers and had normal or corrected-to-normal vision. Forty-three subjects were right-handed.

### 4.2. Stimuli

The Potsdam Sentence Corpus (PSC) comprises 144 German sentences (1138 words) with a large variety of grammatical structures. The mean sentence length is 7.9 words with a range from 5 to 11 words. Words were divided into three categories with respect to the variable frequency and predictability. These categories were used for repeated measures analyses of variance (rmANOVAs) and for the visualization of effects; repeated measures multiple regression analyses (rmMRAs) were based on the continuous values of these predictors.

Word frequencies of the corpus words are based on DWDS norms (Das Digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts), which are computed on a total of 100 million words (Geyken, *in press*; Geyken et al., *in preparation*). Each of three logarithmic frequency classes contains at least 254 words [class 1 (log frequency: 0 to 1): 254 words, mean: 0.46, SD: 0.29;

class 2 (log frequency: 1 to 2.5): 406 words, mean: 1.82, SD: 0.42; class 3 (log frequency: 2.5 to max.): 478 words, mean: 3.57, SD: 0.55].

Predictability of words was collected in an independent norming study from 282 native speakers of German ranging in age from 17 to 80 years. Participants guessed the first word of the unknown sentence and entered it via the keyboard. In return, the computer presented the first word of the original sentence. Thereafter, subjects entered their guess for the second word followed by presentation of the second word of the original sentence. This procedure continued until a period indicated the end of a sentence. Correct words stayed on the screen. The order of sentences was randomized. Twenty subjects generated predictions for all of the 144 sentences. The other participants worked through a quarter of the corpus. Collapsing the complete and partial protocols across participants yielded 83 complete protocols. The obtained predictability values were logit-transformed [ $\text{logit} = 0.5 * \ln(\text{pred} / (1 - \text{pred}))$ ]. Predictabilities of zero were replaced with  $1/(2*83)$  and those of perfectly predicted words with  $(2*83 - 1)/(2*83)$ , where 83 represents the number of complete predictability protocols (Cohen and Cohen, 1975). That means that for a word with predictability 0.5 the odds of guessing are  $0.5/0.5 = 1$ , and consequently the log odds of guessing are  $\ln(1) = 0$ . Thus, words with predictabilities larger than 0.5 yield positive logits and predictabilities smaller than 0.5 negative logits. The logit transformation corrects for the dependency of mean probabilities ( $P$ ) and associated standard deviations ( $SD$ ) [i.e.,  $SD = P(1 - P)$ ] by stretching the tail of the distribution (see also Kliegl et al., 2004). The corpus contains at least 254 words in each of three logit-based predictability classes [class 1 (–2.553 to –2.0): 464 words, mean: –2.47 SD: 0.14; class 2 (–2.0 to –1.0): 254 words, mean: –1.46, SD: 0.29; class 3 (–1.0 to 2.553): 420 words, mean: –0.04, SD: 0.77].

#### 4.3. Procedure

Subjects were seated at a distance of 60 cm from the monitor and were instructed to read the sentences for comprehension. After ten practice trials, the 144 sentences were presented word by word (Font: Courier New, Size: 12) in randomized order. The first word of each sentence was preceded by a fixation cross presented for 500 ms in the middle of the monitor and followed by a blank screen for another 500 ms. Stimuli together with the adjacent punctuation were displayed for 250 ms in black on a white screen in the center of the monitor. The stimulus onset asynchrony (SOA) was 700 ms. A multiple-choice question was presented after 27% of the sentences; subjects pressed one of three buttons to indicate their answer. After the remaining sentences, an array of asterisks appeared for 2000 ms (preceded and followed by a 1000 ms blank screen) in the center of the screen. During the presentation of a question or asterisks, subjects were allowed to blink. They took a break of 10 min after the first half of the experiment. Sessions lasted about 1.5 h.

#### 4.4. Electrophysiological recording

An electrode cap (ElectroCap International) was used to record EEG data on 26 scalp locations (FP1, FP2, AFZ, FZ, F3, F4, F7, F8, FC3, FC4, FC5, FC6, CZ, C3, C4, T7, T8, CP5, CP6, PZ, P3, P4, P7, P8,

O1, O2) corresponding to the revised 10/20 International System. All scalp electrodes and one electrode on the right mastoid were originally referenced to one electrode on the left mastoid. Data were converted offline to average reference. In addition, two horizontal (situated at the outer left and outer right canthus) and two vertical EOG electrodes (above and below the right eye) recorded bipolarly eye movements and blinks. Impedances of scalp electrodes were kept below 5 k $\Omega$ . Data were recorded continuously with a sampling rate of 256 Hz. The recordings were high- and low-pass filtered by amplifier adjustment of 0.1 and 100 Hz, respectively.

#### 4.5. Analyses

EEG data contaminated by artifacts were rejected offline via an automatic algorithm and visual inspection. Data of two subjects had to be completely removed, one because of loss of data due to technical problems and one because of a former neurological disease. From the remaining 48 subjects, a total of 11.43% of trials was eliminated. The continuous EEG recording was divided into 800 ms epochs beginning 100 ms before stimulus onset. Data were analyzed relative to a baseline of 100 ms preceding each stimulus.

In order to reduce effects due to the large variability of sentence lengths those with less than 7 and more than 9 words were excluded. Only open-class words were included in the analyses; closed-class words were eliminated. Additionally, sentence-final words were removed from the data set. This left us with a total of 105 sentences comprising 497 open-class words for statistical analyses. Number of words, mean values, and standard deviations of three categories of frequency and predictability are listed in Table 3.

Correlations between frequency and predictability ( $r = 0.41$ ), predictability and word position ( $r = 0.41$ ), and frequency and position ( $r = 0.12$ ) were significant ( $P \leq 0.01$ ). Descriptive statistics for the distribution of words across the positions in sentences are presented in Table 4.

Two time windows were chosen for analyses. The selection of the first window was based on the hypothesis that an early frequency effect would occur well within the first 200 ms after the presentation of a stimulus. Visual inspection of the data in an epoch between 100 and 200 ms post-stimulus revealed differences between frequency classes on a fronto-central

**Table 4 – Number of words, mean values, and standard deviations (SD) of logarithmic frequency and logit-transformed predictability for open-class words on eight word positions in the Potsdam Sentence Corpus**

Word position	Number of words	Frequency		Predictability	
		Mean	SD	Mean	SD
1	38	1.58	0.90	–2.42	0.30
2	90	1.54	0.98	–2.29	0.59
3	90	1.76	1.18	–1.95	0.79
4	72	1.97	1.28	–1.74	0.89
5	72	1.73	1.11	–1.54	1.08
6	76	2.08	1.32	–1.24	1.08
7	37	1.76	1.12	–1.41	1.07
8	22	2.02	0.94	–1.06	1.02

positivity peaking at 170 ms<sup>1</sup>. This component was identified as P200. We defined the first time window in an interval between 140 to 200 ms (peak amplitude at 170 ms ± 30 ms) on fronto-central electrodes (AFZ, FZ, F3, F4, FC3, FC4, FC5, FC6, CZ, C3, C4). The second epoch ranged from 300 to 500 ms over centro-occipital electrodes (CZ, C3, C4, CP5, CP6, PZ, P3, P4, P7, P8, O1, O2), a time window often used in N400 research.

Effects of frequency, predictability, and position on P200 and N400 amplitudes were analyzed in rmANOVAs. According to the classification in Table 3, words were divided into three categories of frequency and three categories of predictability. Also, three classes of word position were generated (positions 1–2:  $N = 128$ ; positions 3–5:  $N = 234$ ; positions 6–8:  $N = 135$ ). On the basis of these categories, ERP single-averages were computed for each subject and were submitted to two separate  $3 \times 3$  rmANOVAs for each of the components: one with frequency (1, 2, 3) and predictability (1, 2, 3), and one with frequency (1, 2, 3) and position (1–2, 3–5, 6–8) as within-subject factors. The high correlation between predictability and position did not permit an ANOVA including both factors at the same time (i.e., the lack of high predictable words on early positions would have caused empty cells). Where appropriate, the Huynh–Feldt correction for the violation of sphericity (Huynh and Feldt, 1976) was used to adjust degrees of freedom.

The results of the rmANOVAs for each of the components were scrutinized in separate rmMRAs (Lorch and Myers, 1990, method 3). In these analyses, the influence of frequency, predictability, and position together with their interactions could be tested within one model. Mean single-trial EEG amplitudes, collapsed across the selected electrodes of the P200 and the N400 as well as across the sampling points corresponding to each of the time intervals of the two components, were computed for each open-class word within each subject. This resulted in a total of 21,176 amplitude values per epoch serving as dependent variables. The rmMRAs were used to examine P200 and N400 effects of the following six predictors: Frequency, frequency × frequency, predictability, position, predictability × frequency, and position × frequency. Additionally, position × position was included in the rmMRA on P200 amplitudes since visual inspection suggested a quadratic trend. All analyses are based on continuous predictor values instead of the categories utilized in the rmANOVAs. The redundancy of the predictors was checked by removing one predictor at a time and by computing the decrease of explained variance for the reduced model.

#### 4.6. Plots

Figs. 1 and 2 present grand average plots for three frequency and predictability classes corresponding to the categories defined in Table 3, respectively. The grand averages were computed for open-class words collapsed across all word positions except for final ones in sentences comprising 7 to 9 words.

The effects of the predictors on ERP amplitudes in rmMRAs for open-class words are visualized in Figs. 3 and 4. For

purposes of illustration, the predictor frequency (panels 1) was divided into six quantiles, each containing roughly the same number of data points ( $N \geq 3506$ ). The same procedure was applied to predictability (panels 2): six quantiles were computed, but since the proportion of words not predictable at all was very high, the lowest quantile contained more data ( $N = 9405$ ) than the rest of the quantiles ( $N \geq 1859$ ). Consequently, the second quantile did not capture any data at all. Thus, only five points are plotted for predictability. No quantiles were computed for word position (panels 3), since each word was uniquely attributable to one value. For the purpose of noise reduction, the number of bins was reduced in the interaction plots (panels 4 and 5): frequency and predictability were categorized according to three classes defined in Table 3; word position (panels 5) was divided into three classes (positions 1–2, 3–5, and 6–8).

Open symbols in each panel of Figs. 3 and 4 reflect the mean of empirical single-trial EEG amplitudes collapsed across the selected electrodes and across sampling points for each time window, across words of corresponding categories, and across subjects. Error bars reflect the 99% within-subject confidence intervals (Loftus and Masson, 1994).

#### Acknowledgments

We thank Mario Braun for research assistance as well as Olaf Dimigen and two anonymous reviewers for their helpful and constructive comments. This research was supported by grant KL 955/6 from Deutsche Forschungsgemeinschaft.

#### REFERENCES

- Assadollahi, R., Pulvermüller, F., 2001. Neuromagnetic evidence for early access to cognitive representations. *NeuroReport* 12, 207–213.
- Becker, C.A., 1979. Semantic context and word frequency effects in visual word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 5, 252–259.
- Besson, M., Fischler, I., Boaz, T., Raney, G., 1992. Effects of automatic associative activation on explicit and implicit memory tests. *J. Exper. Psychol., Learn., Mem., Cogn.* 18, 89–105.
- Brown, C., Hagoort, P., 1993. The processing nature of the N400: evidence from masked priming. *J. Cogn. Neurosci.* 5, 34–44.
- Cohen, J., Cohen, P., 1975. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.
- Coulson, S., Federmeier, K.D., in press. Words in context: ERPs and the lexical/postlexical distinction. *J. Psycholing. Res.*
- Deacon, D., Hewitt, S., Yang, C.-M., Nagata, M., 2000. Event-related potential indices of semantic priming using masked and unmasked words: evidence that the N400 does not reflect a post-lexical process. *Cogn. Brain Res.* 9, 137–146.
- Fischler, I., Bloom, P.A., 1979. Automatic and attentional processes in the effects of sentence contexts on word recognition. *J. Verbal Learn. Verbal Behav.* 18, 1–20.
- Fodor, J.A., 1983. *The Modularity of Mind*. MIT Press, Cambridge, MA.
- Forster, K.I., 1979. Levels of processing and the structure of the language processor. In: Cooper, W., Walker, E. (Eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*. Erlbaum, Hillsdale, NJ, pp. 25–87.

<sup>1</sup> We also tested for amplitude differences in the epoch from 100 to 140 ms post-stimulus; however, effects in this time window appeared to be unstable.

- Forster, K.I., Chambers, S.M., 1973. Lexical access and naming time. *J. Verbal Learn. Verbal Behav.* 12, 627–635.
- Friedman, D., Simson, R., Ritter, W., Rapin, I., 1975. The late positive component (P300) and information processing in sentences. *Electroencephalogr. Clin. Neurophysiol.* 38, 255–262.
- Geyken, A., in press. The DWDS-corpus: A reference corpus for the German language of the 20th century. In: Fellbaum, C. (Ed.), *Collocations and Idioms: Linguistic, Lexicographic, and Computational Aspects*. Continuum Press, London.
- Geyken, A., Hanneforth, T., Kliegl, R., in preparation. Corpus matters: A comparison of DWDS and CELEX lexical and sublexical frequency norms for the prediction of fixation durations during reading.
- Grainger, J., Jacobs, A.M., 1996. Orthographic processing in visual word recognition: a multiple read-out model. *Psychol. Rev.* 103, 518–565.
- Hagoort, P., 2003. Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *J. Cogn. Neurosci.* 15, 838–899.
- Hauk, O., Pulvermüller, F., 2004. Effects of word length and frequency on the human event-related potential. *Clin. Neurophysiol.* 115, 1090–1103.
- Holcomb, P.J., 1993. Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology* 30, 47–61.
- Hudson, P.T., Bergman, M.W., 1985. Lexical knowledge in word recognition: word length and word frequency in naming and lexical decision tasks. *J. Mem. Lang.* 24, 46–58. <http://www.dwds.de>.
- Huynh, H., Feldt, L.S., 1976. Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs. *J. Educ. Stat.* 1, 69–82.
- Inhoff, A.W., Rayner, K., 1986. Parafoveal word processing during eye fixations in reading: effects of word frequency. *Percept. Psychophys.* 40, 431–439.
- Jacobs, A.M., Grainger, J., 1994. Models of visual word recognition—Sampling the state of the art. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 1311–1334.
- King, J.W., Kutas, M., 1998. Neural plasticity in the dynamics of human visual word recognition. *Neurosci. Lett.* 244, 61–64.
- Kleiman, G.M., 1980. Sentence frame contexts and lexical decisions: sentence-acceptability and word-relatedness effects. *Mem. Cogn.* 8, 336–344.
- Kliegl, R., Grabner, E., Rolfs, M., Engbert, R., 2004. Length, frequency, and predictability effects of words on eye movements in reading. *Eur. J. Cogn. Psychol.* 16, 262–284.
- Kliegl, R., Nuthmann, A., Engbert, R., 2006. Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *J. Exp. Psychol. Gen.* 135, 12–35.
- Kutas, M., Federmeier, K.D., 2000. Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470.
- Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205.
- Kutas, M., Hillyard, S.A., 1983. Event-related brain potentials to grammatical errors and semantic anomalies. *Mem. Cogn.* 11, 539–550.
- Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163.
- Kutas, M., Van Petten, C., 1994. Psycholinguistics electrified: event-related brain potential investigations. In: Gernsbacher, M.A. (Ed.), *Handbook of Psycholinguistics*. Academic Press, San Diego, pp. 83–143.
- Kutas, M., Van Petten, C., Besson, M., 1988. Event-related potential asymmetries during the reading of sentences. *Electroencephalogr. Clin. Neurophysiol.* 69, 218–233.
- Loftus, G.R., Masson, M.E.J., 1994. Using confidence intervals in within-subject designs. *Psychon. Bull. Rev.* 1, 476–490.
- Lorch, R.F., Myers, J.L., 1990. Regression analyses of repeated measures data in cognitive research. *J. Exper. Psychol., Learn., Mem., Cogn.* 16, 149–157.
- McClelland, J.L., 1987. The case of interactionism in language processing. In: Coltheart, M. (Ed.), *Attention and Performance XII: The Psychology of Reading*. Erlbaum, Hillsdale, NJ, pp. 3–36.
- Monsell, S., Doyle, M.C., Haggard, P.N., 1989. Effects of frequency on visual word recognition tasks: where are they? *J. Exp. Psychol. Gen.* 118, 43–71.
- Morton, J., 1969. Interaction of information in word recognition. *Psychol. Rev.* 76, 165–178.
- Münte, T.F., Wieringa, B.M., Weyerts, H., Szentkuti, A., Matzke, M., Johannes, S., 2001. Differences in brain potentials to open and closed class words: class and frequency effects. *Neuropsychologia* 39, 91–102.
- Neville, H.J., Mills, D.L., Lawson, D.S., 1992. Fractionating language: different neural subsystems with different sensitive periods. *Cereb. Cortex* 2, 244–258.
- Osterhout, L., 1997. On the brain response to syntactic anomalies: manipulations of word position and word class reveal individual differences. *Brain Lang.* 59, 494–522.
- Osterhout, L., Holcomb, P.J., 1995. Event-related potentials and language comprehension. In: Rugg, M.D., Coles, M.G.H. (Eds.), *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition*. Oxford Univ. Press, Oxford, pp. 171–215.
- Osterhout, L., Nicol, J., 1999. On the distinctiveness, independence, and time course of the brain response to syntactic and semantic anomalies. *Lang. Cogn. Processes* 14, 283–317.
- Osterhout, L., Bersick, M., McKinnon, R., 1997. Brain potentials elicited by words: word length and frequency of an early negativity. *Biol. Psychol.* 46, 143–168.
- Rayner, K., Duffy, S.A., 1986. Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity. *Mem. Cogn.* 14, 191–201.
- Rayner, K., Well, A.D., 1996. Effects of contextual constraint on eye movements in reading: a further examination. *Psychon. Bull. Rev.* 3, 504–509.
- Rayner, K., Binder, K.S., Ashby, J., Pollatsek, A., 2001. Eye movement control in reading: word predictability has little influence on initial landing positions of words. *Vision Res.* 41, 943–954.
- 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. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 720–732.
- Reichle, E.D., Pollatsek, A., Fisher, D.L., Rayner, K., 1998. Toward a model of eye movement control in reading. *Psychol. Rev.* 105, 125–157.
- Rubenstein, H., Garfield, L., Millikan, J.A., 1970. Homographic entries in the internal lexicon. *J. Verbal Learn. Verbal Behav.* 9, 487–494.
- Rugg, M.D., 1990. Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Mem. Cogn.* 18, 367–379.
- Schilling, H.E., Rayner, K., Chumbley, J.I., 1998. Comparing naming, lexical decision, and eye fixation times: word frequency effects and individual differences. *Mem. Cogn.* 26, 1270–1281.
- Schuberth, R.E., Spoehr, K.T., Lane, D.M., 1981. Effects of stimulus and contextual information on lexical decision process. *Mem. Cogn.* 9, 68–77.
- Sereno, S.C., Rayner, K., Posner, M.I., 1998. Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *NeuroReport* 9, 2195–2200.
- Sereno, S.C., Brewer, C.C., O'Donnell, P.J., 2003. Context effects in word recognition: evidence for early interactive processing. *Psychol. Sci.* 14, 328–333.
- Van Petten, C., 1993. A comparison of lexical and sentence-level

- 
- context effects in event-related potentials. *Lang. Cogn. Processes* 8, 485–531.
- Van Petten, C., 1995. Words and sentences: event-related brain potential measures. *Psychophysiology* 32, 511–525.
- Van Petten, C., Kutas, M., 1990. Interactions between sentence context and word frequency in event-related brain potentials. *Mem. Cogn.* 18, 380–393.
- Van Petten, C., Kutas, M., 1991. Influences of semantic and syntactic context on open- and closed-class words. *Mem. Cogn.* 19, 95–112.