Running head: CONTEXTUAL DIVERSITY VS. WORD FREQUENCY

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

Word count: 3,457 words.

Corresponding Author:

James S. Adelman

Department of Psychology,

University of Warwick,

Gibbet Hill Road,

COVENTRY,

CV4 7AL,

UK.

Telephone: +44 (0) 24 7652 3158

Electronic mail: J.S.Adelman@warwick.ac.uk

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. Using a normative, corpus-based, measure of contextual diversity, word frequency effects were eliminated by contextual diversity (but not vice versa) across three naming and three lexical decision datasets, using any of three corpora to derive the frequency and contextual diversity values. This result is incompatible with existing models of visual word recognition, which attribute frequency effects directly to frequency, and is particularly problematic for accounts in which frequency effects reflect learning. It is argued that the result reflects the importance of likely need in memory, and that the continuity with memory suggests using principles from memory research to inform theorizing about reading.

Contextual Diversity Not Word Frequency Determines Word Naming and Lexical Decision Times

What determines how quickly a word can be read? Empirically, in both word naming and lexical decision, frequency of occurrence is among the strongest known factors: Frequent words are read more quickly than infrequent words (Forster & Chambers, 1973; Frederiksen & Kroll, 1976; Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). 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 the effect is due to the number of experiences with a word; each (and every) exposure has a long-term influence on accessibility.

In learning-based accounts of reading, such as connectionist models (e.g., Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg, & Patterson, 1996; 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 with 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 upon the separation of the exposures in time and context (Glenberg, 1976, 1979). Indeed, under some conditions, if neither changes, 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, that is, word frequency, will be diminished or abolished when these repetitions occur in the same context. Instead, the number of contexts in which a word is experienced, its contextual diversity (CD), should determine its accessibility and hence response times 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 contain that word; such a measure has shown CD to have effects on recognition memory that are distinguishable from WF effects (Steyvers & Malmberg, 2003). Here we compare the ability of CD and WF to predict six existing sets of data regarding response times in word naming and lexical decision, basing our analyses on CD and WF measures from each of three corpora.

Method

Dependent Variables

Item mean RTs from the word naming (reading aloud) and lexical decision (judging whether the stimulus is a word or not) tasks from six datasets made available by Balota and colleagues were used. Two of these datasets contain data for the word naming of 2,820 (2,776 analysed here) uninflected one-syllable words by young adults (Spieler & Balota, 1997) and older adults (Balota & Spieler, 1998); a further two contain data on lexical decision for young and older groups for the same words (Balota, Cortese, & Pilotti, 1999); and the last two contain data for young adults on both tasks for a broader selection of 40,481 (39,383 analysed here) words (Balota et al., 2000).

Independent Variables

Word frequency (number of occurrences) and contextual diversity (number of passages/documents in which a word occurs) were calculated from three corpora. Kučera and Francis (1967) provide these counts for the Brown corpus. This is a samples corpus containing 500 samples (target length 2000 tokens) from distinct documents spread evenly over 15 genres. These have a mean length of 2030 tokens (*SD* 42). Counts were compiled by the present authors from the 12th grade level portion of the LSA/TASA (Landauer, Foltz, & Laham, 1998) corpus which is formed from texts used in the compilation of the Zeno, Ivens, Millard, and Duvvuri (1995) frequency norms, which are designed to sample the likely experience of students through the American school system[1]. The 28,882 samples from distinct documents in this section of the corpus have a mean length of 286 tokens (*SD* 25). The authors also compiled counts from the written portion of the BNC (British National Corpus Consortium, 2000). This corpus is designed have the largest possible samples, ideally whole texts. There are 3144 samples of various forms and lengths from pamphlets through book chapters to whole issues of newspapers[2]. The mean number of tokens in each passage is 26,892 (*SD* 25,914). Where logarithm and power-law fits are calculated, all counts are incremented by 1, to avoid problems from zero counts. Items with zero counts are excluded from the rank analyses.

The following factors from CELEX[3] (Baayen, Piepenbrock, & Gulikers, 1995) were covaried out in the analyses: letter length; orthographic neighborhood size; for the monosyllabic databases only, rime consistency; where applicable, number of syllables; and, for word naming only, initial phoneme.

Results

In the next section of the paper 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. Subsequent sections exclude 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 CD measure.

Does corpus CD or corpus WF predict word naming and lexical decision times?

Table 1 presents the results of analyses using log-transformations of WF and CD; log-WF is generally agreed to approximate a linear predictor of naming and lexical decision RTs. After variance attributed to covariates, introducing either WF or CD accounted for significant additional variance (lines two and three of the table give these squared semipartial correlations (ΔR^2) after the covariates described above), 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. When the unique effects are examined, in all 18 analyses there was a unique effect of CD (line five of the table gives squared semipartial correlations after covariates and log-WF). Six analyses showed a unique effect of WF (line four of the table giving squared semipartial correlations after covariates and log-CD), all such that high WF led to slow RTs, that is, WF acted as a suppressor variable. These results suggest not only that CD, rather than WF, best predicts 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.

Since it is the addition of CD to the regression equation that eliminates the unique effect of WF, CD must be a critical component of the confound. 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 of WF. The raw correlations between the variables[4] suggest that (letter) length is a likely candidate for a contributor to the confound, as its correlation with log-WF is greater than with log-CD. Consistent with this, the analyses summarized in Table 2 with log-WF, log-CD and length (only) as predictors showed no unique facilitatory effect of log-WF, but a unique facilitatory effect of log-CD. Moreover, for the critical analyses where log-WF had appeared to have an effect when length was omitted, there was evidence of a unique (inhibitory) effect of length. Figure 1(a) illustrates the semipartial correlations of log-WF and log-CD with response times on a length-by-length basis for the Elexicon data. Facilitatory effects of CD are consistently present, but this is not so for WF.

Do semantic variables account for the effect of CD?

Of course, CD may itself be confounded with some variable that has not been controlled in this analysis. Whilst WF is more strongly subject to effects of structural variables, CD seems more likely to be influenced by semantic variables. Ambiguity, for instance, might be important here, 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. Indeed, Galbraith and Underwood (1973) find that abstract words are rated to have more different contextual uses than concrete words, and Schwanenflugel and Shoben (1983) find that context availability and diversity of contexts are correlated with concreteness, and predict lexical decision response times. 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 1812 words they had in common with the Elexicon database. The correlations[5] between concreteness and CD appeared to be more negative than those between concreteness and WF, although TASA, in general, appeared to be biased towards more concrete words. Despite this, none of these variables eliminated the CD effect: As can be seen in Table 3, after these variables' effects, there remained a significant facilitatory effect of CD, and none of WF. Also, here the BNC counts accounted for more variance than the TASA counts, consistent with its being a larger corpus (by tokens); this may indicate that the apparent advantage for TASA comes from its relationship to imageability and concreteness, not its greater number of passages. High CD is associated with faster responses regardless of imageability, concreteness, ambiguity and other factors, and high WF is not.

Can the results be explained away by the high correlation between WF and CD?

The high correlation between log-WF and log-CD might cause concern to some readers in the context of our regressions, although the inferential logic is unaffected by the collinearity[6]. One way to illustrate that no simple problem exists is to remove the correlation, by examining the effect of one variable whilst holding the other constant. Figure 1(b) shows the effect of log-WF on response times for individual values of CD; there is little or no evidence for a unique effect of WF. By contrast, Figure 1(c), showing log-CD effects for individual values of WF, demonstrates a consistent (and necessarily unique) facilitatory effect. Moreover, all the analyses described in Table 1 give evidence for a unique effect of CD in the facilitatory direction. Such a pattern would be unlikely even if Type I errors were occurring at random in every analysis (as the signs should be inconsistent between the analyses in this case).

Nonetheless, the high correlations between measures of WF and CD also raise the possibility that this result comes about despite WF being the better predictor because log-CD correlates better (more linearly) with the 'correct' transformation of WF than does log-WF; both Balota et al. (2004) and Murray and Forster (2004) have found evidence of nonlinearity in the prediction of latencies in reading from log-WF.

One such possibility is that the rank of a word's WF is a more linear predictor of these RTs. Murray and Forster (2004) provide some evidence that rank-WF is a better predictor than log-WF from Kučera and Francis (1967) for lexical decision times; this is what they predict from their model of lexical access, as it serially searches for lexical entries in lists that are frequency-ordered. Across the 18 dataset-corpus combinations examined here, however, rank-WF[7] accounted for more variance than log-WF for only eight, including all six analyses with Kučera and Francis (1967); this count is the least reliable estimate of WF, is the least predictive of RTs, has the smallest range of CD values, and is most subject to negative bias in the estimation of ranks of low-frequency words[8]. In Table 4, the comparison between rank-WF and rank-CD may be seen for the eight cases where rank-WF accounted for more variance than log-WF. In all eight of these analyses there is a unique effect of rank-CD, such that high rank-CD leads to fast responses. Six of the eight analyses yield a significant unique effect of rank-WF. The three of these involving monosyllabic data use the K-F frequency count, but for these data TASA accounts for more variance than K-F, even when ranked, and log TASA accounts for even more variance. This suggests that K-F counts and a ranking transformation are both inappropriate here. Moreover, the power transformation discussed below accounts for much more variance in all cases. Furthermore, rank WF does not in these instances eliminate a unique effect of rank CD, with the consequence that the resulting regression formula does not correspond to any simple (or readily interpretable) version of a rank-hypothesis serial search model. Moreover, in every case, CD is a stronger predictor than WF, even when ranked measures are used.

Another possibility is that the power law of practice (e.g., Newell & Rosenbloom, 1981) is followed by WF effects in reading (Kirsner & Speelman, 1996). This would mean that the best transformation of WF is some (negative) power function. 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 as a negative free parameter). Broadly speaking, using this transformation led to large increases in the variance accounted for by WF or CD. As can be seen in Table 5, in 17 of the 18 analyses, CD accounted for more residual variance than WF. In all 18 analyses, CD led to a significant increase in R^2 based on a relationship such that high CD was predictive of fast responses, whilst significant increases in R^2 from WF were based on low WF being predictive of fast responses.

Is corpus CD just a better indicator of real world WF?

A final possibility that we consider is that these results come about because the CD measure from a corpus is more correlated with real world WF (the frequency in the language as a whole) than is the WF measure from the same corpus[9]. This could occur as a result of WF being more influenced by idiosyncratic properties of individual passages than is CD, as one obscure word might occur many times in one passage[10], inflating WF greatly, but CD only slightly.

As an extreme example of this, if words did cluster, but not to differing degrees, this would necessarily be the case. Suppose that each word occurred in a particular document with a probability proportional to its frequency, and if it did occur, it occurred 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 CD are both unbiased as estimators of real world WF, but corpus CD has much lower variance[11], 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 a clustering index. This index correlates well between the different corpora (K-F vs. TASA: .362; K-F vs. BNC: .485; TASA vs. BNC: .414, words in Elexicon data), indicating that much of the clustering here is not idiosyncratic to any particular corpus, that is, CD is reliable for reasons unrelated to corpus WF.

The preceding does not, however, address the more subtle possibilites; corpus WF might be biased as an estimate of real world WF due to contextual factors, and corpus CD more unbiased, leading to its better predictions. One way to approach the question of whether corpus CD better reflects real world WF is to use pairs of corpora to see whether (i) WF is consistently predicted better by CD than by WF, and (ii) CD consistently predicts WF better than CD; either such eventuality would be damaging for the case that there are true CD effects. Examination of the raw correlations does not yield consistent answers (3 out of 6 pairs are consistent with (i), and 2 out of 6 with (ii)), and hence does not support the suggestion that CD is consistently acting as a 'better' measure of WF, but nor does this pattern allow us to reject the suggestion. 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, and investigated the predictions of log-WF, and the predictions by log-CD. WF is predicted slightly (but highly significantly) better by WF than CD (K-F: .7816 vs. .7798, SE_{diff} .00019; TASA: .9340 vs. .9333, SE_{diff} .00003; BNC: .9721 vs. .9553, SE_{diff} .00012) , and CD predicts CD somewhat (and highly significantly) better than it does WF (K-F: .7928 vs. .7812, SE_{diff} .00019; TASA: .9423 vs. .9338, SE_{diff} .00003; BNC: .9790 vs. 9549, SE_{diff} .00011). These results appear to exclude the possibility that CD is a better indicator of WF than observations of WF itself.

Finally, we used a standard adjustment for clustered sampling of word frequency estimates, Carroll's U: This adjusts frequency estimates downwards for words occurring in few contexts. Table 5 presents the relevant analyses, analogously to Table 1. Essentially, the same pattern of results obtains: All 18 analyses show a unique facilitatory effect of CD, and none shows a unique facilitatory effect of adjusted WF (U), with many showing unique inhibitory effects.

Discussion

In both word naming and lexical decision contextual diversity was more predictive of reaction times than word frequency. Moreover, CD had a unique effect such that high CD led to fast responses, whilst WF had no unique effect or a suppressor effect with high WF leading to slow responses. This implies there is a CD effect, but no facilitatory effect of WF per se. This (i) held even when ambiguity, imagery and concreteness were controlled, (ii) was not artefactual of the strong correlation between the CD and WF variables, and (iii) did not appear to be a result the clustering properties of corpora as CD did not better predict WF, and the result held 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 an item occurring in many contexts is more likely to be needed in any new context, and since different words cluster 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). Since CD is a good indicator of the probable recency of an item, it is feasible that recency, and not CD per se, that drives the CD effect. However, when the recency of items is controlled by introducing recent repetitions, the (apparent) WF effect is diminished, but not abolished (Kirsner & Speelman, 1996; Balota & Spieler, 1999), which would not occur if recency were the key factor in the CD effect.

Previously, attempts to link contextual diversity 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 this variable did have an effect distinct from WF, it did not entirely eliminate the WF effect. Possibly this occurs because temporal, as well as semantic, aspects of context contribute to the CD effect.

Learning-based models of reading cannot accommodate these results without modifications to learning mechanisms to make them 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 to be 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 in the phonological lexicon (Coltheart, 2004). By contrast, on a view that reading uses the same kind of memorial resources as recall, the result is natural. The present results motivate a theory of reading based on principles from memory research.

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: Linguistic Data Consortium, University of Pennsylvania.

Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., et al. (2000). *The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords.* Retrieved 11th December, 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 Psychonomics Society* (p. 44). Los Angeles, CA: 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: Humanities Computing Unit, University of Oxford.

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 and 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.

Author Note

This work was supported by a Warwick Postgraduate Research Fellowship to JSA 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 Merkx, Chris Kent, Elizabeth Maylor, Neil Stewart and Matthew Roberts for comments on this work.

Footnotes

[1]This corpus is also described at `http://lsa.colorado.edu/spaces.html` . We use this grade level ("TASA12") because frequency computed from it is a better predictor of RTs than frequency counted across the whole corpus. This probably reflects the fact that the full corpus is too heavily weighted toward college-level texts to be representative of undergraduate participants or education-matched controls.

[2]See also `http://www.natcorp.ox.ac.uk/what/balance.html` .

[3]CELEX was not used as a frequency count 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.

[4]These have been made available at `http://www.warwick.ac.uk/~psrcaj/cd/correls.html` for reasons of space.

[5]These are also presented at the web page referenced in Footnote 4.

[6]Estimated coefficients are in this context unbiased, but are subject to higher error. Power is therefore reduced, but Type I error rates are not thereby inflated. The non-independence of estimates and comparatively high sensitivity to small changes in the data speak against interpretation of coefficient magnitudes, but allows null-hypothesis significance testing. The negative effects on power are mitigated by large sample sizes.

[7]Our calculations of rank differ somewhat from those of Murray and Forster (2004) because we do not consider any entries to be 'spurious'.

[8]The relationship between ranks estimated from different corpora is nonlinear.

[9]We thank David Balota for highlighting this possibility.

[10]This would be especially problematic for longer passages. The BNC is the only corpus with sizeable variability in passage size. To counter this, each occurrence was weighted by the reciprocal of the length of the passage, so that each passage gave an equal contribution the WF count. However, this decreased the correlation with RTs, and the

analyses still favored CD.

[11]It is generally the case that CD estimates are more stable: This is why CD correlates better with itself than WF correlates with itself over split halves of a corpus.

Table 1

Overall (after covariates) and unique effects (ΔR^2 in %) of log-transformed Word Frequency (log-WF) and log-transformed Contextual Diversity (log-CD) from three corpora across six datasets.

Data	SB97 Young Naming			BS98 Older Naming			BCP99 Young LDT		
Corpus	K-F	TASA	BNC	K-F	TASA	BNC	K-F	TASA	BNC
Effect									
Covariates	45.95^c	45.95^c	45.95^c	26.72^c	26.72^c	26.72^c	.99^c	.99^c	.99^c
log-WF	5.16^c	6.73^c	5.85^c	10.39^c	13.73^c	11.74^c	27.89^c	38.15^c	32.60^c
log-CD	5.35^c	6.82^c	6.72^c	10.90^c	13.82^c	13.29^c	30.05^c	38.79^c	37.84^c
log-WF unique	.00	.00	*.07*†	*.00*	.02	*.08*†	*.21*^{b}	*.00*	*.45*^{c}
log-CD unique	.19^b	.09^a	.94^c	.51^c	.11^a	1.63^c	2.37^c	.64^c	5.69^c

Data	BCP99 Older LDT			Elexicon Naming			Elexicon LDT		
Corpus	K-F	TASA	BNC	K-F	TASA	BNC	K-F	TASA	BNC
Effect									
Covariates	.76^c	.76^c	.76^c	37.24^c	37.24^c	37.24^c	32.04^c	32.04^c	32.04^c
log-WF	22.67^c	32.42^c	27.80^c	8.66^c	12.66^c	12.17^c	14.87^c	19.66^c	20.14^c
log-CD	34.63^c	32.66^c	32.52^c	9.07^c	12.90^c	13.12^c	15.53^c	20.03^c	21.05^c
log-WF unique	*.12*^{a}	.02	*.48*^{c}	*.00*	*.00*†	*.03*^{c}	*.00*	*.01*^{a}	.00†
log-CD unique	1.76^c	.26^c	5.20^c	.41^c	.24^c	.98^c	.66^c	.38^c	.91^c

Numbers in italics refer to improvement in prediction from an *inhibitory* effect of word frequency. SB97 refers to Spieler and Balota (1997); BS98 to Balota and Spieler (1998); BCP99 to Balota et al. (1999); and Elexicon to Balota et al. (2000). K-F refers to Kučera and Francis (1967). TASA refers to Landauer et al. (1998) 12th grade. BNC refers to British National Corpus Consortium (2000). $^\dagger p < .1$; $^a p < .05$; $^b p < .01$; $^c p < .001$.

Table 2

Unique effects (ΔR^2 in %) in analyses of Word Frequency (log-WF), Contextual Diversity (log-CD) and Length from three corpora across six datasets.

Data		SB97 Young Naming			BS98 Older Naming			BCP99 Young LDT	
Corpus	K-F	TASA	BNC	K-F	TASA	BNC	K-F	TASA	BNC
Effect									
log-WF (after CD)	$.05^{\dagger}$	$.27^b$	$.27^b$	.03	$\dot{3}1^b$	.06	*$.21^b$*	*.01*	*$.45^c$*
log-CD (after WF)	.04	*.02*	$.12^{\dagger}$	$.31^b$	*.00*	$.79^c$	2.37^c	$.70^c$	5.85^c
log-WF (after CD, Length)	*.02*	*.01*	*.04*	*.05*	.01	*$.13^a$*	*$.24^b$*	*.01*	*$.55^c$*
log-CD (after WF, Length)	$.37^c$	$.16^a$	$.93^c$	$.79^c$	$.13^a$	1.97^c	2.44^c	$.65^c$	6.02^c
Length (after CD, WF)	12.56^c	11.35^c	12.65^c	8.82^c	7.36^c	9.08^c	$.08^a$	.00	$.07^b$

Data		BCP99 Older LDT			Elexicon Naming			Elexicon LDT	
Corpus	K-F	TASA	BNC	K-F	TASA	BNC	K-F	TASA	BNC
Effect									
log-WF (after CD)	*$.10^a$*	.02	*$.43^c$*	$.02^b$	$.02^b$	$.04^c$	$.05^c$	$.05^c$	$.26^c$
log-CD (after WF)	1.71^c	$.28^c$	5.16^c	$.41^c$	$.21^c$	$.73^c$	$.53^c$	$.21^c$	$.55^c$
log-WF (after CD, Length)	*$.13^a$*	.02	*$.54^c$*	*$.04^c$*	*$.07^c$*	*$.15^c$*	*$.01^b$*	*$.03^c$*	*.00*
log-CD (after WF, Length)	1.78^c	$.27^c$	5.37^c	$.63^c$	$.49^c$	1.43^c	$.78^c$	$.47^c$	1.14^c
Length (after CD, WF)	$.11^a$	*.00*	$.21^b$	21.36^c	13.15^c	19.49^c	20.36^c	11.57^c	17.94^c

Notation as in Table 1.

Table 3

Unique effects of CD, WF and semantic variables, after covariates, Elexicon data for which semantic variables are available.

	Data	Elexicon Naming			Elexicon LDT		
	Corpus	K-F	TASA	BNC	K-F	TASA	BNC
Effect							
Concreteness		*.60*^{c}	*.77*^{c}	*.21*^{b}	*.74*^{c}	*1.14*^{c}	*.21*^{b}
Imagery		3.10^{c}	1.77^{c}	1.70^{c}	3.18^{c}	1.48^{c}	1.52^{c}
Ambiguity		.14^{a}	.12^{a}	.08†	.30^{c}	.28^{c}	.12^{a}
log-WF		*.00*	*.01*	*.02*	.01	*.04*	.00
log-CD		.39^{c}	.24^{b}	.86^{c}	.50^{c}	.14^{a}	1.14^{c}
Total		59.11^{c}	60.24^{c}	60.68^{c}	59.63^{c}	61.90^{c}	62.65^{c}

Notation as in Table 1.

Table 4

Overall (after covariates) and unique effects (ΔR^2 in %) of rank Word Frequency (rank-WF) and rank Contextual Diversity (rank-CD) for dataset-corpus combinations where rank-WF performed better than log-WF.

Data	SB97 Young Naming	BS98 Older Naming	BCP99 Young LDT
Corpus	K-F	K-F	K-F
Effect			
Covariates	46.56^c	27.20^c	1.16^c
rank-WF	5.04^c	10.94^c	29.24^c
rank-CD	5.09^c	11.34^c	30.77^c
rank-WF unique	$.09^a$	$.09^a$	$.12^a$
rank-CD unique	$.14^b$	$.49^c$	1.65^c

Data	BCP99 Older LDT		Elexicon Naming		Elexicon LDT
Corpus	K-F	TASA	K-F	TASA	K-F
Effect					
Covariates	$.94^c$	$.66^c$	38.71^c	36.17^c	34.32^c
rank-WF	26.07^c	31.66^c	8.55^c	12.10^c	13.37^c
rank-CD	27.86^c	32.85^c	8.85^c	12.38^c	13.71^c
rank-WF unique	.04	*.05*	$.10^c$	$.02^b$	$.19^c$
rank-CD unique	1.85^c	1.38^c	$.40^c$	$.30^c$	$.53^c$

Notation as in Table 1.

Table 5

Overall (after covariates) and unique effects (ΔR^2 in %) of power-law-transformed Word Frequency (pow-WF) and power-law-transformed Contextual Diversity (pow-CD), and log-transformed Adjusted Word Frequency (log-U) and log-transformed Contextual Diversity (log-CD) from three corpora across six datasets.

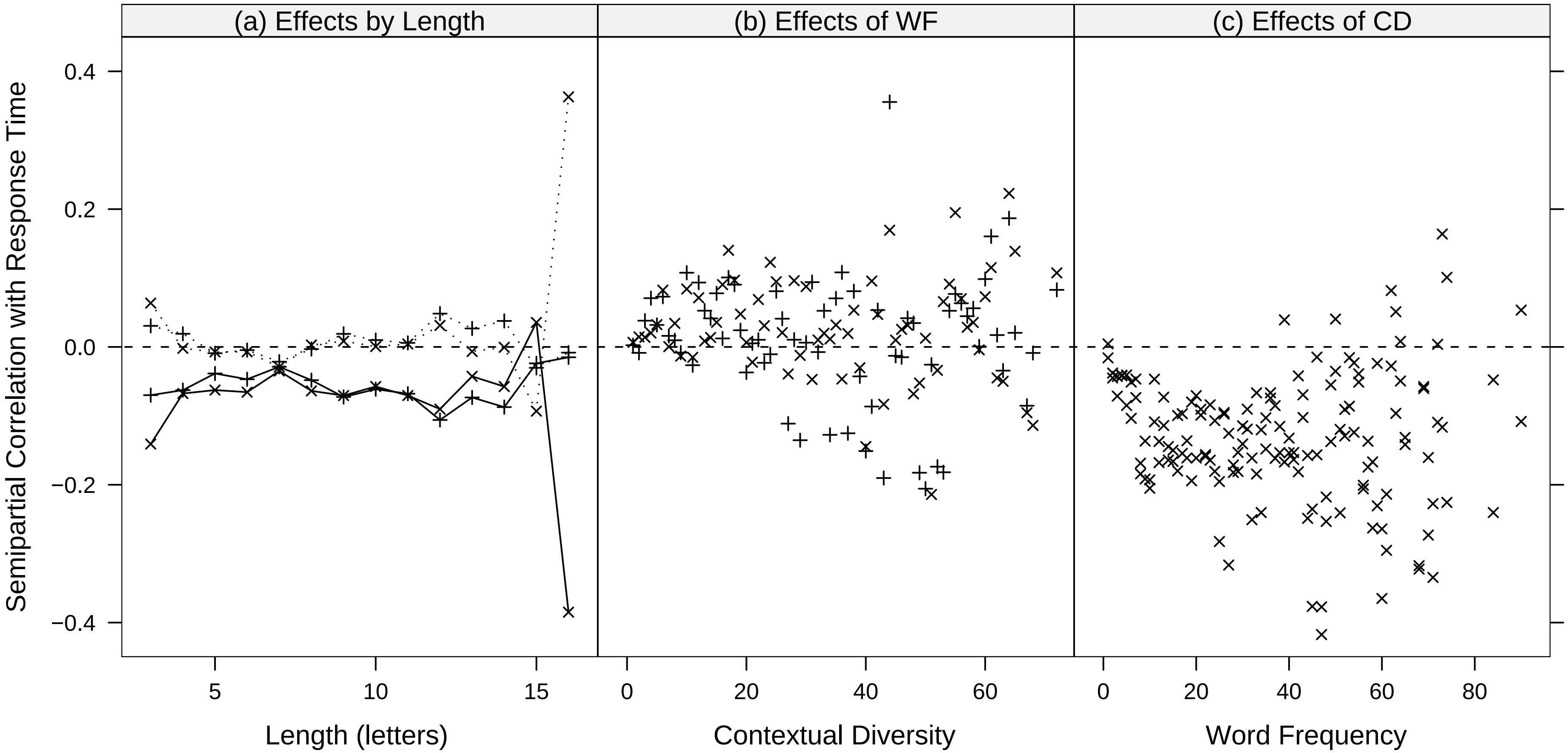
Data	SB97 Young Naming			BS98 Older Naming			BCP99 Young LDT		
Corpus	K-F	TASA	BNC	K-F	TASA	BNC	K-F	TASA	BNC
Effect									
pow-WF	5.89^c	7.16^c	6.33^c	11.70^c	14.50^c	12.31^c	33.12^c	41.35^c	35.21^c
pow-CD	5.67^c	7.29^c	6.74^c	11.93^c	14.76^c	13.29^c	35.46^c	42.49^c	38.88^c
pow-WF unique	*.21*^{b}	*.13*^{a}	*.16*^{b}	*.60*^{c}	*.19*^{b}	*.24*^{c}	*.40*^{c}	*2.26*^{c}	*.92*^{c}
pow-CD unique	.19^b	.25^c	.57^c	.82^c	.45^c	1.23^c	1.75^c	3.40^c	3.59^c
log-U	4.84^c	6.63^c	6.03^c	9.71^c	13.37^c	11.92^c	26.18^c	38.12^c	33.71^c
log-U unique	*.08*^{a}	.00	*.04*	*.26*^{c}	.04	*.08*†	*1.32*^{c}	.00	*.28*^{c}
log-CD unique	.59^c	.19^c	.73^c	1.45^c	.49^c	1.45^c	5.19^c	.65^c	4.41^c

Data	BCP99 Older LDT			Elexicon Naming			Elexicon LDT		
Corpus	K-F	TASA	BNC	K-F	TASA	BNC	K-F	TASA	BNC
Effect									
pow-WF	28.04^c	37.20^c	31.33^c	9.78^c	14.27^c	12.27^c	16.20^c	20.74^c	20.21^c
pow-CD	29.22^c	38.16^c	33.16^c	9.99^c	14.61^c	13.12^c	16.56^c	21.20^c	21.05^c
pow-WF unique	*1.31*^{c}	*.69*^{c}	*.18*^{a}	*.35*^{c}	*.34*^{c}	*.25*^{c}	*.37*^{c}	*.20*^{c}	*.24*^{c}
pow-CD unique	2.48^c	1.65^c	2.01^c	.58^c	.50^c	1.10^c	.73^c	.66^c	1.08^c
log-U	20.92^c	31.11^c	28.58^c	8.66^c	12.66^c	12.17^c	14.87^c	19.66^c	20.14^c
log-U unique	*1.71*^{c}	*.37*^{c}	*.40*^{c}	*.00*	*.00*†	*.03*^{c}	*.00*	*.01*^{a}	.00†
log-CD unique	5.41^c	1.93^c	4.35^c	.41^c	.24^c	.98^c	.66^c	.38^c	.91^c

Notation as in Table 1. Power-law and U analyses are conducted separately. Power-law analyses are after covariates (as presented in Table 1). U analyses are after covariates, and unique log-U effects are after unique log-CD and covariates (as in Table 1).

Figure Captions

Figure 1. Semipartial Correlations of TASA log-WF and log-CD with Response Times in Word Naming (+) and Lexical Decision (×), Elexicon Data, with orthographic N, number of syllables, and onset (for naming only) partialled. Points based on fewer than 50 words are omitted. (a) For each length, effects of WF after CD and covariates (dotted lines), and effects of CD after WF and covariates (solid lines). (b) For each value of CD, effects of WF after covariates. (c) For each value of WF, effects of CD after covariates.



(a) Effects by Length
(b) Effects of WF
(c) Effects of CD
Semipartial Correlation with Response Time
0.4
0.2
0.0
−0.2
−0.4
5
10
15
Length (letters)
0
20
40
60
Contextual Diversity
0
20
40
60
80
Word Frequency