Journal of Experimental Psychology:
Learning, Memory, and Cognition
2014, Vol. 40, No. 1, 275–283

0278-7393/14/$12.00 DOI: 10.1037/a0034058

RESEARCH REPORT

The Influence of Contextual Diversity on Eye Movements in Reading

Patrick Plummer
University of California, San Diego

Manuel Perea
Universitat de València

Keith Rayner
University of California, San Diego

Recent research has shown contextual diversity (i.e., the number of passages in which a given word appears) to be a reliable predictor of word processing difficulty. It has also been demonstrated that word-frequency has little or no effect on word recognition speed when accounting for contextual diversity in isolated word processing tasks. An eye-movement experiment was conducted wherein the effects of word-frequency and contextual diversity were directly contrasted in a normal sentence reading scenario. Subjects read sentences with embedded target words that varied in word-frequency and contextual diversity. All 1st-pass and later reading times were significantly longer for words with lower contextual diversity compared to words with higher contextual diversity when controlling for word-frequency and other important lexical properties. Furthermore, there was no difference in reading times for higher frequency and lower frequency words when controlling for contextual diversity. The results confirm prior findings regarding contextual diversity and word-frequency effects and demonstrate that contextual diversity is a more accurate predictor of word processing speed than word-frequency within a normal reading task.

Keywords: word-frequency, contextual diversity, eye movements, reading

Words that are encountered more frequently are processed faster than their less frequent counterparts (Forster & Chambers, 1973; Inhoff & Rayner, 1986). As such, the word-frequency effect is generally considered a benchmark in psycholinguistic research. Word-frequency has been reliably shown to modulate the ease of lexical processing across a broad range of tasks and is a critical component in implementations of models of visual-word recognition such as the Interactive Activation model (McClelland & Rumelhart, 1981), the Dual Route Cascaded model (Coltheart, Rastle, Perry, Ziegler, & Langdon, 2001), the Multiple Read-Out model (Grainger & Jacobs, 1996), the Bayesian Reader model (Norris, 2006), and the Spatial Coding model (Davis, 2010), as well as models of eye movement control in reading such as the E-Z Reader model (Reichle, Pollatsek, Fisher, & Rayner, 1998) and SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005). In E-Z Reader, the frequency and predictability of the fixated word influences the duration of both the first and second stage of lexical access. In SWIFT, word-frequency is used to instantiate word difficulty, although word processing speed is determined by a combination of linguistic and oculomotor factors.

Word-frequency effects have an intuitive theoretical explanation: The repetition of exposure inherent in more frequent words corresponds to "stronger" or more accessible internal representations. The role of word recognition in the course of sentence processing is obvious, as words must be identified before they are integrated into more complex meanings conveyed by sentences. Interestingly, there is also considerable evidence suggesting that preceding linguistic context routinely influences word processing. The construction of syntactically and semantically complex representations in the course of sentence comprehension modulates lexical processing difficulty. In a variety of studies, the impact of contextual information on word recognition was apparent in the earliest stages of processing, as indicated by shorter fixation durations and higher first-pass skipping rates for highly predictable words (Balota, Pollatsek, & Rayner, 1985; Drieghe, Rayner, & Pollatsek, 2005; Ehrlich & Rayner, 1981; Kliegl, Grabner, Rolf, & Engbert, 2004; Rayner & Well, 1996). It has also been demonstrated that context has clear effects on the ease of integrating identified words into a sentence discourse. In several studies, word plausibility, as determined by sentence context, exerted a signifi-

This article was published Online First August 12, 2013.

Patrick Plummer, Department of Psychology, University of California, San Diego; Manuel Perea, Departamento de Metodología, Universitat de València, Valencia, Spain; Keith Rayner, Department of Psychology, University of California, San Diego.

This work was supported by National Institutes of Health Grant HD065829 and by a supplement to that grant supporting Patrick Plummer, with additional support by Spanish Ministry of Economy and Competitiveness Grant PSI2011-26924. We thank Harald Baayen, Marc Brysbaert, and Paul Hoffman for their helpful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Patrick Plummer, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093. E-mail: pplummer@ucsd.edu

cant influence on eye-movement measures that reflect early and later stages of word processing (see Rayner, Warren, Juhasz, & Liversedge, 2004; Warren & McConnell, 2007).

The strong effect of word-frequency notwithstanding, recent research has established measures based at the level of context that account for lexical characteristics potentially confounded with word-frequency. Many of these variables have been successfully utilized as predictors of response latencies in isolated word processing tasks by cataloging and/or contrasting the contexts in which a word can be found (Adelman & Brown, 2008; Adelman, Brown, & Quesada, 2006; Hoffman, Lambon Ralph, & Rogers, 2012; Hoffman, Rogers, & Lambon Ralph, 2011; Johns, Gruenenfelder, Pisoni, & Jones, 2012; Jones, Johns, & Recchia, 2012; Steyvers & Malmberg, 2003; Yap, Tan, Pexman, & Hargreaves, 2011) or formally computing information content of a word (relative entropy) using specific probabilistic information (Baayen, 2010; McDonald & Shillcock, 2001). Despite considerable conceptual and operational differences, each measure captures the apparent relationship between contextual information and word recognition. Proposing that semantic context is more important than frequency assumes that the better predictor of word processing difficulty is derived by counting the total number of contexts in which a word appears rather than the mere number of occurrences for the word. This measure, *contextual diversity* (a.k.a., dispersion), was examined by Adelman et al. (2006). Across several corpora, Adelman et al. demonstrated that lexical decision and naming latencies were more accurately predicted using contextual diversity when compared to word-frequency: In the two tasks; words that appeared in more passages were responded to more quickly. Furthermore, more variability in response times was accounted for by contextual diversity than word-frequency, and when both were included in regression analyses, word-frequency showed little to no effect on response times (see also Perea, Soares, & Comesaña, 2013, for a similar finding with developing readers).

Any demonstrated superiority of contextual diversity over word-frequency is impressive given that the two factors are highly correlated: Words that occur more often have a higher likelihood of occurring across many contexts. Despite this fact, as indicated above, contextual diversity has been shown to have a clear predictive advantage over word-frequency in studies with isolated word identification tasks. One important step moving forward from the previous studies is an investigation of the relationship between these two variables in a more normal reading situation. Generalizing from isolated word recognition tasks to normal reading is not straightforward, and it is reasonable to assume that a variable such as contextual diversity might behave differently in reading compared to when employed as predictors in isolated word processing tasks.

The Current Study

Our aim in the present study was to contribute to existing research by contrasting contextual diversity and word-frequency in reading. To the extent that previous findings generalize to silent reading, contrasts between the two variables should correspond to differences in their accuracy as predictors of eye movement behavior during reading. Specifically, in the current experiment we compared three independent groups of words that contrasted in word-frequency and contextual diversity. The contrast group was composed of words that occurred commonly across printed English documents, and they were distributed across a below average number of documents or distinct samples of text (i.e., contexts). The two comparison groups were established using the same corpus materials. The first comparison group was composed of words with a similar word-frequency as those in the contrast group but with higher values in contextual diversity [higher CD group]—this enabled us to assess the effects of contextual diversity. The second comparison group was composed of words with a similar contextual diversity as the contrast group, but with lower values in word-frequency [lower WF group]—this enabled us to assess the effects of word-frequency. By ensuring that the three conditions were closely matched with regard to other important lexical and linguistic characteristics, the influence of the two critical variables can be evaluated in a reading context. Thus, the design takes advantage of the small population of words that occur somewhat frequently but in relatively few contexts. Importantly, the design also compensates for the prohibitive dearth of experimentally viable words that are relatively high in contextual diversity and relatively low in word-frequency.

In sum, conducting an assessment of contextual diversity and word-frequency during reading has the potential to help clarify how the two constructs correspond to representational characteristics of lexical items in the cognitive system. In addition to guiding the development of theoretical accounts of lexical representation, the findings might also inform the development of computational models of word retrieval and sentence processing. In particular, there are two basic contrasts of interest in the design. First, finding that reading times for words in the higher CD group condition are significantly different than the reading times in the contrast group condition would indicate that contextual diversity is a factor that affects normal reading, not only isolated word recognition. Second, if reading times on words in the contrast group condition significantly differ from than reading times in the lower WF group condition, it would further confirm that word-frequency is an accurate measure of word processing difficulty during reading—even when contextual diversity is controlled. Importantly, if the first comparison is positive, whereas the second is negative (as Adelman et al., 2006, demonstrated in lexical decision and naming tasks), this would demonstrate that contextual diversity is a superior metric for word processing difficulty during reading when compared to word-frequency.

Method

Subjects

Forty-eight undergraduates at the University of California, San Diego (UCSD), participated in the experiment for course credit. All were native English speakers with normal or corrected-to-normal vision.

Stimuli

Critical words were obtained from the SUBTLEX (US) Corpus (Brysbaert & New, 2009) via the English Lexicon Project (Balota et al., 2007). The SUBTLEX (US) Corpus provides measures of word-frequency based on the $\log_{10}$ transformed raw number of occurrences (i.e., from the total 51 million words) and contextual diversity measures based on the $\log_{10}$ transformed number of films in which the word appears (i.e., out of the total 8,388 films). Prior to $\log_{10}$ transformation of word-frequency and contextual diversity, one token count was added to each word's raw count, that is, $\log_{10}$(frequency count + 1). Fifty-four items were selected (18 per experimental condition) relative to the average log-transformed word-frequency and contextual diversity values in the corpus (see Table 1). The average word-frequency in the SUBTLEX (US) Corpus was 1.66 (approximately 0.88 occurrences per million words), whereas average contextual diversity was 1.54 (occurring in approximately 0.41% of total films). Critical items ranged from four to seven letters in length. Three groups of words were created: a contrast group and two comparison groups. One of the comparison groups, the *higher CD group*, was composed of words with a similar word-frequency as the contrast group [$t < 0.3$] but with higher contextual diversity [$t(34) = 9.2$, $p < .01$]—this enabled us to examine the effect of contextual diversity with respect to the contrast group. The other comparison group, the *lower WF group*, was composed of words with lower word-frequency than the contrast group [$t(34) = 11.0$, $p < .01$] but with similar contextual diversity [$t < 0.2$]—this enabled us to examine the effect of word-frequency. It is important to note that the average word-frequency in the three groups is considerably lower than what is conventionally considered high frequency in psycholinguistic research—there is no such convention for contextual diversity, and corpus wide averages were employed in order to establish the groups. Obviously, there are vastly more low frequency words than high frequency words, and the lower range of the frequency distribution offers many more words that afford opportunities to contrast word-frequency and contextual diversity. In this light, a recent analysis conducted by Keuleers, Diependaele, and Brysbaert (2010) found across English, French, and Dutch items that word-frequency exerts the strongest effect between 0.1 and 50 occurrences per million. Table 2 presents average word lengths, in addition to average orthographic, phonological, and phonographic neighborhood sizes, corresponding neighborhood frequencies, letter bigram frequencies, and number of phonemes; none of which showed differences across the three conditions [all ts < 1.66].

Each word was embedded into a unique sentence frame (see the Appendix). Subjective judgments for the sentences were recorded from a separate group of UCSD undergraduates ($n = 10$) who rated each sentence on grammaticality and sensibility using a 1–7 scale, with 1 being *very poorly written* and 7 being *very well written*. The ratings for sentences in the contrast group ($M = 5.57$, $SD = 0.5$) were not statistically different from ratings in the lower WF group ($M = 5.30$, $SD = 0.6$) or the higher CD group ($M = 5.62$, $SD = 0.5$) [both ts < 1.4]. Word predictability measurements for each item were obtained from another group of UCSD undergraduates ($n = 15$) using a cloze task in which they responded to an incomplete sentence fragment by continuing the sequence with their best guess for the next upcoming word. Overall, subjects responded with the critical words in less than 1% of the trials with no significant differences in the proportion of stimulus word responses between group conditions [all ts < 0.6].

Table 2
Mean Lexical Characteristics for Target Words in Each Group

Characteristic	Higher CD group	Lower WF group	Contrast group
Word length	5.72 (1.07)	5.72 (1.07)	5.76 (1.15)
Ortho_N	1.83 (2.28)	1.61 (1.94)	2.00 (2.76)
Phono_N	2.89 (4.55)	6.50 (10.83)	6.35 (7.64)
OG_N	1.06 (1.63)	1.00 (1.64)	1.00 (1.66)
Freq_N	7.02 (1.25)	6.96 (1.43)	7.11 (2.07)
Freq_P_N	7.43 (1.18)	7.20 (1.13)	6.87 (1.64)
Freq_OG_N	6.97 (1.29)	6.73 (1.31)	7.23 (2.66)
BG_Freq	1946 (1011)	1758 (815)	1800 (572)

Note. All standard deviations are provided in in parentheses. CD = contextual diversity; WF = word-frequency; Word length = number of characters; Ortho_N = orthographic neighborhood size; Phono_N = phonological neighborhood size; OG_N = phonographic neighborhood size; Freq_N = average word frequency of orthographic neighborhood; Freq_P_N = average word frequency of phonological neighborhood; Freq_OG_N = average word frequency of phonographic neighborhood; BG_Freq = average letter bigram frequency.

Procedure

Eye movements were recorded using an SR EyeLink 1000 eye-tracker (1000-Hz sampling rate). Sentences were presented 60 cm from the subject on a 20-in. (50.8-cm) HP CRT p1230 monitor with a display resolution of 1024 × 768 using 14-point Courier New font (2.4 characters per degree of visual angle). Head movements were minimized via a head-rest. After initial calibration, subjects silently read each sentence for comprehension and answered a subsequent true–false question using a handheld trigger. Average comprehension accuracy was 93% ($SD = 4\%$; range = 83%–100% correct). Experimental sentences were presented in randomized order intermixed with 42 unrelated filler sentences and five initial practice trials. All 101 sentences were presented from the left-center of the display screen after the eyes focused on a fixation point in that location between each trial. At this time, the validity of the calibration was checked by the experimenter and recalibrated when necessary.

Table 1
Mean Target Word-Frequency (WF) and Contextual Diversity (CD) for Each Group

Group	WF (Log)		WF (Standard)		CD (Log)		CD (Standard)	
	M (*SD*)	Range	*M* (*SD*)	Range	*M* (*SD*)	Range	*M* (*SD*)	Range
Higher CD group	1.87 (0.14)	1.71–2.16	1.50 (0.55)	0.96–2.82	1.73 (0.14)	1.58–2.04	0.67% (0.24)	0.44–1.3
Lower WF group	1.27 (0.17)	0.95–1.57	0.37 (0.16)	0.16–0.71	1.19 (0.14)	0.95–1.38	0.18% (0.06)	0.05–0.32
Contrast group	1.86 (0.15)	1.63–2.15	1.49 (0.54)	0.82–2.73	1.20 (0.20)	0.70–1.45	0.19% (0.08)	0.10–0.27

Note. Log_{10} transformed WF (based on raw frequency count out of 51 million total words) and CD averages (based on film occurrence count out of 8,388 total films) are provided with standard deviations in parentheses. Conventional WF (per million) and CD (percentage) measures are provided for reference with standard deviations in parentheses.

Analysis

Reading times on critical items from each of the three conditions were analyzed as target words in their respective sentence frames. All fixation durations greater than 699 ms or less than 50 ms were removed prior to analysis[1]—a restriction that affected 1.2% of all fixations. For analysis on the target word, all trials in which tracking of the right eye was lost before, during, or after any fixation on the target word were removed prior to analysis; consequently, 7.2% of the trials were excluded from analysis (186 of 2,592 trials). Results from seven standard eye-movement measures (Rayner, 1998, 2009) are reported. First-fixation duration (FFD) is the duration of the initial fixation on the target word during first-pass reading. Single-fixation duration (SFD) is the duration for trials wherein the target word received exactly one fixation during first-pass reading. Gaze duration (GZD) includes all first-pass fixations on the target word during first-pass reading including the initial fixation and any refixations on the target until the eyes move off of the word in either direction. Go-past time (a.k.a., regression path duration) includes all first-pass fixation durations on the target word in addition to all fixations subsequent to regressions into prior text before the eyes move beyond the right-boundary of the target word. Skipping rate (Skip) is the proportion of trials in which the target word does not receive a direct fixation during first-pass reading. Regression rate (RegIn) is the proportion of trials in which the target word receives one or more fixations as the result of regressive eye-movements landing on the target word region. Total reading time (TTime) includes all first-pass fixation durations and any subsequent fixation durations on the target word for the entire trial period. Planned comparisons were conducted using *t* tests across subjects ($t1$) and items ($t2$).[2]

Results

All reading measures are provided in Table 3.

Effect of Contextual Diversity (Contrast Group vs. Higher CD Group)

Target words with lower contextual diversity were fixated longer than words with higher contextual diversity. This effect was observed for all first-pass measures: FFD [$t_1(47) = 2.60$, $p < .05$; $t_2(34) = 2.75$, $p < .05$], SFD [$t_1(47) = 3.18$, $p < .01$; $t_2(34) = 2.89$, $p < .01$], and GZD [$t_1(47) = 3.86$, $p < .01$; $t_2(34) = 2.78$, $p < .01$]. Target words with lower contextual diversity were skipped less often than words with higher contextual diversity during first-pass reading; however, this effect was only marginal [$t_1(47) = 1.93$, $p = .06$; $t_2(34) = 1.65$, $p = .11$].

There were also significant effects of contextual diversity for go-past and total reading times. Go-past times were longer for words with lower contextual diversity than for words with higher contextual diversity [$t_1(47) = 3.57$, $p < .01$; $t_2(34) = 2.61$, $p < .05$]. Likewise, total reading times for target words with lower contextual diversity were longer than those with higher contextual diversity [$t_1(47) = 5.20$, $p < .01$; $t_2(34) = 2.89$, $p < .01$]. In addition, target words with lower contextual diversity received a higher proportion of regressive fixations than those with higher contextual diversity [$t_1(47) = 3.46$, $p < .01$; $t_2(34) = 2.39$, $p < .05$].

Table 3
Mean Reading Measures on Target Words Across Each Group

Measure	Higher CD group	Lower WF group	Contrast group
FFD	242 (31)	263 (35)	260 (37)
SFD	245 (34)	273 (41)	270 (40)
GZD	281 (42)	313 (56)	320 (57)
Skip	0.10 (0.08)	0.08 (0.07)	0.06 (0.07)
GoPast	331 (66)	389 (92)	380 (89)
RegIn	0.18 (0.12)	0.27 (0.14)	0.30 (0.17)
TTime	335 (69)	415 (100)	427 (101)

Note. All reading times on target word are presented in milliseconds. All standard deviations are provided in in parentheses. CD = contextual diversity; WF = word-frequency; FFD = first-fixation duration; SFD = single-fixation duration; GZD = gaze duration; Skip = skipping rates (presented as the proportions of trials in which the target word did not receive a first-pass fixation across subjects); GoPast = go-past time; RegIn = regressions in (presented as the proportion of regressions made into the target word across subjects); TTime = total reading time.

Effects of Word-Frequency (Contrast Group vs. Lower WF Group)

Critically, higher frequency target words did not significantly differ from lower frequency words in any first-pass measures for subjects or items [all *t*s < 1.0]. Furthermore, there were no significant differences across word-frequencies for go-past time, proportion of regressions in, or total reading time [all *t*s < 0.9]. This pattern of effects provides converging evidence for the absence of a reliable word-frequency effect when contextual diversity is controlled in reading.[3]

Supplemental Analysis

The present results demonstrate that contextual diversity is a superior predictor of word processing difficulty than word-frequency in reading. As a further control, we collected the average lexical decision and word naming latencies for words in each condition from the English Lexicon Project. Average lexical decision latencies for words in the contrast group ($M = 737$ ms, $SD = 55$) were slower than latencies for words in the higher CD group ($M = 672$ ms, $SD = 43$) [$t(34) = 3.93$, $p < .01$] (i.e., a significant effect of contextual diversity), whereas latencies for words in the lower WF group and the contrast group ($M = 742$ ms, $SD = 61$) did not significantly differ [$t(34) < 0.3$] (i.e., no effect of word-frequency). In the naming task, and to control for the influence of outliers, words with average naming latencies above

[1] The results remained very stable over various cutoff procedures used involving minimum and maximum cutoff points.

[2] In addition to analysis of the designed contrasts reported here, linear mixed effects regressions were performed on the data using the *lme4* package in *R* statistical computing program. The regression model included fixed effects of the log-transformed word-frequency of both the target word and the word that immediately preceded the target, the log-transformed contextual diversity of the target word, and the length of the target word (all as continuous predictors), in addition to random intercepts for items and subjects as well as random subject slopes for the effects of target word-frequency and contextual diversity. This model yielded the same pattern of results as the reported planned group comparisons (i.e., contextual diversity effects in the absence of word-frequency effects).

[3] Contrasts of the higher CD and lower WF groups revealed significant differences on all reading measures with the exception of skipping rate.

950 ms were removed (one item from the contrast group and one from the lower WF group) prior to analysis. Naming latencies for words in the contrast group (M = 678 ms, SD = 52) were marginally longer than for words in the higher CD group (M = 643 ms, SD = 58) [$t(33)$ = 1.92, p = .064], whereas latencies for words in the contrast group did not differ from words in the lower WF group (M = 720 ms, SD = 97) [$t(32)$ = 1.55, p = .13]. Thus, the lexical decision and word naming averages across conditions reflect the pattern observed in Adelman et al.'s (2006) corpus analysis.

The measurement of contextual diversity has some shortcomings; specifically, it does not fully take into account the relative frequency of a given context or the semantic similarity between contexts in which a word appears. Certainly, some contexts are more frequently encountered than others. It is also undoubtedly the case that contextual diversity has substantial sensitivity to the frequency of a given context. Frequent contexts should correspond to a high number of passages that would, presumably, contain relatively high word co-occurrence rates. However, words occurring in only a few, highly frequent, contexts will have indistinguishable contextual diversities when compared to words distributed across a large variety of semantic contexts. Additionally, there are many instances where specific word meanings or the sense in which a word is used vary across contexts.

A recently established variable, *semantic diversity*, accounts for many of these, potentially crucial, lexical characteristics by measuring the semantic variability across contexts for a given word (Hoffman et al., 2012, 2011). Semantic similarity for passages is measured by overall word-by-word co-occurrence rates utilizing document-to-document comparison in Latent Semantic Analysis (LSA; Landauer & Dumais, 1997).[4] Semantic diversity captures the similarity across contexts (i.e., documents in a corpus) in which a word appears. Words that occur in a variety of semantically distinct documents have relatively high semantic diversity values; words with lower semantic diversity values occur in fewer contexts. Experimental results found that semantic diversity was a reliable predictor for word processing difficulty in healthy subjects as well as populations with semantic aphasia (Hoffman et al., 2012, 2011).

Thus, semantic diversity values were obtained for 40 of the 54 target words. These semantic diversity values ranged from 1.95 to 0.74 with an average of 1.38 (SD = 0.3). For these 40 items, semantic diversity was significantly correlated with log-transformed contextual diversity (r = –.251, $p < .01$) and word-frequency (r = –.498, $p < .01$), as well as with word length (r = –.264, $p < .01$). Partial correlations were computed in order to examine the effect of semantic diversity while controlling for these potential confounds. Correlation coefficient estimates for semantic diversity after controlling for the effects of contextual diversity, word-frequency, and word length can be found in Table 4. For words in the current study, semantic diversity had no statistically significant relationship with any of the first-pass measures. Crucially, semantic diversity did have a significant effect on the two later reading measures: go-past duration and total reading time. For reading measures in which semantic diversity exerted a significant, independent influence, the effect is such that items of higher semantic diversity are associated with inflated reading times (i.e., a positive correlation between semantic diversity and reading measures). Additionally, the corresponding partial correlations were computed for contextual diversity and word-frequency for these 40 items using the same residualization technique (see Table 4). Partial correlations generally confirmed the results of the planned group contrasts. Contextual diversity was a significant predictor of all first-pass and later reading times even when statistically controlling for word-frequency, semantic diversity, and word length; in contrast, when controlling for other important lexical properties, word-frequency did not have an influence on reading times (with the exception of total reading time). Interestingly, in this analysis, word-frequency demonstrates a positive correlation with total reading time indicating an inhibitory effect similar to the effects observed in Adelman et al.'s (2006) corpus analysis. Collectively, these data offer some evidence that semantic diversity has effects on later eye-movement measures, but not on first-pass reading measures. Crucially, even when accounting for the effects of other important lexical characteristics, contextual diversity exerts a significant influence on eye-movement behavior during reading.

Discussion

The present experiment extends prior findings and indicates that contextual diversity has a strong influence on word recognition in reading. Furthermore, the findings demonstrate that, at least for the range of low-frequency words employed in the experiment, sheer frequency of occurrence has little predictive value for reading times when accounting for a variable that is sensitive to the distribution of words across different contexts (for similar findings using word identification tasks, see Adelman et al., 2006; Jones et al., 2012). At least two plausible alternative interpretations appear warranted by the data. First, the results may indicate that contextual diversity is simply a more accurate or realistic measure of the frequency of word occurrence in the mental lexicon—in a similar way as subtitle word-frequency is a better predictor of the ease of lexical access than printed word-frequency because it may better reflect everyday language (Brysbaert & New, 2009). Idiosyncrasies in word-frequency measurements could lead to considerable deviations from the subjective number of exposures to a word across individuals. The change in sampling granularity from word token counts to document counts could potentially avoid an appreciable amount of this measurement error.

The second account has more noteworthy theoretical implications relating to representations in the mental lexicon and intrinsic lexical properties. The relative effects of contextual diversity and word-frequency suggest that lexical access is mediated by the number of contexts in which a word tends to occur rather than pure repetition of occurrence. The overall predictive supremacy of contextual diversity over word-frequency in eye-movement measures indexing the earliest processing stages provides compelling evidence that, when accounting for confounded word characteristics, mere frequency of occurrence has little effect on word recognition during reading. Further support is offered by the fact that the contextual diversity effect replicates across a variety of word processing tasks beyond normal reading. The empirical relationship of contextual diversity and word-frequency offers evidence of

[4] Previous research has used latent semantic analysis as a reliable predictor of eye movement behavior during first-pass reading (Pynte, New, & Kennedy, 2008; Wang, Pomplun, Chen, Ko, & Rayner, 2010).

Table 4
Partial Correlations of Predictors and Reading Times

Predictor	SemD After Log-CD, Log-WF, and Length	Log-CD After SemD, Log-WF, and Length	Log-WF After SemD, Log-CD, and Length
FFD	−0.0003	−0.062*	−0.007
SFD	0.025	−0.093**	0.007
GZD	0.010	−0.074**	0.016
GoPast	0.065**	−0.073**	0.023
TTime	0.140**	−0.148**	0.062**

Note. Estimates are partial correlation coefficients between reading times and continuous predictors of semantic diversity (SemD), log-transformed contextual diversity (Log-CD), and log-transformed word-frequency (Log-WF) after partialing out other relevant predictors computed using Pearson correlations. First-fixation duration (FFD), gaze duration (GZD), and go-past duration (GoPast) correlations are each based on 1,636 observations; total reading time (TTime) correlations are based on 1,770 observations; and single-fixation duration (SFD) correlations are based on 1,297 observations.
* $p < .05$. ** $p < .01$.

stable contextual information in addition to other semantic features associated with lexical representations. The findings lend themselves to the assumption that supralexical information is a relevant level of lexical organization. Specifically, representing information about contexts in which a given word should appear based on prior experience with the word's distribution across contexts. Moreover, these long-term memory resources must be considered stable to the extent that they affect lexical processing in a specific linguistic context. This relationship also coerces examination and possible reinterpretation of the roles orthographic and phonological encoding play in lexical access, as well as the relationship between sublexical characteristics and the well-established word-frequency effect.

In addition to representational-level implications, the contrast of word-frequency and contextual diversity may indicate systematic differences across the intrinsic properties of the experimental items. More specifically, the critical population of words having above average frequency of occurrence yet only occurring in a relatively small number of contexts may differ with regard to information content when compared to words that do not instantiate this contrast. Such words might tend to carry more information in general. Alternatively, these words might tend to carry particular forms of information that lead to words being focal points of processing difficulty. Taking the observed effects in these data to a theoretical conclusion suggests that indexing the number of discourses in which a word form occurs is more important than capturing how often the specific word form is encountered for the task of predicting processing difficulty in normal language contexts.

By either of the two broad interpretations, theoretical accounts and computationally explicit models of lexical access that modulate word processing speed and difficulty as a function of repeated exposure to word form should accommodate the effect of contextual diversity (see Jones et al., 2012, for a word-learning model that includes contextual diversity; see also Steyvers & Malmberg, 2003). Computational models of eye-movement control during reading typically use word-frequency as a surrogate for the speed of lexical access rather than instantiating explicit word recognition modules. As such, these models could easily substitute word-frequency with contextual diversity without any serious theoretical implications.

The current experimental results also suggest that semantic diversity contributes to the ease with which a word is integrated into the preceding context. During sentence comprehension, identified words must incrementally update a corresponding representation of the message being conveyed. In reading research, eye-movement measures such as go-past times and regressions into a word are generally associated with the construction of semantic and syntactic interpretations. Semantic diversity formally expresses the variety of contexts in which a word appears and will, in general, positively correlate with contextual diversity; nonetheless, it does not represent any information regarding the relative frequency of distinct semantic contexts and thus does not directly encode dispersion or occurrence information. In fact, there is a negative correlation between contextual diversity and semantic diversity for items in the current experiment. The characteristics of semantic diversity fit intuitively with the observed significant effect on later measures, which correspond to integration, as well as the absence of first-pass effects, which correspond to meaning retrieval. Inasmuch as words with higher semantic diversity generally occur in a variety of semantically distinct contexts, ambiguity associated with selecting the appropriate semantic interpretation for words (given their specific context) may generate difficulty during postlexical integration. This would suggest that influences owed to the distinctiveness or similarity across contexts typically containing a word may be less pronounced during earlier lexical processing stages. It may also be the case that once facilitatory influences of occurring across higher numbers of distinct contexts is accounted for (ostensibly by contextual diversity), only an inhibitory influence of increased semantic variability across the distinct contexts remains to be accounted for (ostensibly by semantic diversity). However, any interpretation of semantic diversity effects in the current experiment must be looked upon with caution. While words in this study had a variable range of semantic diversity values, the distribution of semantic diversity in the experimental stimuli was less well controlled than that of word-frequency or contextual diversity. Additionally, the effects of semantic diversity might be affected by the current design wherein sentences containing the experimental items were intended to be neutral (i.e., semantically unconstrained contexts). Moreover, in these data there was no direct control of word co-occurrence based semantic properties.[5] In order to license any strong inferences regarding semantic diversity effects, more in depth consideration and quantification of the extent to which particular linguistic contexts share semantic similarity with the typical contexts containing experimental items would be necessary. Certainly, a more direct and carefully controlled experimental manipulation of semantic diversity is required to clarify the nature of its influence on lexical access and sentence processing.

In sum, the current experiment has clearly shown that the number of passages in which a particular word appears has a stronger influence than sheer frequency of occurrence on eye-movement behavior during reading. While the current experiment contributes to assessing contextual diversity and word-frequency effects, there are number of fruitful avenues for future research on the topic. Classification of the items in terms of word-frequency was based on corpus averages, the group of higher frequency words employed here is considerably lower in occurrence rate than words conventionally classified as high frequency in the psycholinguistic literature. Further research is necessary to examine whether the observed contextual diversity effects will differ in items across the full range of the word-frequency distribution. More or less response variance might be accounted for at the extremes of the frequency index. Likewise, the effect of contextual diversity may behave differently for the relatively small population of words that are extremely high in frequency or the vast population of words that are extremely low in frequency. Examining the independent effects and relative predictive power of contextual diversity in words typically classified as high frequency in psycholinguistic research is certainly another crucial inquiry for future empirical studies. Furthermore, it would also be valuable to investigate how these effects are modulated by changes in within-sentence contextual constraint. Thus, the present experiment validates endeavors to develop and refine measures of a word's information content and distributional characteristics across distinct contexts during reading.

[5] In a post hoc analysis, indices of co-occurrence based semantic similarity between target words and their corresponding sentence frames—including all words in the sentence with the exception of the target word itself—were obtained from LSA using the default 300 dimension semantic space (available at http://lsa.colorado.edu/). Based on a median split such that 20 items had semantic diversity values below the median of 1.39 ($M = 1.15$, $SD = 0.19$), and 20 items had semantic diversity values above the median ($M = 1.6$, $SD = 0.19$). The contrast revealed that for target words lower in semantic diversity, the average sematic similarity to their respective sentence frames ($M = 0.21$, $SD = 0.12$) did not significantly differ from the average semantic similarity of higher semantic diversity words and their respective sentence frames ($M = 0.19$, $SD = 0.1$) [$t(37) < 0.6$].

References

Adelman, J. S., & Brown, G. D. A. (2008). Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review, 115,* 214–227. doi:10.1037/0033-295X.115.1.214

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17,* 814–823. doi:10.1111/j.1467-9280.2006.01787.x

Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon, 5,* 436–461. doi:10.1075/ml.5.3.10baa

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17,* 364–390. doi:10.1016/0010-0285(85)90013-1

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445–459. doi:10.3758/BF03193014

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41,* 977–990. doi:10.3758/BRM.41.4.977

Coltheart, M., Rastle, K., Perry, C., Ziegler, J., & Langdon, R. (2001). DRC: A dual-route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256. doi:10.1037/0033-295X.108.1.204

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review, 117,* 713–758. doi:10.1037/a0019738

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 954–969. doi:10.1037/0096-1523.31.5.954

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

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777–813. doi:10.1037/0033-295X.112.4.777

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12,* 627–635. doi:10.1016/S0022-5371(73)80042-8

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518–565. doi:10.1037/0033-295X.103.3.518

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2012, December 13). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*. Advance online publication. doi:10.3758/s13428-012-0278-x

Hoffman, P., Rogers, T. T., & Lambon Ralph, M. A. (2011). Semantic diversity accounts for the "missing" word frequency effect in stroke aphasia: Insights using a novel method to quantify contextual variability in meaning. *Journal of Cognitive Neuroscience, 23,* 2432–2446. doi:10.1162/jocn.2011.21614

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40,* 431–439. doi:10.3758/BF03208203

Johns, B. T., Gruenenfelder, T. M., Pisoni, D. B., & Jones, M. N. (2012). Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *The Journal of the Acoustical Society of America, 132,* EL74–EL80. doi:10.1121/1.4731641

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 66,* 115–124. doi:10.1037/a0026727

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). 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 Psychology, 1,* 174. doi:10.3389/fpsyg.2010.00174

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16,* 262–284. doi:10.1080/09541440340000213

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240. doi:10.1037/0033-295X.104.2.211

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407. doi:10.1037/0033-295X.88.5.375

McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech, 44,* 295–322. doi:10.1177/00238309010440030101

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113,* 327–357. doi:10.1037/0033-295X.113.2.327

Perea, M., Soares, A. P., & Comesaña, M. (2013). Contextual diversity is a main determinant of word-identification times in young readers. *Journal of Experimental Child Psychology, 116,* 37–44. doi:10.1016/j.jecp.2012.10.014

Pynte, J., New, B., & Kennedy, A. (2008). A multiple regression analysis of syntactic and semantic influences in reading normal text. *Journal of Eye Movement Research, 2,* 1–11.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62,* 1457–1506. doi:10.1080/17470210902816461

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1290–1301. doi:10.1037/0278-7393.30.6.1290

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

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105,* 125–157. doi:10.1037/0033-295X.105.1.125

Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 760–766. doi:10.1037/0278-7393.29.5.760

Wang, H.-C., Pomplun, M., Chen, M., Ko, H., & Rayner, K. (2010). Estimating the effect of word predictability on eye movements in Chinese reading using latent semantic analysis and transitional probability. *The Quarterly Journal of Experimental Psychology, 63,* 1374–1386. doi:10.1080/17470210903380814

Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review, 14,* 770–775. doi:10.3758/BF03196835

Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review, 18,* 742–750. doi:10.3758/s13423-011-0092-y

Appendix

Stimuli Sentences With Target Words Underlined

Lower Word-Frequency (WF) Group

Kara’s answer on the math quiz was the exact inverse of the correct answer.

The natives of the remote island always adorn newcomers with flowers and jewels.

The newly engineered life form was placed in complete stasis by the researchers.

When taking attendance the substitute was sure not to omit reading anyone’s name.

Last night Tim slept on an uncomfortable pallet that was made of blankets and straw.

Harold tried his hardest to quickly stifle his laughter during the business meeting.

Before cooking it, Harrison had to impale the meat onto a fork.

While hiking, Jackson found many aloe plants along the trails.

Todd noticed that the dark tint of each building’s windows prevented reflections.

Last summer, Terry became a dedicated vegan after learning of the health benefits.

Many kindergarteners tend to believe in sorcery and other imaginary things.

While taking his exam, Mike tried to feign illness when he realized he would fail.

Over the summer, Kara watched the hilarious parody video about high school.

Karen never understood her father’s strict myopic viewpoints about social issues.

For the presentation Henry would need to carefully collate his detailed notes.

Chris’ behavior caused family and friends to spurn him since they felt betrayed.

While shopping Susan bought the necessary utensil for the dinner party he was hosting.

The encoded message was mostly jumbled letters without any meaning.

Higher Contextual Diversity (CD) Group

During the day Melanie works for an export agency that manages product shipping.

Michael runs daily to strengthen his flabby torso and get into better shape.

When Ted reached the very steep hill he was discouraged by the thought of climbing it.

The delivery truck that carried timber ran into traffic on the way to the factory.

Tania and her friends found three wasp hives hidden in a tree and carefully avoided them.

Stewart traveled to see his oldest sibling for the first time in ten years.

The explorer was searching for the dangerous reptile species in the jungle.

Before going to bed, Elizabeth began to shiver with nervousness thinking about the task.

(*Appendix continues*)

Last weekend Jackie dyed her hair a light almond color to match her skin tone.

On Tuesday, Sarah showed her friends her splits and they were very impressed.

Barry is required to visit several clinics each week to monitor his patients' health.

While walking to class Tommy saw a large perch crowded with seagulls.

Ellen's new jacket was very snug and looked to be a little small for her.

While at school Billy began to spasm uncontrollably and was sent to the nurse.

Mary and Ben are looking for cheap rugs to put in their living room.

Yesterday Tom had to check the enormous boiler and make sure it was running properly.

Walking into the restaurant, Beth saw the unclean plate lying on the table.

After a tiring day, Sara treated herself to a fresh avocado and chicken pizza for dinner.

Contrast Group

In class, Howard learned the role of the neutron in an atom's structure and function.

Sam and Rick went to see the mock joust held at the resort.

Steve spent hours leading the energetic horde of students through campus.

Janet was fascinated by the ubiquitous kelp being washed onto the shoreline.

Holly was approached by a lost beagle that seemed to be searching for its owner.

After many years of service the two most senior porters were fired abruptly.

While taking a stroll Tony saw a rare condor flying around looking for food.

After construction was completed, the voyager was finally ready for its trip into space.

The strange animals Harry saw were all mutants created in laboratory experiments.

The professor skillfully explained the importance of the watt as a unit of measurement.

Fans were excited to meet the world renowned striker before the game started.

Last Friday, Kevin spent hours on his new flan recipe before the potluck dinner.

Hippos do not run fast because of their stumpy legs and fat bodies.

The hunter kept the sharpened talon of an adult eagle as a trophy.

At the sale, Kristi found an elegant cravat which was being sold for a dollar.

Eric and Steven admired the intricate quill found at the collector shop.

A few nights ago, Kelly found a small piglet abandoned in her backyard.

Last week, Bill bought a custom bobsled since he was interested in trying new things.

Received February 19, 2013
Revision received June 7, 2013
Accepted June 27, 2013 ■