HOW STRONGLY DO WORD READING TIMES AND LEXICAL DECISION TIMES CORRELATE?

COMBINING DATA FROM EYE MOVEMENT CORPORA AND MEGASTUDIES

Victor Kuperman [1], Denis Drieghe [2], Emmanuel Keuleers [3], & Marc Brysbaert [3]

[1] McMaster University, Canada

[2] University of Southampton, UK

[3] Ghent University, Belgium

Keywords: visual word recognition, lexical decision, naming, gaze duration, eye movements, megastudy, corpus

Address:
Marc Brysbaert
Department of Experimental Psychology
Ghent University
Henri Dunantlaan 2
B-9000 Gent
Belgium
Tel. +32 9 264 94 25
Fax. +32 9 264 64 96
marc.brysbaert@ugent.be

Abstract

We assess the amount of shared variance between three measures of visual word recognition latencies: eye movement latencies, lexical decision times and naming times. After partialling out the effects of word frequency and word length, two well-documented predictors of word recognition latencies, we see that 7-44% of the variance is uniquely shared between lexical decision times and naming times, depending on the frequency range of the words used. A similar analysis of eye movement latencies shows that the percentage of variance they uniquely share either with lexical decision times or with naming times is much lower. It is 5 – 17% for gaze durations and lexical decision times in studies with target words presented in neutral sentences, but drops to .2% for corpus studies in which eye movements to all words are analysed. Correlations between gaze durations and naming latencies are lower still. These findings suggest that processing times in isolated word processing and continuous text reading are affected by specific task demands and presentation format, and that lexical decision times and naming times are not very informative in predicting eye movement latencies in text reading once the effect of word frequency and word length are taken into account. The difference between controlled experiments and natural reading suggests that reading strategies and stimulus materials may determine the degree to which the immediacy-of-processing assumption and the eye-mind assumption apply. Fixation times are more likely to exclusively reflect the lexical processing of the currently fixated word in controlled studies with unpredictable target words rather than in natural reading of sentences or texts.

One of the ways in which theories of visual word processing aim to explain the cognitive processes involved in reading, is by accounting for differences in the time required to recognize different words. Hence, recognition times for large samples of words in different languages are collected to validate (computational) models of word processing. The question arises as to whether different measures of word recognition time are equally valid estimates when testing these theories.

At first sight, the most ecologically valid estimate of the time needed to recognize a printed word is the time the eyes remain fixated on that word in silent reading, possibly over multiple fixations. However, these measures, which are typically obtained by registering the reader's eye movements with an eye-tracking system, are not necessarily the best estimates of word processing durations. As Kliegl, Nuthmann, and Engbert (2006) noted, the usefulness of reading times based on eye movements hinges on two assumptions. The first is the *immediacy-of-processing* assumption, or the idea that words are interpreted as soon as they are encountered (i.e., there are no instances in which word interpretation is temporarily suspended). The second is the *eye-mind* assumption, or the idea that readers continue to inspect a word for as long as that word is being processed. Several authors have noted that if both of these assumptions held to the full extent, reading would be considerably slower than we observe (e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Pollatsek, Fisher, & Rayner, 1998). In addition, the time spent on a word in reading continuous text is also influenced by linguistic factors beyond the word level: Some words are more predictable from the preceding context than others and some sentences are syntactically or pragmatically more challenging than others (see references in Rayner, 2009). Additionally, the time eyes spend on a word in continuous reading is influenced by the visuo-oculomotor planning of upcoming saccades, spill-over effects of the difficulty of earlier visual and cognitive processing and under some conditions by the material available for pre-processing via parafoveal preview.

The most common alternatives to measures derived from eye tracking, are lexical decision times and naming times. In the lexical decision task, participants are presented with letter strings and have to decide whether the string is an existing word or not. The time needed to reach a decision is the main dependent variable of interest. In the naming task, single words are presented on a computer screen and participants have to read the word aloud as rapidly as possible. The time required to start speaking the word (the voice onset time) is the dependent variable of interest here and is registered using a voice key (a microphone in combination with specialized hardware or software). In both tasks, the order of the stimuli is typically randomized.

Naming times and lexical decision times have been questioned as measures of word recognition. For example, many words can be named through sublexical grapheme-phoneme correspondences, making it unclear to what extent word naming times reflect lexical processing. A worrying finding is this respect is that the first phoneme is the most important variable in naming times of monosyllabic words (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Cortese & Khanna, 2007), whereas no such effect is found for lexical decision times or eye-movement times. A weakness of lexical decision times is that they depend on the nonwords used (e.g., Gibbs & Van Orden, 1998; Keuleers & Brysbaert, 2011). Moreover, the frequency effect, i.e., the difference in reaction time between frequent words and infrequent words, is large in the lexical decision task than in naming (Balota & Chumbley, 1984). Furthermore, the lexical decision task incorporates a decision-making component, which may decrease the construct validity of this variable as an index of the cognitive processes underlying reading-for-comprehension or pure word-recognition tasks. Finally, typical naming times and lexical decision times are more than twice as long as typical fixation times in word reading (500-600 ms vs. 200-250 ms), raising the question as to what confounds this extra time may introduce to the lexical processing latencies (Rayner & Pollatsek, 1989).

A systematic exploration of the relationship between the measures obtained by lexical decision, naming, and eye tracking is motivated by two reasons. First, finding high correlations between the three measures would be reassuring, because it would indicate that each task assesses the same construct (word recognition) and is not subject to excessive task-specific variability, or, as stated by Rayner and Pollatsek (1989, p. 68): “Each technique has its own problems, but if enough techniques appear to converge on a common answer, we can have a reasonable degree of confidence in our conclusion”.

The second reason is that the correlations between measures derived from eye-tracking and measures derived from single word recognition tasks are informative for theories of eye movement control in reading, in particular with respect to the issue of serial vs. parallel processing. Throughout the history of eye movement research, some investigators have defended a position close to the immediacy-of-processing assumption and the eye-mind assumption. In their view, words in text are processed and integrated one by one and eye movements closely mimic this process. An example is the Reader model of Just and

Carpenter (1980; for a comparison of models, see Reichle, Rayner & Pollatsek, 2003). In contrast, proponents of the parallel processing account have postulated a much looser coupling between eye movements and lexical processing, such that the fixations on word *n* are influenced not only by the processing of word *n*, but also by processing of the previous word *n-1* and the following words, for instance, *n+1* and *n+2*. A recent example of this account is the SWIFT model (Engbert et al., 2005; Kliegl et al., 2006; Kliegl, 2007; also see Kennedy, Pynte, & Ducrot, 2002; Kennedy & Pynte, 2005, 2009), which argues that the words going from *n-1* to *n+2* are processed in parallel and influence eye movement control. In-between these extremes is the E-Z Reader model (Rayner, Pollatsek, Drieghe, Slattery, & Reichle, 2007; Reichle, Liversedge, Pollatsek, & Rayner, 2009), which assumes word processing in text reading to be serial, but accepts that eye movements do not always accurately reflect this processing due to spill-over effects from the previous word, inflated fixation durations prior to skipping the next word, saccade execution errors, and effects from sentence and discourse processing (Rayner, 2009; Reichle, McConnell, & Warren, 2009). Both the SWIFT and E-Z Reader model have parameters that allow researchers to vary the degree of coupling between foveal word processing and eye movement control. We believe that having knowledge of the correlations between single word processing times, coming from naming, lexical decision, and eye movement data will be informative in estimating this degree of coupling. Finding high correlations would be more in line with a tight coupling (a serial model with a position close to the immediacy-of-processing assumption and the eye-mind assumption), whereas low correlations would be more in line with a loose coupling (parallel processing or frequent discrepancies between eye position and word processing).

A further interesting question is to what extent the correlations between the various measures go beyond those that may be expected on the basis of word frequency and word length. Word frequency is by far the most important variable in word processing (Brysbaert, Buchmeier, Conrad, Jacobs, Bölte, & Böhl, 2011; Brysbaert & New, 2009; Murray & Forster, 2004) and is known to influence all dependent variables of word processing tasks. Word length is an easy to control variable included in all studies, even though its effect is much smaller and not well established (e.g., New, Ferrand, Pallier, & Brysbaert, 2006). Knowing how strongly word processing times correlate in various tasks once the effects of word frequency and word length are partialled out, informs us about the degree to which lexical processing in the various tasks is affected by other word characteristics.

In summary, by correlating eye movement measures with lexical decision times and naming times, we may get better insight not only into the factors playing a role in eye movement control in reading, but also into the factors influencing word recognition times in various tasks. Finding high correlations would be easiest to interpret, because these would be in line with the assumption that the various tasks assess the same construct (lexical processing). In contrast, low correlations would be more difficult to interpret, because they could be due either to a loose coupling between eye movements and word processing or to the fact that naming times and lexical decision times are less valid estimates of word processing (not to mention the possibility that low correlations could also be the result of badly run experiments). Below, we explore the correlations in four different datasets and interpret the findings in terms of the serial vs. parallel processing debate in eye movement control and in terms of the validity of different measures of lexical processing.

Setting the stage: an analysis of the Schilling et al. (1998) dataset

The most often cited study comparing word naming latencies, lexical decision times, and reading times was published by Schilling, Rayner, and Chumbley (1998). In this study, 24 low-frequency and 24 high-frequency words, matched on length (6-9 letters, M = 6.9), were presented in a naming task, a lexical decision task, and a reading task in which participants read short, neutral sentences for meaning comprehension while their eye-movements were recorded. There were three groups of 16 participants, each group taking part in two of the tasks. We correlated the eye movement data of Schilling et al. (1998) with the naming times and the lexical decision times registered for these stimuli in the English Lexicon Project (ELP, Balota et al., 2007). The ELP is a database of lexical decision times and naming times for over 40,000 English words. The use of the ELP data allowed us to assess the correlations when the data were obtained from independent observers (remember that participants in Schilling et al.'s study always took part in two experiments).

There was only one word missing in the ELP database (HORNETS), giving us chronometric (lexical decision and naming) and eye-movement measures for 47 words from the original data set. Available eye movement data included first fixation duration (FFD), single fixation duration and gaze duration (GD). FFD is the duration of the first fixation on a word, SFD is the fixation duration in cases where the word was only fixated once, and GD refers to the total time a participant fixated the word before leaving it for the first time (i.e., in first-pass reading). For other data sets we also had estimates of total fixation time (TT), the summed duration of all fixations on the word both in first-pass reading and in rereading.

Table 1 shows the correlations between all the behavioural measures. The correlations are either based on the raw measurements (in ms), or on residualised values (the behavioural latencies from which the effects of word frequency and word length were partialled out). Instead of the Brown corpus frequencies (Kucera and Francis, 1967) used by Schilling et al., we used the SUBTLEX-US word frequencies, which are based on a 51 million word corpus of English-language film and television subtitles and are superior to the Brown corpus in predicting word processing times (Brysbaert & New, 2009).

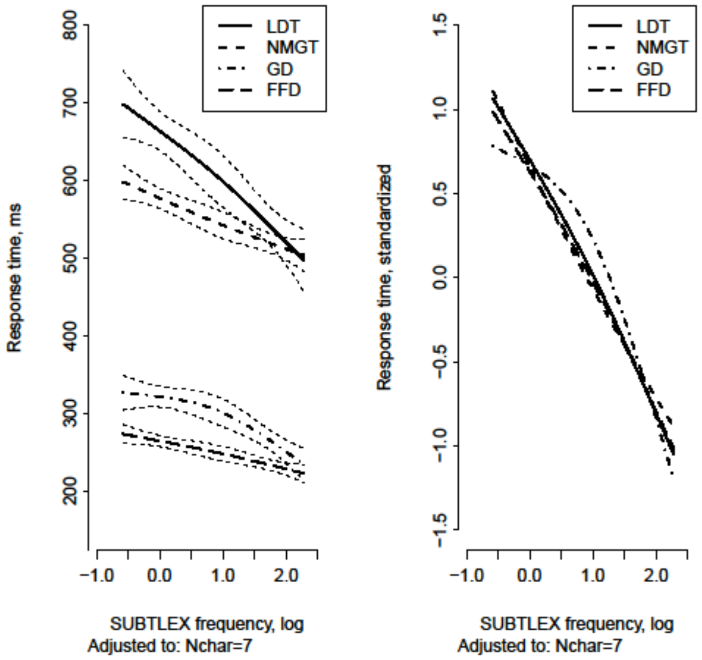
Table 1: Data from Schilling et al. (1998, Table 2), 47 words. Above the diagonal: Pearson's correlations between by-word averages on naming time (NMGT), lexical decision time (LDT), single fixation duration (SFD), first fixation duration (FFD), and gaze duration (GD), as well as with the residualised values (labelled with the prefix "r") of these variables from which word frequency and word length effects were partialled out. Below the diagonal: p-values of the correlations. Correlations between residualised chronometric and eye-movement measures are presented in bold.

	LDT	NMGT	FFD	SFD	GD	rLDT	rNMGT	rFFD	rSFD	rGD
LDT	*****	0.849	0.671	0.705	0.724	0.643	0.429	0.251	0.258	0.267
NMGT	<0.001	*****	0.547	0.566	0.545	0.434	0.651	0.157	0.120	0.062
FFD	<0.001	<0.001	*****	0.931	0.784	0.265	0.164	0.678	0.583	0.422
SFD	<0.001	<0.001	<0.001	*****	0.830	0.263	0.120	0.561	0.653	0.468
GD	<0.001	<0.001	<0.001	<0.001	*****	0.284	0.066	0.426	0.491	0.684
rLDT	<0.001	0.002	0.072	0.075	0.053	*****	0.667	**0.391**	**0.402**	**0.416**
rNMGT	0.003	<0.001	0.270	0.422	0.661	<0.001	*****	**0.242**	**0.184**	**0.096**
rFFD	0.088	0.290	<0.001	<0.001	0.003	0.007	0.101	*****	0.859	0.622
rSFD	0.079	0.423	<0.001	<0.001	<0.001	0.005	0.216	<0.001	*****	0.717
rGD	0.069	0.676	0.003	0.001	<0.001	0.004	0.521	<0.001	<0.001	*****

From Table 1 we see that the correlations between naming and lexical decision times were substantially higher than those with the eye movement measures. This was true both for the raw correlations and the partial correlations. For instance, the raw correlation between lexical decision and naming was .849, whereas that between lexical decision and gaze duration was .724. Of the eye movement measures gaze duration tended to have the numerically largest correlations with chronometric measures, and lexical decision times had a numerically larger correlation with eye movement data than naming times. The correlation between gaze duration and lexical decision time was .724, against .545 for the correlation between gaze duration and naming time. The closer relationship between lexical decision time and gaze duration compared to naming times and gaze duration became particularly clear when the effects of word frequency and word length were partialled out. Then, the correlation between gaze duration and naming time was no longer significant (r = .096, p = .521), whereas that with lexical decision time still was (r = .416, p = .004). There was also shared variance left between naming and lexical decision (r = .667, p < .001) after the effects of word frequency and word length were partialled out (in contrast to Forster & Chambers, 1973 who found no significant correlation between these two measures after partialling out frequency).

The left part of Figure 1 shows the frequency effect for various dependent variables (controlled for word length). It shows that the frequency effect is strongest for the variable with the longest RTs (lexical decision) and smallest for the variable with the shortest RTs (first fixation duration). To find out whether the differences in effect sizes were exclusively due to differences in overall RTs, we plotted the frequency effects on standardized (z-) scores (right part of Figure 1). This analysis showed that the frequency effect was very similar for the different dependent variables.

Figure 1: Effects of log word frequency on raw (left panel) and standardized (right panel) by-word averages of lexical decision times (LDT), naming times (NMGT), first fixation durations (FFD), gaze durations (GD), and total fixation time (TT), for 47 words from Schilling et al's (1998) dataset. Partial effects are adjusted to the median word length (NChar) of 7 letters. Dotted lines in the left panel indicate the 95%-confidence intervals. Word frequency is measured as log10(frequency per million words). So, -1 equals to a frequency of .1 per million (pm), 0 equals to 1 pm, and 2 refers to 100 pm.



Validation with a new study

To find out whether Schilling et al.'s (1998) data can be generalized, we analyzed a second, similar dataset, which had not been set up in the first place to test the effect of word frequency (although the words did vary considerably on this variable). The dataset came from a study run at the University of Southampton[1], in which the effects of caffeine on eye movements in reading were investigated (there were none, making the entire dataset useful for our purpose). Seventeen participants read 80 sentences each. There was one six-letter target word per sentence that was analysed. As in the previous tables, we compared eye movement data with the lexical decision and naming times from the English Lexicon Project. Table 2 reports the correlations between the by-word averages of the raw and residualised (i.e. those from which word frequency has been partialled out) behavioural measures for the Southampton eye movement study.

Table 2: Data set of 80 English six-letter content words from the Southampton eye movement study. Above diagonal: Pearson correlations between by-word averages of naming time (NMGT), lexical decision time (LDT), first fixation duration (FFD), single fixation duration (SFD), gaze duration (GD), and total fixation time (TT), as well as between residualised values (labelled with the prefix "r") of these variables from which word frequency effects have been partialled out. Below the diagonal: p-values for correlations. Correlations between residualised chronometric and eye-movement measures are presented in bold.

---

[1] The authors thank Valerie Benson, Hannah Walker and Julie Kirkby for kindly giving them access to the data.

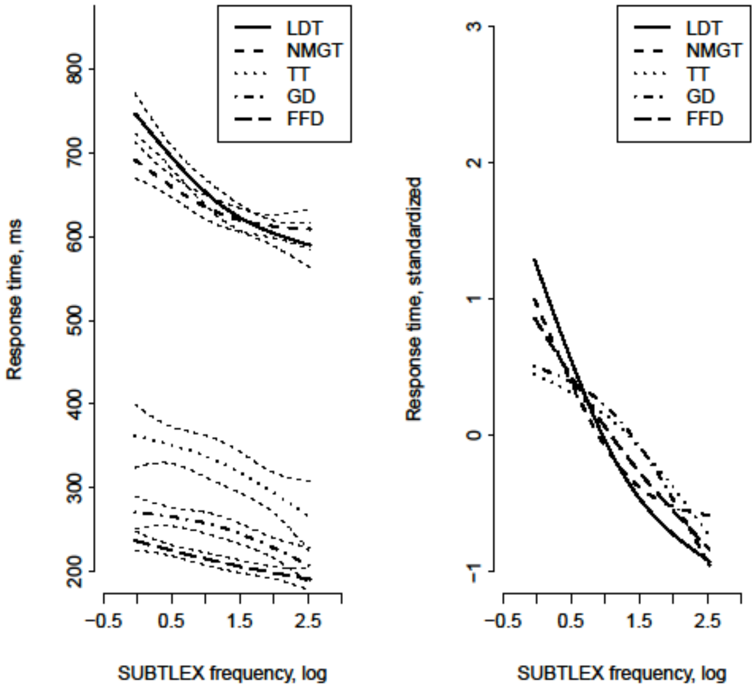
	LDT	NMGT	FFD	SFD	GD	TT	rLDT	rNMGT	rFFD	rSFD	rGD	rTT
LDT	*****	0.614	0.460	0.495	0.439	0.417	0.666	0.245	0.064	0.157	0.143	0.170
NMGT	<0.001	*****	0.372	0.399	0.251	0.302	0.306	0.834	0.092	0.152	0.035	0.128
FFD	<0.001	0.001	*****	0.704	0.655	0.450	0.079	0.092	0.834	0.507	0.463	0.267
SFD	<0.001	<0.001	<0.001	*****	0.528	0.419	0.207	0.160	0.535	0.877	0.379	0.278
GD	<0.001	0.025	<0.001	<0.001	*****	0.584	0.188	0.037	0.488	0.379	0.879	0.435
TT	<0.001	0.006	<0.001	<0.001	<0.001	*****	0.236	0.142	0.297	0.293	0.457	0.924
rLDT	<0.001	0.006	0.485	0.065	0.094	0.035	*****	0.367	**0.095**	**0.236**	**0.214**	**0.256**
rNMGT	0.029	<0.001	0.419	0.155	0.743	0.209	0.001	*****	**0.110**	**0.183**	**0.042**	**0.154**
rFFD	0.574	0.418	<0.001	<0.001	<0.001	0.008	0.400	0.332	*****	0.609	0.555	0.321
rSFD	0.163	0.177	<0.001	<0.001	0.001	0.008	0.035	0.105	<0.001	*****	0.431	0.317
rGD	0.207	0.757	<0.001	0.001	<0.001	<0.001	0.057	0.711	<0.001	<0.001	*****	0.495
rTT	0.131	0.258	0.016	0.012	<0.001	<0.001	0.022	0.174	0.004	0.004	<0.001	*****

In general the correlations were lower than in the analysis of the Schilling et al.'s data, but for the rest entirely in line with those of Table 1. Again, correlations of eye-movement latencies with naming times were lower than those with lexical decision times, and first fixation duration correlated less strongly with lexical decision and naming than the other eye movement variables. Once the effect of word frequency was taken into account, there was no significant variance left to explain between eye movement data and naming latencies. The partial correlations with lexical decision hovered around .22 (5% variance), which was borderline significant.

To get a better picture of the frequency effect on the different dependent variables, we again plotted the effect against raw RTs and z-scores (Figure 2). We defer discussion of this figure until we have analysed all datasets.

Figure 2: Effects of log word frequency on average raw (left panel) and standardized (right panel) lexical decision times (LDT), naming times (NMGT), first fixation durations (FFD), gaze durations (GD), and total fixation time (TT), based on the Southampton reading

experiment data. Dotted lines in the left panel indicate the 95%-confidence intervals indicated in the left panel. Word frequency is measured as log10(frequency per million words).



Extension to natural reading: Findings from the Dundee corpus

Part of the argument in the serial vs. parallel debate is whether eye movement data should come from a limited set of carefully controlled target words in purpose-built carrier sentences, or from natural reading conditions. For instance, Rayner et al. (2007, p. 524) asserted that: “we believe that controlled experiments are the better source of data because there are always serious problems interpreting correlational analyses (e.g., the relevant variables are often quite confounded and may thus be virtually impossible to unconfound through the use of such techniques) and because controlled experiments allow for much stronger inferences about causality". On the other hand, Kliegl et al. (2006, p. 13, 31) argued that: "We refer to reading without experimental manipulations of target words and reading without gaze-contingent display changes as *natural reading*. ... The unique contribution of eye-movement analyses of natural reading is the possibility to examine simultaneously a large number of variables (and their potential redundancy) as well as interactions among them (given sufficient statistical power)." This is the long-standing question about whether natural conditions allow enough control to disentangle possible confounds, vs. whether the introduction of controls impoverishes the situation to such an extent that specific biases are introduced.

Again, we aim to inform this discussion by a direct comparison of lexical decision and naming times involving single word processing with reading data from connected text, such as sentences and passages. For a study of connected text-reading we consider in this section the Dundee corpus (Kennedy & Pynte, 2005), which currently is the largest corpus of eye movement data available in the English language. It contains the eye tracking records of 10 participants each reading 51,000 words from English newspaper editorials. There were 6,817 word types attested in this corpus for which the ELP also provided lexical decision and naming times. Table 3 reports the correlations between the various behavioural measures for those words.

Table 3: The data set included 6,817 English words common to the Dundee corpus and the English Lexicon Project. Above the diagonal: Pearson correlations between by-word averages of naming time (NMGT), lexical decision time (LDT), first fixation duration (FFD), single fixation duration (SFD), gaze duration (GD), and total fixation time (TT), as well as between residualised values (labelled with the prefix "r") of these variables from which

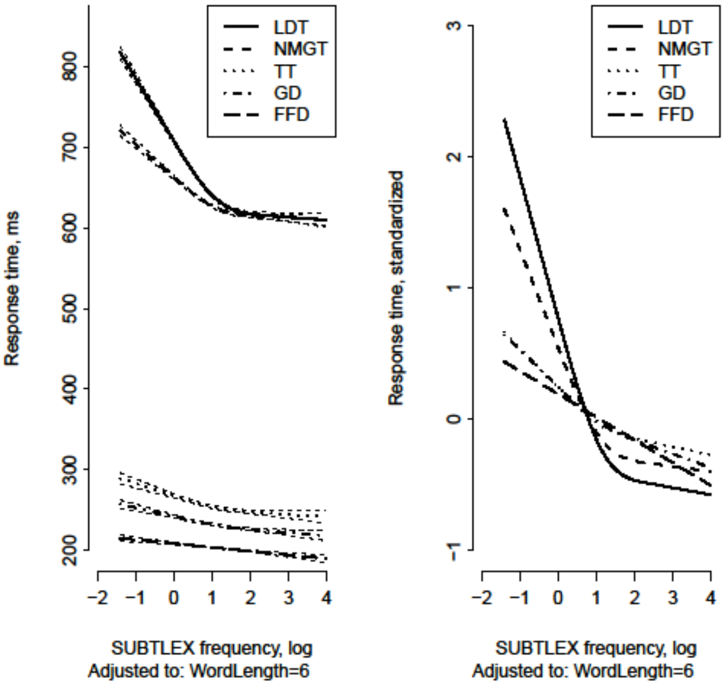
word frequency and word length effects have been partialled out. Below the diagonal: p-values for correlations. Correlations between residualised chronometric and eye-movement measures are presented in bold.

	LDT	NMGT	FFD	SFD	GD	TT	rLDT	rNMGT	rFFD	rSFD	rGD	rTT
LDT	*****	0.542	0.115	0.167	0.240	0.233	0.748	0.209	0.013	0.030	0.032	0.033
NMGT	<0.001	*****	0.103	0.140	0.227	0.229	0.230	0.820	0.018	0.024	0.033	0.043
FFD	<0.001	<0.001	*****	0.898	0.710	0.570	0.017	0.022	0.986	0.882	0.706	0.555
SFD	<0.001	<0.001	<0.001	*****	0.690	0.583	0.039	0.028	0.873	0.975	0.663	0.549
GD	<0.001	<0.001	<0.001	<0.001	*****	0.829	0.040	0.038	0.670	0.637	0.936	0.754
TT	<0.001	<0.001	<0.001	<0.001	<0.001	*****	0.041	0.050	0.531	0.531	0.759	0.942
rLDT	<0.001	<0.001	0.264	0.012	0.011	0.008	*****	0.280	**0.018**	**0.040**	**0.043**	**0.044**
rNMGT	<0.001	<0.001	0.164	0.071	0.015	0.001	<0.001	*****	**0.022**	**0.029**	**0.041**	**0.053**
rFFD	0.397	0.247	<0.001	<0.001	<0.001	<0.001	0.257	0.158	*****	0.895	0.716	0.563
rSFD	0.055	0.129	<0.001	<0.001	<0.001	<0.001	0.010	0.064	<0.001	*****	0.680	0.563
rGD	0.041	0.033	<0.001	<0.001	<0.001	<0.001	0.006	0.009	<0.001	<0.001	*****	0.806
rTT	0.036	0.005	<0.001	<0.001	<0.001	<0.001	0.005	0.001	<0.001	<0.001	<0.001	*****

Although Table 3 largely replicates the pattern of the previous analyses, there is a considerable drop in percentage of variance accounted for. Indeed, when word frequency and word length are partialled out, there is only .2% of shared variance left between lexical decision and gaze duration. Although this association is still statistically significant due to the large number of observations, it is becoming negligibly small in practical terms.

Figure 3 shows the effect of word frequency on the various dependent variables (both raw RTs and standardized scores). A comparison with Figures 1 and 2 indicates that the range of word frequencies in the Dundee corpus is much larger, going from .01 per million (pm) to 10,000 pm, than that in the small-scale factorial designs described in the first two analyses, going from about 1 pm to slightly over 100 pm. At the same time, two new findings emerge. First, in connected text the word frequency effect is smaller than in isolated word processing (both naming and lexical decision; see also Forster, Guererra, & Elliot, 2009). Second, for lexical decision and word naming, but not for word reading data, the frequency effect levels off for frequencies above 50 pm (log = 1.7). The floor effect for high frequencies in the lexical decision task has been reported before by other authors (Balota et al., 2004; Baayen et al., 2006; Ferrand et al., 2010; Keuleers, Diependaele, & Brysbaert, 2010b; Keuleers, Lacey, Rastle, & Brysbaert, in press).

Figure 3: Partial effects of log word frequency on raw (left panel) and standardized (right panel) lexical decision times (LDT), naming times (NMGT), first fixation durations (FFD), gaze durations (GD), and total fixation time (TT), based on eye-movement data from the Dundee corpus and the chronometric data from the English Lexicon Project for 6,817 words. Partial effects are adjusted to the median word length (NChar) of 6 letters. Dotted lines in the left panel indicate the 95%-confidence intervals. Word frequency is measured as log10(frequency per million words).



A cross-validation in Dutch

To avoid drawing conclusions from a single dataset or a single language, we repeated the corpus study with recently gathered Dutch data. The eye-movement data were obtained from the Dutch Eye-Movement Online Internet Corpus (Kuperman, Dambacher, Nuthmann, & Kliegl, 2010).[2] In this study, 28 participants read 220 isolated sentences, which contained a total of 893 different words (excluding the first and the last word of each sentence). For each word the mean gaze duration and the mean single fixation duration were calculated. The lexical decision times came from the Dutch Lexicon Project (Keuleers et al., 2010b), in which 39 participants made responses to 14,000 monosyllabic and disyllabic words. There were 545 word types in common between the eye movement corpus and the Dutch Lexicon Project. Because all words were monosyllabic or disyllabic, the ranges of the variables were more restricted. In general, this resulted in lower correlations.

Table 4 shows the correlations between the various lexical processing measures: both for the raw RTs and the residualised values from which the effects of log word frequency and word length were partialed out. Estimates of word frequency were based on the SUBTLEX-NL corpus (a corpus of 43.7 million words derived from film subtitles in Dutch; Keuleers, Brysbaert, & New, 2010a).

Table 4: The data set included 545 words common to the Dutch eye-movement corpus and the Dutch Lexicon Project. Above diagonal: Pearson's correlations between by-word

---

[2] Correlations of LD times with EM obtained in another eye-tracking study of sentence reading in Dutch (Kuperman, Bertram, & Baayen, 2010) were virtually identical to the ones reported in Table 8, so this study is not presented further.

averages of lexical decision time (LDT), first fixation duration (FFD), single fixation duration (SFD), gaze duration (GD), and total fixation time (TT), as well as between residualised values (labelled with the prefix “r”) of these variables from which word frequency effects have been partialled out. Below the diagonal: p-values for correlations. Correlations between residualised chronometric and eye-movement measures are presented in bold.

	LDT	FFD	SFD	GD	TT	rLDT	rFFD	rSFD	rGD	rTT
LDT	*****	0.170	0.189	0.224	0.217	0.849	0.036	0.045	0.040	0.036
FFD	<0.001	*****	0.979	0.833	0.594	0.041	0.963	0.942	0.834	0.565
SFD	<0.001	<0.001	*****	0.834	0.590	0.051	0.936	0.957	0.829	0.555
GD	<0.001	<0.001	<0.001	*****	0.772	0.042	0.766	0.766	0.884	0.629
TT	<0.001	<0.001	<0.001	<0.001	*****	0.038	0.522	0.516	0.634	0.890
rLDT	<0.001	0.336	0.235	0.325	0.381	*****	**0.043**	**0.053**	**0.048**	**0.042**
rFFD	0.396	<0.001	<0.001	<0.001	<0.001	0.318	*****	0.978	0.866	0.586
rSFD	0.292	<0.001	<0.001	<0.001	<0.001	0.215	<0.001	*****	0.866	0.580
rGD	0.345	<0.001	<0.001	<0.001	<0.001	0.266	<0.001	<0.001	*****	0.712
rTT	0.404	<0.001	<0.001	<0.001	<0.001	0.325	<0.001	<0.001	<0.001	*****

Figure 4: Partial effects of log word frequency on raw (left panel) and standardized (right panel) average lexical decision times (LDT), first fixation durations (FFD), gaze durations (GD), and total fixation time (TT), based on the data from the Dutch eye-movement corpus and the chronometric data from the Dutch Lexicon Project for 545 words. Partial effects are adjusted to the median word length (NChar) of 5 letters. Dotted lines in the left panel indicate the 95%-confidence intervals.

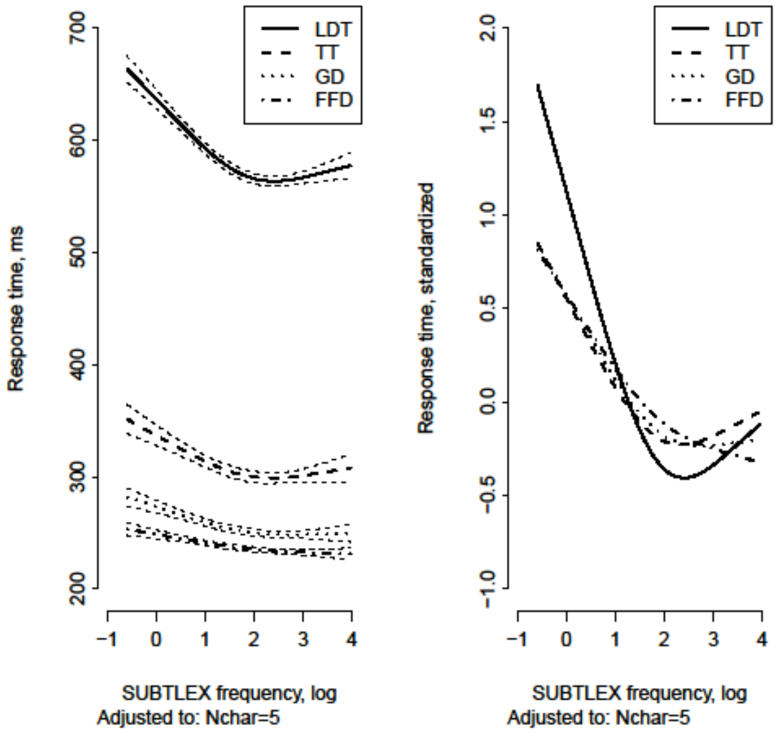


Table 4 and Figure 4 largely confirm the findings of the Dundee corpus (Table 3 and Figure 3). Gaze duration shows the numerically largest correlation with lexical decision times, but even for this variable there is virtually no shared variance with lexical decision times (.2%) once the effect of word frequency (and to a lesser extent word length) is controlled for. Again, the word frequency effect is numerically larger in lexical decision than in continuous text reading, and it levels off at high frequency values. This is more so in lexical decision than in eye-movement measures of sentence reading.

Discussion

We examined the correlations between eye movement data, lexical decision times, and naming times, to see to what extent word recognition measures based on eye movements in reading capture the same processes as those based on vocal and manual responses to individually presented words (for a similar examination of eye-movement latencies and components registered in the event-related potentials, see Dambacher & Kliegl, 2007).

Our examination of four datasets resulted in the following findings:

1. Eye movement latencies correlate less strongly with lexical decision times and naming times than the latter two measures correlate with each other; that is, there is more variance shared between naming and lexical decision than between either of these and eye movement measures. We also observed a stronger correlation of eye-movement latencies with lexical decision than with naming times, which we discuss below.
2. Eye movement data correlate more strongly with lexical decision times in factorial experiments that focus on a small set of target words than in corpus studies that include a variety of words. The shape of the frequency effect on standardized latencies in the former studies is also more similar to that in lexical decision (see Figures 1 and 2 vs. Figures 3 and 4). Across studies, gaze duration is the eye-movement measure that shows the strongest correlation with both types of chronometric latencies.
3. The correlations between gaze durations and naming or lexical decision times are largely due to word frequency. When the effect of word frequency (and word length) are partialled out, the percentages of shared variance decreased to 17% in Schilling et al. (down from 52%), 5% in the Southampton study (down from 19%), and .2% in the two corpus studies (down from 5%). Interestingly, correlations between raw and residualised behavioural latencies are strong (e.g., r = .849 between raw and residualised lexical decision times in the Dutch data). This suggests that a large portion of the residual variance in each behavioural measure is systematic rather than noise (e.g., due to the speed of information uptake,

perceptual discrimination, preparation and execution of responses, ...). Importantly, however, this variance is unique to each measure and not shared with the other tasks.

These findings allow us to draw several conclusions with respect to eye movement data. First, gaze durations correlate more strongly with lexical decision times and naming times when they are based on a limited number of target words that have been included in neutral carrier sentences (Schilling et al., the Southampton study) than when they are based on correlational analyses of unselected materials (Dundee, the Dutch study). This result is best understood in the context of recent findings of Radach et al. (2008) who in an eye-tracking study manipulated the presentation format of the sentences (isolated sentence versus passage) and the reader's task (comprehension versus verification). Among other phenomena, Radach et al. observed that gaze durations were shorter in passage-embedded sentences than in same sentences presented in isolation, but that total reading times were longer. Our study corroborated the first part of their pattern, as the gaze durations were considerably longer in the single-sentence Schilling et al. study (M = 291 ms) than in the Dundee corpus (M = 252 ms; word lengths limited to 6 letters and longer).[3] Radach et al. (2008) concluded that the preferred strategy for passage reading is a rapid assessment of content in the first pass, potentially followed by multiple re-reads (on the role of presentation format see also Britt et al., 1992; Binder et al., 2001). In our study, we compared words presented in isolation, to words presented in isolated sentences, to words presented in sentences embedded in larger contexts. It appears that reading isolated sentences for comprehension - with a more elaborate processing of words in first-pass

---

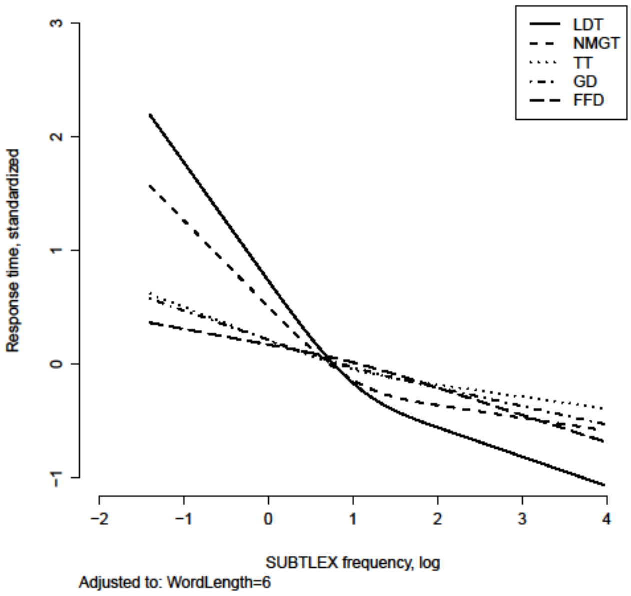
[3] The differences between Schilling et al. and the Dundee corpus reported in this ms remained when we removed the 1,800 words with lengths lower than 6 letters from the Dundee corpus.

reading and relatively uncommon re-reading - yields eye-movement patterns more similar to the ones elicited by reading isolated words.

Several factors may underlie this similarity. First, in factorial experiments target words are usually included in sentences with a low contextual constraint and thus have a low predictability (unless effects of predictability are explicitly tested). Of course, no context (apart of potential list effects) is available in isolated word reading. This is different from natural texts where words may be predictable from the preceding words. In this respect, it is interesting to see that the frequency effect is smaller in continuous text reading (Figures 3 and 4) than in target words not predicted by the preceding context (Figures 1 and 2), arguably because rare words are easier to process in a coherent context than in isolation. This converges with a weaker effect of word frequency in passage-embedded sentences as compared to same sentences presented in isolated (Radach et al., 2008). This finding may also be in line with claims that the word frequency effect is better understood as an effect due to the ease with which a context for the word can be activated (i.e. the contextual diversity; Adelman, Brown, & Quesada, 2006; Baayen, 2011; McDonald & Schillcock, 2001). Likewise, the degree of semantic integration of words into sentences and words and sentences into discourse may vary between single-line sentence reading (Schilling et al.'s data, Southampton data and Dutch data), and passage reading (the Dundee corpus). In summary, the present evidence expands on Radach et al. (2008) by showing a stronger similarity between the speed of processing and word frequency effects observed in words in isolation and words in isolated (rather than passage-embedded) sentences. We note that the present study conflates the presentation format with task (isolated words in lexical decision and naming; isolated sentences in reading for comprehension verified by questions; and passage reading for unverified comprehension): future research is required to make a more controlled comparison.

Second, the word frequency effect in lexical decision and naming levels off for high frequencies (roughly those of more than 50 pm), whereas this floor effect is less pronounced in eye movement latencies. Our conjecture at present is that this is largely due to the function words, such as prepositions and conjunctions. Many of these words have a high frequency but tend to be responded to slowly in lexical decision and naming (possibly because they are not expected in lists of words). To test this hypothesis, we excluded 120 function words from the word list of the Dundee corpus and replotted the functional relationship between behavioural latencies and log frequency. Figure 5 to some extent confirms our conjecture as it reveals that the levelling-off of lexical decision and naming times for the highest frequencies is less pronounced in the absence of function words, compare to Figure 3.

Figure 5: Partial effects of log word frequency on standardized lexical decision times (LDT), naming times (NMGT), first fixation durations (FFD), gaze durations (GD), and total fixation time (TT) in the absence of function words. Figure based on the eye-movement data from the Dundee corpus and the chronometric data from the English Lexicon Project for 6,693 content words. Partial effects are adjusted to the median word length (NChar) of 6 letters. Word frequency is measured as log10(frequency per million words).



Another possibility is that the differential floor effects are due to different lower bounds of processing times used in the various tasks. For instance, Kinoshita and Lupker (2002) reported that word naming times differ as a function of the context in which the words are presented. In particular, the authors showed that the size of the frequency effect was reduced in an environment of low frequency regular word fillers relative to an environment of low-frequency exception filler words. They interpreted this finding as evidence for a flexible time criterion in word naming. We leave the exploration of this possibility to future research.

A last important finding of our analyses with respect to eye movement data is that gaze durations correlate most strongly with lexical decision times. The rather low correlations with first fixation duration (FFD) are noteworthy because first fixation duration is often considered as an early index of lexical access while measures including subsequent fixations are thought to be more sensitive to the higher level syntactic and discourse processing, as well as to the semantic integration of the word into the sentence framework (for a discussion see e.g., Inhoff, 1989, and references therein). On the basis of this reasoning we could have expected that FFD would correlate most with lexical decision times and naming latencies. One reason for the low correlation between FFD and lexical decision may be that some first fixations tend to be short because of landing errors, resulting in a quick saccade towards the intended landing position: this scenario is not found when there is a single fixation on the word. Another reason may be that lexical decision times include semantic processing, so that their interpretation comes closer to that of GD than of FFD.

Single fixation duration is sometimes proposed as a better alternative of FFD, because it is not subject to cases in which words are refixated as a result of faulty landing. Except for the Southampton study with six-letter target words, there was little evidence for the advantage for SFD over GD in our data. On the basis of our experiences with eye movement research, we think that the best eye movement measure for examining word processing may be contingent on word length. For words of 3 letters and less, skipping rate is the preferred measure, as these words are more often skipped than fixated (Brysbaert, Drieghe & Vitu, 2005). For words between 4 and 8 letters single fixation duration is indicated, as these words are shorter than average saccade length (about 8 letters) and, therefore, are likely to be processed in a single fixation. For words longer than 8 letters refixations are frequent enough to assume that gaze duration is the most sensitive measure of word recognition time (see e.g. Rayner, Sereno & Raney, 1996 for refixation averages).

Our findings also have implications for the word recognition literature. First, they confirm that naming latencies are a less good measure of lexical processing than lexical decision times. Indeed, in all analyses naming latencies correlated less with gaze durations than lexical decision times. This agrees with the fact that words can be named on the basis of sublexical grapheme-phoneme correspondences and with the observation that naming times are strongly influenced by the first phoneme (Balota et al., 2004), at least for monosyllabic words (for multisyllabic words the stress pattern seems to be particularly important; Yap & Balota, 2009). It also dovetails well with earlier arguments that sounding out pronounceable words in the word naming task may not require lexical access and, on some theories, would be successful even in the individuals with impaired access to lexical semantics (e.g. Coltheart et al., 2001).

Another important finding is the confirmation that the correlation between gaze duration and lexical decision/naming time is overwhelmingly due to word frequency. This agrees with recent research showing the impact of this variable for the prediction of word processing times in megastudies (Baayen et al., 2006; Balota et al., 2004; Brysbaert et al., 2011; Ferrand et al., 2010; Keuleers, Diependaele, & Brysbaert, 2010b; Keuleers, Lacey, Rastle, & Brysbaert, in press). It also agrees with Murray and Forster's (2004, p. 721) assertion that: "Of all the possible stimulus variables that might control the time required to recognize a word pattern, it appears that by far the most potent is the frequency of occurrence of the pattern ... Most of the other factors that influence performance in visual word processing tasks, such as concreteness, length, regularity and consistency, homophony, number of meanings, neighborhood density, and so on, appear to do so only for a restricted range of frequencies or for some tasks and not others". It will be interesting to search for the variables that predict the remaining shared variance between gaze durations and lexical decision times for the controlled studies with target words (Schilling et al., the Southampton study) and to examine whether these variables also predict some shared variance in the uncontrolled eye movement corpus studies. Another interesting question for further research is whether the relationship between word processing time and log word frequency is indeed more linear for reading data than for lexical decision times and word naming times. The nonlinear relationship for lexical decision times has been used by some authors to propose word frequency measures other than log frequency (Adelman, Brown, & Quesada, 2006; Murray & Forster, 2004, 2008). It will be important to find out to what extent this deviation (and the need for alternative frequency measures) is task dependent.

With respect to the link between eye movement data and lexical processing, we forwarded the hypothesis that knowledge concerning the correlations between a single word recognition task and eye movement measures would be informative in gauging the degree of coupling between the lexical processing of a specific word and the eye movement measures observed on that word. Finding a high correlation would be more in line with a tight coupling as proposed by theories which adhere strongly to the immediacy-of-processing assumption and the eye-mind assumption (e.g. Just & Carpenter, 1980), whereas low correlations would indicate a more loose coupling.

The rather low observed correlations between fixation times and single word recognition measures cast doubt on the feasibility of a very strong position on the eye-mind assumption and are more in line with models that allow for effects originating from previous words (such as spill-over effects and effects of for instance syntactic integration), or upcoming materials to influence fixation times such as the E-Z Reader model and the SWIFT model. Moreover, our data suggest that this coupling is dependent on the stimulus materials and the reader's purposes. The fact that the correlations are even lower in corpus studies compared to studies examining fixation durations on unpredictable words embedded in single sentences, is compatible with this statement. It is reasonable to assume that effects of predictability (though not word frequency, as shown here and in Radach et al., 2008) will play a bigger role in coherent samples of natural text, compared to single sentences designed to be meaningful on their own.

A comparison which might allow further teasing apart predictions from the E-Z Reader model and the SWIFT model would be examining parafoveal-on-foveal effects in isolated sentences versus paragraphs (for a similar study of the effect of task on parafoveal-on-foveal effects see Wotschak & Kliegl, in press). E-Z Reader, which regards parafoveal-on-foveal effects as resulting from a number of potential sources of noise (Drieghe, 2011), would not predict a difference between these two different reading environments. A difference might be predicted by for instance the SWIFT model but it would require a simulation to examine whether this prediction would be an increase of the parafoveal-on-foveal effect (i.e. by more parallel lexical processing), a decrease in the effect (i.e. because parafoveal-on-foveal effects are numerically small they might be drowned out by all the other influences impacting the eye movement record during passage reading), or no difference.

Finally, our analyses have confronted us with the scarcity of corpus data about eye movements in reading. In the past years authors have invested heavily in the collection of lexical decision data for tens of thousands of words in so-called megastudies. Data of comparable size are lacking for eye movements. The closest is the Dundee corpus with its 6,817 distinct words. However, this corpus is limited because many low frequency words only occurred in a single sentence, making the estimates of the processing times sentence-dependent. To improve the quality of the eye movement data, it would be better to make sure that each word appears in a number of sentences presented at different times in the study, just like in lexical decision and naming experiments word order is randomised anew for each participant.

Having data for many words from well-controlled eye movement studies would allow us to include these data in the Dutch and English Lexicon Projects, so that it becomes easy for researchers to investigate which word variables have an effect across tasks and which are task-specific. An additional advantage would be that not only the statistical significance of a variable can be assessed, but also the precise curve of the relationship, as shown in Figures 1-4, and the percentage of variance accounted for by each variable. Such an approach is likely to be particularly fruitful for a better understanding of the processes underlying word recognition.

References

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

Baayen (2011). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon, 5(3)*, 436-461.

Balota, D.A., & Chumbley, J.I. (1984). Are lexical decisions a good measure of lexical access: The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 340-357.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459.

Binder, K. S., Duffy, S. A., & Rayner, K. (2001). The effects of thematic fit and discourse context on syntactic ambiguity resolution. *Journal of Memory and Language, 4*, 297–324.

Britt, M. A., Perfetti, C. A., Garrod, S., & Rayner, K. (1992). Parsing in discourse: Context effects and their limits. *Journal of Memory & Language, 31*, 293-314.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A.M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology, 58,* 412-424.

Brysbaert, M., Drieghe, D., & Vitu, F., (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive Processes in Eye Guidance*. Oxford: University Press, pp. 53 - 77.

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

Carr, T.H., Pollatsek, A., Posner, M.I., & Snyder, C.R.R. (1979). Orthography and familiarity effects in word processing. *Journal of Experimental Psychology: General, 108*, 389-414.

Coltheart, M., Rastle, K., Perry, C., Ziegler, J. & Langdon, R. (2001). DRC: A dual route cascaded model of word recognition and reading aloud. *Psychological Review, 108*, 204-256.

Dambacher, M., & Kliegl, R. (2007). Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. *Brain Research, 1155*, 147-162.

Drieghe, D. (2011). Parafoveal-on-Foveal effects in Eye Movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *Oxford Handbook on Eye Movements*. Oxford: University Press, pp. 839 – 855.

Drieghe, D., Rayner, K., & Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *Quarterly Journal of Experimental Psychology, 61*, 1239 – 1246.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Meot, A., Augustinova, M., & Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods, 42*, 488-496.

Forster, K., Guererra, C. & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods, 41*, 163 – 171.

Gibbs, P., & Van Orden, G.C. (1998). Pathway selection's utility for control of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1162-1187.

Harrell, F.E (2001). *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression and Survival Analysis.* New York: Springer-Verlag.

Inhoff, A.W. (1989). Lexical Access during Eye Fixations in Reading: Are Word Access Codes Used to Integrate Lexical Information across Interword Fixations? *Journal of Memory and Language, 28*, 444-461.

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. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research, 45*, 153-168.

Keuleers, E., Brysbaert, M., & New, B. (2010a). SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods, 42*, 643-650.

Keuleers, E., Diependaele, K. & Brysbaert, M. (2010b). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Language Sciences*.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (in press). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. Behavior Research Methods.

Kinoshita, S., & Lupker, S.J. (2002). Effects of filler type in naming: Change in time criterion or attentional control of pathways? *Memory & Cognition, 30*, 1277-1287.

Kinoshita, S., & Mozer, M.C.(2006). How lexical decision is affected by recent experience: Symmetric versus asymmetric frequency-blocking effects. *Memory & Cognition, 34*, 726-742.

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

Kuperman, V., Bertram, R., & Baayen, R.H. (2010). Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language, 62*, 83-97.

Kuperman, V., Dambacher, M., Nuthmann, A., & Kliegl, R. (2010). The effect of word position on eye-movements in sentence and paragraph reading. *Quarterly Journal of Experimental Psychology, 63*, 1838-1957.

McDonald, S., & Shillcock, R. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech, 44*, 295-323.

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

Murray, W.S., & Forster, K.I. (2008). The rank hypothesis and lexical decision: A reply to Adelman and Brown (2008). *Psychological Review, 115*, 240-251.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Re-examining word length effects in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review, 13*, 45-52.

Proctor, R.W. (1981). A unified theory for matching-task phenomena. *Psychological Review, 88*, 291-326.

Radach, R., Huestegge, L., & Railey, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research, 72*, 675–688.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*, 1457-1506.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. New York: Prentice Hall.

Rayner, K., Sereno, S., C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models.

Reichle, E.D., McConnell, K., & Warren, T. (2009). Using E-Z Reader to Model the effects of Higher-Level Language Processing on Eye Movements During Reading. *Psychonomic Bulletin & Review, 16 (1)*, 1 -21.

R Development Core Team. (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

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 & Cognition, 26*, 1270-1281.

Wotschak, C., & Kliegl, R. (in press). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *Quarterly Journal of Experimental Psychology*.