

## Research Article

# Contextual Diversity, Not Word Frequency, Determines Word-Naming and Lexical Decision Times

James S. Adelman, Gordon D.A. Brown, and José F. Quesada

University of Warwick, Coventry, United Kingdom

**ABSTRACT**—*Word frequency is an important predictor of word-naming and lexical decision times. It is, however, confounded with contextual diversity, the number of contexts in which a word has been seen. In a study using a normative, corpus-based measure of contextual diversity, word-frequency effects were eliminated when effects of contextual diversity were taken into account (but not vice versa) across three naming and three lexical decision data sets; the same pattern of results was obtained regardless of which of three corpora was used to derive the frequency and contextual-diversity values. The results are incompatible with existing models of visual word recognition, which attribute frequency effects directly to frequency, and are particularly problematic for accounts in which frequency effects reflect learning. We argue that the results reflect the importance of likely need in memory processes, and that the continuity between reading and memory suggests using principles from memory research to inform theories of reading.*

What determines how quickly a word can be read? Empirically, in both word-naming and lexical decision tasks, frequency of occurrence is among the strongest known influences: Frequent words are read more quickly than infrequent words (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Forster & Chambers, 1973; Frederiksen & Kroll, 1976). Thus, it appears that repeated experience with or exposure to a particular word

makes it more readable or identifiable. A key assumption of theoretical explanations of the word-frequency (WF) effect is that it is due to the number of experiences with a word—that each (and every) exposure to a word has a long-term influence on its accessibility.

In learning-based accounts of reading, such as connectionist models (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Zorzi, Houghton, & Butterworth, 1998), learning occurs upon each experience of a word, strengthening the connections needed to process that word and allowing it to be processed more quickly. In lexicon-based models, the accessibility of individual lexical entries (words) is governed directly by frequency, either by thresholds of activation based on WF (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) or by a serially searched frequency-ranked list (e.g., Murray & Forster, 2004).

Research on memory, however, has found that the extent to which the number of repeated exposures to a particular item affects that item's later retrieval depends on the separation of the exposures in time and context (Glenberg, 1976, 1979). Indeed, under some conditions, if neither time nor context changes substantially, there may be no benefit of repetition at all (Verkoeijen, Rikers, & Schmidt, 2004). If the memory for words that subserves word recognition operates in the same fashion, then the effect of repetitions (i.e., WF) will be diminished or eliminated when these repetitions occur in the same context. Accordingly, the number of contexts in which words are experienced, their contextual diversity (CD), should determine their accessibility and hence response times (RTs) in word naming and lexical decision.

A normative measure of a word's CD may be obtained by counting the number of passages (documents) in a corpus that

Address correspondence to James S. Adelman, Department of Psychology, University of Warwick, Gibbet Hill Rd., Coventry, CV4 7AL, United Kingdom, e-mail: j.s.adelman@warwick.ac.uk.

contain that word; measured this way, CD has been shown to have effects on recognition memory that are distinguishable from WF effects (Steyvers & Malmberg, 2003). In the study reported here, we compared the ability of CD and WF to predict six existing sets of data regarding RTs in word naming and lexical decision, basing our analyses on measures of CD and WF from each of three corpora.

## METHOD

### Dependent Variables

The dependent variables were mean RTs for word naming (reading aloud) and lexical decision (judging whether or not the stimulus is a word) in six data sets. Two of these data sets contain word-naming data for 2,820 uninflected one-syllable words (data for 2,776 words were analyzed in the present study) from studies of young adults (Spieler & Balota, 1997) and older adults (Balota & Spieler, 1998). Another two data sets contain lexical decision data for the same words; again, one set contains data obtained with a young-adult group, and the other contains data obtained with an older-adult group (Balota, Cortese, & Pilotti, 1999). The last two data sets, from the Elexicon project, contain both word-naming and lexical decision data obtained from young adults using a broader selection of 40,481 words (Balota et al., 2000; data for 39,383 words were analyzed in the present study).

### Independent Variables

The independent variables were WF (number of occurrences) and CD (number of passages or documents in which a word occurs), calculated from three corpora. First, we used Kučera and Francis's (1967; KF) counts for the Brown corpus, which contains 500 samples (target length of 2,000 tokens) from distinct documents spread evenly over 15 genres. The samples have a mean length of 2,030 tokens ( $SD = 42$ ). Second, we calculated WF and CD in the 12th-grade portion of the LSA/TASA (Latent Semantic Analysis from Touchstone Applied Science Associates) corpus (Landauer, Foltz, & Laham, 1998), which consists of most of the texts used in the compilation of Zeno, Ivens, Millard, and Duvvuri's (1995) frequency norms; these norms are designed to reflect the likely experience of students schooled in the United States.<sup>1</sup> This section of the corpus has 28,882 samples (from distinct documents) with a mean length of 286 tokens ( $SD = 25$ ). Third, we also compiled counts from the written portion of the British National Corpus (BNC; British National Corpus Consortium, 2000). This corpus is designed to have the largest possible samples—ideally, whole texts. This portion of the BNC contains 3,144

samples, of various forms and lengths, drawn from documents ranging from pamphlets through book chapters to whole issues of newspapers.<sup>2</sup> The mean number of tokens in each passage is 26,892 ( $SD = 25,914$ ).

When calculating logarithm and power-law fits, we increased all counts by 1, to avoid problems from zero counts. Items with zero counts were excluded from the rank analyses.

The following measures from CELEX<sup>3</sup> (Baayen, Piepenbrock, & Gulikers, 1995) were included as covariates in the analyses: word length (number of letters), orthographic neighborhood size, rime consistency (for the monosyllabic databases only), number of syllables (when applicable), and initial phoneme (for word naming only).

## RESULTS

In this section, we show that CD predicts word-processing times independently of WF and, moreover, that there is no evidence for a facilitatory effect of WF independent of CD. We also discuss a number of possible explanations of the results that are inconsistent with our contention that CD per se determines accessibility, and provide evidence for the validity of the measure of CD that we use here.

### Does CD or WF Predict Word-Naming and Lexical Decision Times?

Table 1 presents the results of the 18 analyses (6 data sets  $\times$  3 corpora) using log-transformations of WF and CD; log-WF is generally agreed to approximate a linear predictor of naming and lexical decision RTs. After the covariates were entered into the analysis, introducing either WF or CD accounted for significant additional variance, with high WF and high CD both being associated with faster RTs. Moreover, the improvement in prediction was always greater for CD than for WF. In all 18 analyses, there was a unique effect of CD. Six analyses showed a unique effect of WF; in every case, high WF led to slow RTs, meaning that WF acted as a suppressor variable. These results suggest not only that CD is a better predictor of lexical decision and word-naming times for both young and older participants, but also that WF does not contribute to such RTs, except insofar as it is correlated with CD and the covariates.

Because the addition of CD to the regression equation eliminated the unique effect of WF, CD must be a critical component of a confound that creates an artifactual effect of WF. However, it need not be the only component. When only log-WF and log-CD were entered into the equation, there was always a facilitatory effect of CD, but in some cases there was also a facilitatory effect

<sup>1</sup>This corpus is also described on the Web at <http://lsa.colorado.edu/spaces.html>. We used the 12th-grade level because frequency computed from this level is a better predictor of RTs than frequency counted across the whole corpus, probably because the full corpus is too heavily weighted toward college-level texts to be representative of undergraduate participants or older control subjects education-matched to them.

<sup>2</sup>See also <http://www.natcorp.ox.ac.uk/what/balance.html> for details of the composition of this corpus.

<sup>3</sup>CELEX was not used for frequency counts because the corresponding CD values were not readily obtainable. Additionally, the base corpus consists of only 243 documents, and so would yield a relatively coarse measure of CD.

**TABLE 1**

*Effects of Log-Transformed Word Frequency (Log-WF) and Log-Transformed Contextual Diversity (Log-CD) as Calculated From Each Corpus*

Data set and corpus	Effect ( $\Delta R^2$ in %)				
	Covariates	Log-WF	Log-CD	Log-WF unique	Log-CD unique
SB97: young adults, naming					
KF	45.95***	5.16***	5.35***	0.00	0.19**
TASA	45.95***	6.73***	6.82***	0.00	0.09*
BNC	45.95***	5.85***	6.72***	0.07 <sup>†</sup> (I)	0.94***
BS98: older adults, naming					
KF	26.72***	10.39***	10.90***	0.00 (I)	0.51***
TASA	26.72***	13.73***	13.82***	0.02	0.11*
BNC	26.72***	11.74***	13.29***	0.08 <sup>†</sup> (I)	1.63***
BCP99: young adults, lexical decision					
KF	0.99***	27.89***	30.05***	0.21** (I)	2.37***
TASA	0.99***	38.15***	38.79***	0.00 (I)	0.64***
BNC	0.99***	32.60***	37.84***	0.45*** (I)	5.69***
BCP99: older adults, lexical decision					
KF	0.76***	22.67***	34.63***	0.12* (I)	1.76***
TASA	0.76***	32.42***	32.66***	0.02	0.26***
BNC	0.76***	27.80***	32.52***	0.48*** (I)	5.20***
Elexicon: young adults, naming					
KF	37.24***	8.66***	9.07***	0.00 (I)	0.41***
TASA	37.24***	12.66***	12.90***	0.00 <sup>†</sup> (I)	0.24***
BNC	37.24***	12.17***	13.12***	0.03*** (I)	0.98***
Elexicon: young adults, lexical decision					
KF	32.04***	14.87***	15.53***	0.00 (I)	0.66***
TASA	32.04***	19.66***	20.03***	0.01* (I)	0.38***
BNC	32.04***	20.14***	21.05***	0.00 <sup>†</sup>	0.91***

**Note.** The following data sets were used: SB97 = Spieler and Balota (1997), BS98 = Balota and Spieler (1998), BCP99 = Balota, Cortese, and Pilotti (1999), Elexicon = Balota et al. (2000). The corpora from which the measures of WF and CD were obtained are as follows: KF = Kučera and Francis (1967), TASA = Touchstone Applied Science Associates (Landauer, Foltz, & Laham, 1998), BNC = British National Corpus (British National Corpus Consortium, 2000). All effects of WF and CD were calculated after those of covariates had been included. Unique effects are those calculated with the indicated variable entered last. I = inhibitory effect of WF.

<sup>†</sup> $p < .1$ . \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

of WF. The raw correlations among the variables<sup>4</sup> suggest that word length (in letters) is a likely contributor to the confound, as its correlation with log-WF is greater than its correlation with log-CD. The results summarized in Table 2 are consistent with this idea: In analyses including only log-WF, log-CD, and length as predictors, log-WF showed no unique facilitatory effect, but log-CD did. Moreover, for the critical analyses in which log-WF had appeared to have an effect when length was omitted, there was evidence of a unique (inhibitory) effect of length when it was introduced. Figure 1a illustrates the semipartial correlations of

log-WF and log-CD with response time as a function of word length for the Elexicon data. Facilitatory effects are consistently present for CD, but not for WF.

#### Do Semantic Variables Account for the Effect of CD?

Of course, CD may itself be confounded with some variable that was not controlled in this analysis. Whereas WF is subject to effects of structural variables, CD seems more likely to be influenced by semantic variables. Ambiguity, for instance, might be important in the present case, as words with multiple meanings should be used in multiple contexts. Abstract words are also likely to be used in a larger number of contexts than concrete words are. Indeed, Galbraith and Underwood (1973) found that abstract words are

<sup>4</sup>Because of space limitations, these correlations are not reported here, but they are available on the Web at <http://www.warwick.ac.uk/~psrcaj/cd/correls.html>.

**TABLE 2**

*Unique Effects of Log-Transformed Word Frequency (Log-WF), Log-Transformed Contextual Diversity (Log-CD), and Word Length From the Three Corpora in Each Data Set*

Data set and corpus	Effect ( $\Delta R^2$ in %)				
	Log-WF after CD	Log-CD after WF	Log-WF after CD and length	Log-CD after WF and length	Length after CD and WF
SB97: young adults, naming					
KF	0.05 <sup>†</sup>	0.04	0.02 (I)	0.37***	12.56***
TASA	0.27**	0.02 (I)	0.01 (I)	0.16*	11.35***
BNC	0.27**	0.12 <sup>†</sup>	0.04 (I)	0.93***	12.65***
BS98: older adults, naming					
KF	0.03	0.31**	0.05 (I)	0.79***	8.82***
TASA	0.31**	0.00 (I)	0.01	0.13*	7.36***
BNC	0.06	0.79***	0.13* (I)	1.97***	9.08***
BCP99: young adults, lexical decision					
KF	0.21** (I)	2.37***	0.24** (I)	2.44***	0.08*
TASA	0.01 (I)	0.70***	0.01 (I)	0.65***	0.00
BNC	0.45*** (I)	5.85***	0.55*** (I)	6.02***	0.07**
BCP99: older adults, lexical decision					
KF	0.10* (I)	1.71***	0.13* (I)	1.78***	0.11*
TASA	0.02	0.28***	0.02	0.27***	0.00 (F)
BNC	0.43*** (I)	5.16***	0.54*** (I)	5.37***	0.21**
Elexicon: young adults, naming					
KF	0.02**	0.41***	0.04*** (I)	0.63***	21.36***
TASA	0.02**	0.21***	0.07*** (I)	0.49***	13.15***
BNC	0.04***	0.73***	0.15*** (I)	1.43***	19.49***
Elexicon: young adults, lexical decision					
KF	0.05***	0.53***	0.01** (I)	0.78***	20.36***
TASA	0.05***	0.21***	0.03*** (I)	0.47***	11.57***
BNC	0.26***	0.55***	0.00 (I)	1.14***	17.94***

**Note.** The following data sets were used: SB97 = Spieler and Balota (1997), BS98 = Balota and Spieler (1998), BCP99 = Balota, Cortese, and Pilotti (1999), Elexicon = Balota et al. (2000). The corpora from which the measures of WF and CD were obtained are as follows: KF = Kučera and Francis (1967), TASA = Touchstone Applied Science Associates (Landauer, Foltz, & Laham, 1998), BNC = British National Corpus (British National Corpus Consortium, 2000). F = facilitatory effect of word length; I = inhibitory effect of WF or CD. No other covariates were considered in these analyses.

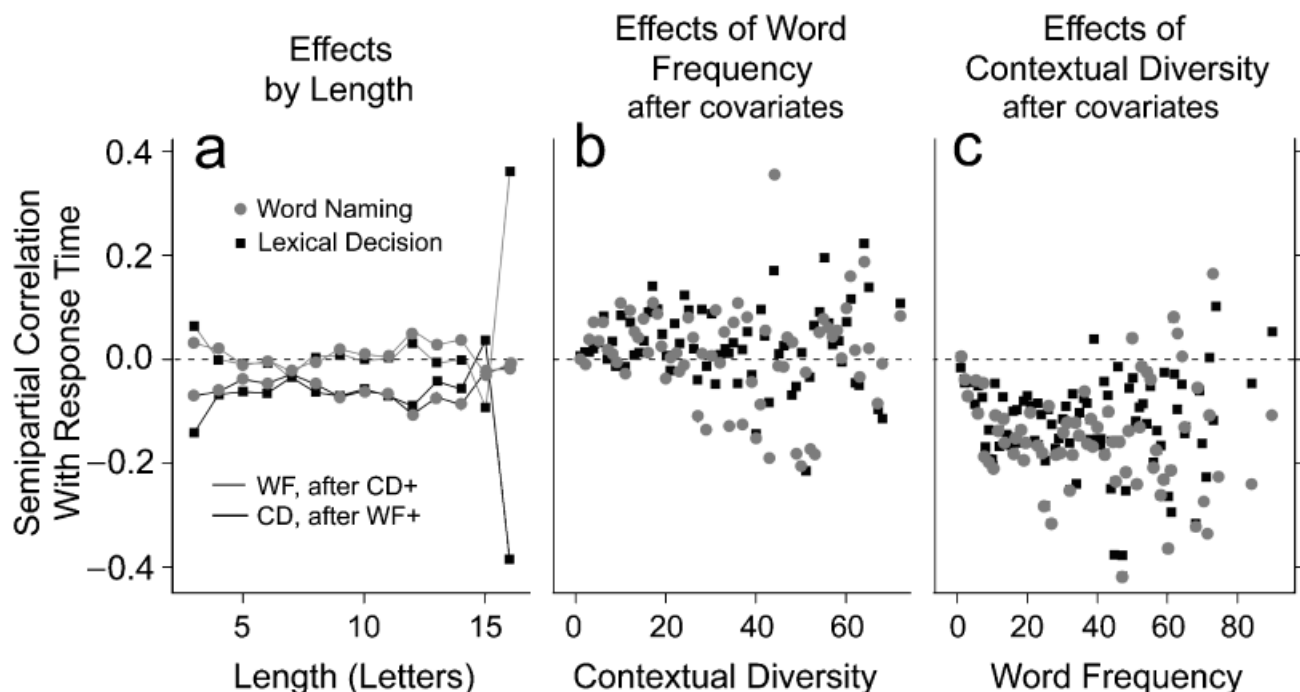
<sup>†</sup>  $p < .1$ . \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

rated by undergraduates to have more different contextual uses than concrete words, and Schwanenflugel and Shoben (1983) found that context availability (ease of remembering a context for a word) and diversity of contexts are negatively correlated with concreteness and predict lexical decision RTs. Imageability is conceptually related to concreteness, and often substituted for it in experimental designs.

We conducted analyses using the concreteness, imagery, and ambiguity norms from Gilhooly and Logie (1980) for the 1,812 words also found in the Elexicon database. The correlations<sup>5</sup>

<sup>5</sup>These correlations are also available on the Web at <http://www.warwick.ac.uk/~psrcj/cd/correls.html>.

between concreteness and CD appeared to be more negative than those between concreteness and WF, although TASA appeared to have greater proportions of types and tokens that were concrete (and to a lesser degree imageable) words than the other corpora did. Despite this relationship, including these variables in the analysis did not eliminate the effect of CD: As shown in Table 3, after these variables' effects were accounted for, CD still had a significant facilitatory effect, and WF still did not. Also, in these analyses, the BNC counts accounted for more variance than the TASA counts, a result consistent with the BNC being a larger corpus (by tokens); this may indicate that an apparent advantage for TASA in predicting the RTs in Table 1 comes from its relationship to concreteness and imageability, not its



**Fig. 1.** Semipartial correlations of TASA (Touchstone Applied Science Associates) log-word frequency (WF) and log-contextual diversity (CD) with response times in word naming and lexical decision in the Elexicon data, with orthographic neighborhood size, number of syllables, and onset (for naming only) partialled out. Points based on fewer than 50 words are omitted. The graph in (a) presents the correlations of both log-WF and log-CD as a function of word length; the correlations of log-WF have log-CD partialled out, and vice versa. The graph in (b) presents the correlations for log-WF for each value of log-CD, and the graph in (c) presents the correlations for log-CD for each value of log-WF.

greater number of passages. These analyses show that high CD is associated with faster responses regardless of imageability, concreteness, ambiguity, and other lexical measures, and high WF is not.

#### Can the Results Be Explained by the High Correlation Between WF and CD?

The obviously high correlation between log-WF and log-CD might cause concern to some readers in the context of our regressions, although the inferential logic of these analyses is unaffected by this collinearity.<sup>6</sup> One way to show that this correlation is not responsible for the results is to remove it by examining the effect of one variable while holding the other constant. Figure 1b shows the effect of log-WF on RT for individual values of CD; there is little or no evidence for a unique effect of WF. By contrast, Figure 1c, which shows the effect of log-CD for individual values of WF, demonstrates a consistent (and necessarily unique) facilitatory effect. Moreover, all the analyses summarized in Table 1 give evidence for a unique fa-

cilitatory effect of CD. Such a pattern would be unlikely even if Type I errors occurred at random in every analysis (because the signs would be inconsistent across analyses).

Nonetheless, the high correlations between measures of WF and measures of CD raise the possibility that WF is the better predictor, but log-CD shows a unique facilitatory effect because log-CD correlates better (more linearly) than log-WF with the most appropriate transformation of WF; both Balota et al. (2004) and Murray and Forster (2004) have found evidence of nonlinearity in the prediction of reading latencies from log-WF.

One possible reason that the logarithmic transformation might lead to misleading results is that the rank of a word's WF is a more linear predictor of these RTs than the logarithm of the word's WF. Murray and Forster (2004) provided some evidence that rank-WF is a better predictor than log-WF for lexical decision times when KF is used to derive these values; this is what they predicted from their model of lexical access, as it serially searches for lexical entries in lists that are frequency ordered. However, rank-WF<sup>7</sup> accounted for more variance than log-WF for only 8 of the 18 combinations of data sets and corpora examined here. These 8 included all 6 analyses involving KF measures of WF; this corpus gives the least reliable estimate of

<sup>6</sup>In instances of high collinearity, estimated coefficients are unbiased, but are subject to higher error than would otherwise be the case. Power is therefore reduced, but Type I error rates are not thereby inflated. The nonindependence of estimates and comparatively high sensitivity to small changes in the data make interpretations of coefficient magnitudes unconvincing, but allow null-hypothesis significance testing. The negative effects on power are mitigated by the large sample sizes used in our analyses.

<sup>7</sup>Our calculations of rank differ somewhat from those of Murray and Forster (2004) because we do not consider any entries to be "spurious."

**TABLE 3**

*Unique Effects of Contextual Diversity (CD), Word Frequency (WF), and Semantic Variables, After Covariates, in the Elexicon Data*

Data set and corpus	Effect ( $\Delta R^2$ in %)					
	Concreteness	Imagery	Ambiguity	Log-WF	Log-CD	Total
Elexicon naming						
KF	0.60*** (I)	3.10***	0.14*	0.00 (I)	0.39***	59.11***
TASA	0.77*** (I)	1.77***	0.12*	0.01 (I)	0.24**	60.24***
BNC	0.21** (I)	1.70***	0.08 <sup>†</sup>	0.02 (I)	0.86***	60.68***
Elexicon lexical decision						
KF	0.74*** (I)	3.18***	0.30***	0.01	0.50***	59.63***
TASA	1.14*** (I)	1.43***	0.28***	0.04 (I)	0.14*	61.90***
BNC	0.21** (I)	1.52***	0.12*	0.00	1.14***	62.65***

**Note.** This analysis was conducted using only those words included in Gilhooly and Logie's (1980) semantic norms. Elexicon = Balota et al. (2000). The corpora from which the measures of WF and CD were obtained are as follows: KF = Kučera and Francis (1967), TASA = Touchstone Applied Science Associates (Landauer, Foltz, & Laham, 1998), BNC = British National Corpus (British National Corpus Consortium, 2000). All effects are those with the indicated variable entered after all others, including covariates. I = inhibitory effect of concreteness or of WF.

<sup>†</sup>  $p < .1$ . \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

WF because of its size, is the least predictive of RTs, gives the smallest range of CD values, and is most subject to negative bias in the estimation of ranks of low-frequency words.<sup>8</sup>

Table 4 presents the results for rank-WF and rank-CD for the eight analyses in which rank-WF accounted for more variance than log-WF. In all eight cases, there was a unique effect of rank-CD, such that high rank-CD led to fast responses. Six of these analyses yielded a significant unique effect of rank-WF. The three of these involving monosyllabic data used KF frequency counts, but for these data TASA accounted for more variance than KF, even when WF was ranked, and log-WF from TASA accounted for even more variance. This suggests that the counts from KF and a ranking transformation were both inappropriate. Moreover, the power transformation we discuss next accounted for much more variance in all cases. Furthermore, because rank-WF did not in any of these instances eliminate a unique effect of rank-CD, the resulting regression formulas did not correspond to any simple (or readily interpretable) version of a rank-hypothesis serial-search model. Moreover, in every case, CD was a stronger predictor than WF, even when ranked measures were used.

The power law of practice (e.g., Newell & Rosenbloom, 1981) represents a further possibility for an appropriate transformation of frequency for predicting word-processing times (Kirsner & Spelman, 1996). That is, the most appropriate transformation of WF might be some power function (with a negative exponent). The analyses presented in Table 5 tested the possibility that the advantage of CD over WF would disappear when both measures underwent a power-law transformation (with the exponent always a negative free parameter). Broadly speaking, using this transformation led to large increases in the variance

accounted for by WF and CD. As Table 5 shows, in 17 of the 18 analyses, CD accounted for more residual variance than WF. In all 18 analyses, including CD led to a significant increase in  $R^2$ , with high CD predicting fast responses; although including WF led to significant increases in  $R^2$  in all 18 analyses, low WF was predictive of fast responses in every case.

### Is Corpus CD Just a Better Indicator of Real-World WF?

A final possibility that we consider is that the results suggesting that CD is the cause of apparent WF effects were due to the CD measure from a corpus being more correlated with real-world WF (the frequency in the language as a whole) than is the WF measure from the same corpus.<sup>9</sup> This could occur as a result of WF being more influenced than CD by idiosyncratic properties of individual passages, as one obscure word might occur many times in one passage,<sup>10</sup> inflating WF greatly, but CD only slightly.

To take an extreme example of how corpus CD could be a better reflection of real-world WF than corpus WF is, suppose words did cluster, but not to differing degrees. That is, suppose that the probability that each word occurs in a particular document at all is proportional to the word's real-world frequency, and if the word does occur, it occurs with equal probability either once or 25 times. In this scenario, (proportional) real-world WF and CD are the same thing, and (proportional) corpus WF and

<sup>9</sup>We thank David Balota for highlighting this possibility.

<sup>10</sup>This issue would be especially problematic for long passages. The BNC is the only corpus examined here that has sizable variability in passage size. We therefore conducted an analysis in which we weighted each occurrence by the reciprocal of the length of the passage, so that all passages contributed equally to the WF count. However, this decreased the correlation with RTs, and the analysis still favored CD.

<sup>8</sup>The relationship between ranks estimated from different corpora is nonlinear.

TABLE 4

*Effects of Rank Word Frequency (Rank-WF) and Rank Contextual Diversity (Rank-CD) for the Data Set–Corpus Combinations in Which Rank-WF Performed Better Than Log-WF*

Data set and corpus	Effect ( $\Delta R^2$ in %)				
	Covariates	Rank-WF	Rank-CD	Rank-WF unique	Rank-CD unique
SB97: young adults, naming KF	46.56***	5.04***	5.09***	0.09*	0.14**
BS98: older adults, naming KF	27.20***	10.94***	11.34***	0.09*	0.49***
BCP99: young adults, lexical decision KF	1.16***	29.24***	30.77***	0.12*	1.65***
BCP99: older adults, lexical decision KF	0.94***	26.07***	27.86***	0.04	1.85***
TASA	0.66***	31.66***	32.85***	0.05 (I)	1.38***
Elexicon: young adults, naming KF	38.71***	8.55***	8.85***	0.10***	0.40***
TASA	36.17***	12.10***	12.38***	0.02**	0.30***
Elexicon: young adults, lexical decision KF	34.32***	13.37***	13.71***	0.19***	0.53***

**Note.** The following data sets were used: SB97 = Spieler and Balota (1997), BS98 = Balota and Spieler (1998), BCP99 = Balota, Cortese, and Pilotti (1999), Elexicon = Balota et al. (2000). The corpora from which the measures of WF and CD were obtained are as follows: KF = Kučera and Francis (1967), TASA = Touchstone Applied Science Associates (Landauer, Foltz, & Laham, 1998). All effects of WF and CD were calculated after those of covariates had been included. Unique effects are those with the indicated variable entered last. I = inhibitory effect of WF.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

CD are both unbiased estimators of real-world WF, but corpus CD has much lower variance,<sup>11</sup> because it is not distorted by low-frequency words that by chance occur 25 times in more than half the passages that they occur in. Consistently different levels of clustering between words are necessary for CD to be conceptually distinct from WF. The ratio of CD to WF can be used as an index of clustering. This index correlates well between the different corpora in this study (correlations calculated for words in the Elexicon data were as follows: KF-TASA, .362; KF-BNC, .485; TASA-BNC, .414), indicating that much of the clustering at play is not idiosyncratic to any particular corpus; that is, CD is reliable for reasons unrelated to corpus WF.

The preceding discussion does not, however, address the more subtle possible ways in which corpus CD could be a better estimate of real-world WF than corpus WF is. For instance, if corpus WF is biased as an estimate of real-world WF because of contextual factors, and if corpus CD is more unbiased, CD could be a better predictor. One way to approach the question of

whether corpus CD reflects real-world WF better than corpus WF does is to use pairs of corpora to see whether (a) WF is consistently predicted better by CD than by WF and (b) CD consistently predicts WF better than CD; either eventuality would be damaging for the case that CD has true effects. The raw correlations<sup>12</sup> did not yield consistent answers (in three of the six cases, CD predicted WF better than WF did; in two of the six cases, CD predicted WF better than CD). Hence, this analysis neither supports nor disconfirms the suggestion that CD consistently acts as a better measure of WF. However, similar analyses can be conducted with randomly chosen halves (half the passages) of each corpus. We conducted 100 such random splits for each corpus, investigating the predictions of log-WF and of log-CD. WF was predicted slightly (but highly significantly) better by WF than by CD (KF: .7816 vs. .7798,  $SE_{\text{diff}} = .00019$ ; TASA: .9340 vs. .9333,  $SE_{\text{diff}} = .00003$ ; BNC: .9721 vs. .9553,  $SE_{\text{diff}} = .00012$ ), and CD predicted CD somewhat (and highly significantly) better than it predicted WF (KF: .7928 vs. .7812,  $SE_{\text{diff}} = .00019$ ; TASA: .9423 vs. .9338,  $SE_{\text{diff}} = .00003$ ;

<sup>11</sup>It is generally the case that CD estimates are more stable than WF estimates. This is why CD correlates better with itself than WF correlates with itself over split halves of a corpus.

<sup>12</sup>These values are presented on the Web site referred to in footnote 5, at <http://www.warwick.ac.uk/~psrcj/cd/correls.html>.

**TABLE 5**

*Effects of Power-Law-Transformed Word Frequency (Pow-WF), Power-Law-Transformed Contextual Diversity (Pow-CD), Log-Transformed Adjusted Word Frequency (Log-U), and Log-Transformed Contextual Diversity (Log-CD) as Calculated From Each Corpus*

Data set and corpus	Effect ( $\Delta R^2$ in %)						Log-CD unique
	Pow-WF	Pow-CD	Pow-WF unique	Pow-CD unique	Log- <i>U</i>	Log- <i>U</i> unique	
SB97: young adults, naming							
KF	5.89***	5.67***	0.21** (I)	0.19**	4.84***	0.08* (I)	0.59***
TASA	7.16***	7.29**	0.13* (I)	0.25***	6.63***	0.00	0.19***
BNC	6.33***	6.74***	0.16** (I)	0.57***	6.03***	0.04 (I)	0.73***
BS98: older adults, naming							
KF	11.70***	11.93***	0.60*** (I)	0.82***	9.71***	0.26*** (I)	1.45***
TASA	14.50***	14.76***	0.19** (I)	0.45***	13.37***	0.04	0.49***
BNC	12.31***	13.29***	0.24*** (I)	1.23***	11.92***	0.08 <sup>†</sup> (I)	1.45***
BCP99: young adults, lexical decision							
KF	33.12***	35.46***	0.40*** (I)	1.75***	26.18***	1.32*** (I)	5.19***
TASA	41.35***	42.49***	2.26*** (I)	3.40***	38.12***	0.00	0.65***
BNC	35.21***	38.88***	0.92*** (I)	3.59***	33.71***	0.28*** (I)	4.41***
BCP99: older adults, lexical decision							
KF	28.04***	29.22***	1.31*** (I)	2.48***	20.92***	1.71*** (I)	5.41***
TASA	37.20***	38.16***	0.69*** (I)	1.65***	31.11***	0.37*** (I)	1.93***
BNC	31.33***	33.16***	0.18* (I)	2.01***	28.58***	0.40*** (I)	4.35***
Elexicon: young adults, naming							
KF	9.78***	9.99***	0.35*** (I)	0.58***	8.66***	0.00 (I)	0.41***
TASA	14.27***	14.61***	0.34*** (I)	0.50***	12.66***	0.00 <sup>†</sup> (I)	0.24***
BNC	12.27***	13.12***	0.25*** (I)	1.10***	12.17***	0.03*** (I)	0.98***
Elexicon: young adults, lexical decision							
KF	16.20***	16.56***	0.37*** (I)	0.73***	14.87***	0.00 (I)	0.66***
TASA	20.74***	21.20***	0.20*** (I)	0.66***	19.66***	0.01* (I)	0.38***
BNC	20.21***	21.05***	0.24*** (I)	1.08***	20.14***	0.00 <sup>†</sup>	0.91***

**Note.** The power-law and *U* analyses were conducted separately. The following data sets were used: SB97 = Spieler and Balota (1997), BS98 = Balota and Spieler (1998), BCP99 = Balota, Cortese, and Pilotti (1999), Elexicon = Balota et al. (2000). The corpora from which the measures of WF and CD were obtained are as follows: KF = Kučera and Francis (1967), TASA = Touchstone Applied Science Associates (Landauer, Foltz, & Laham, 1998), BNC = British National Corpus (British National Corpus Consortium, 2000). Both sets of effects were calculated after those of covariates, whose effects were as in Table 1, had been included. The nonunique effect of log-CD was also as in Table 1. Unique effects are those with the indicated variable entered last. I = inhibitory effect of WF or of *U*.

<sup>†</sup>  $p < .1$ . \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

BNC: .9790 vs. .9549,  $SE_{\text{diff}} = .00011$ ). These results appear to exclude the possibility that CD is a better indicator of WF than are observations of WF itself.

Finally, we used a standard adjustment for clustered sampling of WF estimates, Carroll's *U*, which adjusts frequency estimates downward for words occurring in few contexts. Table 5 presents results of these analyses, which are analogous to the analyses in Table 1. Essentially, the same pattern of results obtains: All 18 analyses showed a unique facilitatory effect of CD, and none showed a unique facilitatory effect of adjusted WF (*U*); many showed adjusted WF had unique inhibitory effects.

## DISCUSSION

In both word naming and lexical decision, CD was more predictive of RTs than WF was. Moreover, CD had a unique effect, with high CD leading to fast responses, whereas WF had either no unique effect or a suppressor effect, with high WF leading to slow responses. This implies there is a facilitatory effect of CD, but no facilitatory effect of WF per se. This pattern was found even when ambiguity, imagery, and concreteness were controlled; was not an artifact of the strong correlation between the CD and WF variables; and does not appear to be due to the



clustering properties of the corpora, both because CD did not predict WF better than WF did and did not predict WF better than CD and because the same pattern of results was obtained even when WF was adjusted for clustering.

According to the rational analysis of memory (Anderson & Milson, 1989; Anderson & Schooler, 1991), number of contexts has an effect because the more contexts an item has occurred in, the more likely that item is to be needed in any new context; and as a result of different words clustering within particular contexts to differing degrees, WF is a relatively poor indicator of likely need. Recently needed items also have high likely need, and recency certainly affects memory (e.g., Rubin & Wenzel, 1996). Because CD is a good indicator of the probable recency of an item, it is feasible that recency, and not CD per se, drives the CD effect. However, when the recency of items is controlled by introducing recent repetitions, the (apparent) WF effect is diminished, but not eliminated (Balota & Spieler, 1999; Kirsner & Speelman, 1996). This would not be the case if recency were the key factor in the CD effect.

Previous attempts to link CD to lexical decision latencies have also used local windows of semantic context to derive (information-theoretic entropy) values based on contextual predictability (McDonald & Shillcock, 2001). Although the variable thus derived did have an effect distinct from that of WF, it did not entirely eliminate the WF effect, possibly because temporal, as well as semantic, aspects of context contribute to the CD effect.

Learning-based models of reading cannot accommodate these results unless they are modified so that learning mechanisms are sensitive to context, not frequency. Models of reading that attribute frequency effects to frequency-sensitive units in dictionary-like lexicons, but do not specify the source of this sensitivity, could be modified so that these units are sensitive to CD. However, such modifications would seem to violate the principle that only orthographic forms are stored in the orthographic lexicon, and only phonological forms are stored in the phonological lexicon (Coltheart, 2004). By contrast, according to a view that reading uses the same kind of memorial resources as recall, the present results are natural. They therefore motivate a theory of reading based on principles from memory research.

**Acknowledgments**—This work was supported by a Warwick Postgraduate Research Fellowship to J.S.A. and by Grants RES 000221558 and PTA 026270716 from the Economic and Social Research Council (UK) and Grant F/215/AY from the Leverhulme Trust. We thank Marjolein Merckx, Chris Kent, Elizabeth Maylor, Neil Stewart, and Matthew Roberts for comments on this work.

## REFERENCES

- Anderson, J.R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, 96, 703–719.
- Anderson, J.R., & Schooler, L.J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396–408.
- Baayen, R.H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (Release 2) [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
- Balota, D.A., Cortese, M.J., Hutchison, K.A., Neely, J.H., Nelson, D., Simpson, G.B., & Treiman, R. (2000). *The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords*. Retrieved December 11, 2004, from <http://elexicon.wustl.edu/>
- Balota, D.A., Cortese, M.J., & Pilotti, M. (1999). Item-level analysis of lexical decision: Results from a mega-study. In *Abstracts of the 40th annual meeting of the Psychonomic Society* (p. 44). Los Angeles: Psychonomic Society.
- Balota, D.A., Cortese, M.J., Sergent-Marshall, S.D., Spieler, D.H., & Yap, M.J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, 133, 283–316.
- Balota, D.A., & Spieler, D.H. (1998). The utility of item-level analyses in model evaluation: A reply to Seidenberg and Plaut. *Psychological Science*, 9, 238–240.
- Balota, D.A., & Spieler, D.H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, 128, 32–55.
- British National Corpus Consortium. (2000). *British National Corpus world edition* [CD-ROM]. Oxford, England: University of Oxford, Humanities Computing Unit.
- Coltheart, M. (2004). Are there lexicons? *Quarterly Journal of Experimental Psychology*, 57A, 1153–1171.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204–256.
- Forster, K.I., & Chambers, S.M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 12, 627–635.
- Frederiksen, J.R., & Kroll, J.F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 361–379.
- Galbraith, R.C., & Underwood, B.J. (1973). Perceived frequency of concrete and abstract words. *Memory & Cognition*, 1, 56–60.
- Gilhooly, K., & Logie, R. (1980). Age of acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, 12, 395–427.
- Glenberg, A.M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, 15, 1–15.
- Glenberg, A.M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, 7, 95–112.
- Kirsner, K., & Speelman, C. (1996). Skill acquisition and repetition priming: One principle, many processes? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 563–575.
- Kučera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Landauer, T.K., Foltz, P.W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25, 259–284.
- 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, 295–323.

- Murray, W.S., & Forster, K.I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, 111, 721–756.
- Newell, A., & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.
- Plaut, D.C., McClelland, J.L., Seidenberg, M.S., & Patterson, K. (1996). Understanding normal and impaired reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56–115.
- Rubin, D.C., & Wenzel, A.E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, 103, 734–760.
- Schwanenflugel, P.J., & Shoben, E.J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 82–102.
- Seidenberg, M.S., & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.
- Spieler, D.H., & Balota, D.A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, 8, 411–416.
- 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.
- Verkoeijen, P.P.J.L., Rikers, R.M.J.P., & Schmidt, H.G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 796–800.
- Zeno, S.M., Ivens, S.H., Millard, R.T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.
- Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-route model. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1131–1161.

(RECEIVED 4/11/05; REVISION ACCEPTED 12/16/05;  
FINAL MATERIALS RECEIVED 1/24/06)