The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kučera and Francis

CURT BURGESS and KAY LIVESAY
University of California, Riverside, California

Word frequency is one of the strongest determiners of reaction time (RT) in word recognition tasks; it is an important theoretical and methodological variable. The Kučera and Francis (1967) word frequency count (derived from the 1-million-word Brown corpus) is used by most investigators concerned with the issue of word frequency. Word frequency estimates from the Brown corpus were compared with those from a 131-million-word corpus (the HAL corpus; conversational text gathered from Usenet) in a standard word naming task with 32 subjects. RT was predicted equally well by both corpora for high-frequency words, but the larger corpus provided better predictors for low- and medium-frequency words. Furthermore, the larger corpus provides estimates for 97,261 lexical items; the smaller corpus, for 50,406 items.

Word frequency (WF) has been referred to as the sine qua non among variables that affect basic word recognition (Chiarello, 1988). All models of lexical access incorporate word frequency into their retrieval mechanisms (frequency-coded models, Forster, 1976; semantic feature models, Smith, Shoben, & Rips, 1974; Logogen models, Morton, 1969; connectionist models, McClelland & Rumelhart, 1985; Monsell, 1991; Plaut, 1996; high-dimensional semantic space models, Lund & Burgess, 1996). Although these models differ with respect to the specific implementation of the WF mechanism, the basic mechanism is that WF sets a threshold for word recognition: WF is a measure of experience with a word which has some direct relationship to its availability in memory.

The notion that one's familiarity with a word is related to response time goes back at least to Cattell (1886). This result, as well as other reaction time (RT) results that he collected, were "important enough to prove those to be wrong who with Kant hold that psychology can never become an exact science" (p. 63). Thus it follows that one of the most basic findings in the memory and psycholinguistic literature is that high-frequency (HF) words are responded to more quickly than low-frequency (LF) words (Dobbs, Friedman, & Lloyd, 1985; Gernsbacher, 1984). For some models, WF plays a more complex role in the lexical access process. Grainger, O'Regan, Jacobs, and Segui (1989) found that LF orthographic neighborhoods facilitate lexical access, whereas HF neighborhoods inhibit access (cf., Sears, Hino, & Lupker, 1995). At a more conceptual level, WF, to the extent that it is correlated with contextual diversity, can be a major determinant in the stability of conceptual representations (Lund & Burgess, 1996; Schwanenflugel & Shoben, 1985). When a word has been first experienced, its representation is essentially episodic. As it is experienced more, and in different contexts, its meaning becomes more of a generalized representation. Typically, WF is correlated with the diversity of contexts in which a word is found, thus making WF estimates an important component in semantic models (Burgess, 1998). Not surprisingly, then, it is a crucial factor in a word's memorability, both in recognition and in recall (Clark & Burchett, 1994).

There is also a relationship between WF and word difficulty. Lower frequency words tend to be more difficult to understand (Breland, 1996; Rudell, 1993). Although this is a robust relationship, a number of caveats suggest important implications for the use of WF estimates in making these conclusions. Word counts for LF words tend to be somewhat unreliable. WF counts are usually made from published literature, and Rudell has suggested that such text is stylized and has been revised so that common, but more conversational, language has been edited out. Furthermore, the corpus size may not adequately sample the language for reliable counts of LF words. This hypothesis is supported by the results of a comparison of two corpora made by Graf and Williams (1987). They had subjects complete three-letter word fragments. Word fragment completion correlated ($r = .30$) with the WF counts from Kučera and Francis (1967). However, word fragment completion correlated more strongly ($r = .48$) with the WF counts from the Thorndike and Lorge (1944) norms. The Thorndike and Lorge norms, although several decades older, are derived from a corpus about 18 times larger than the Brown corpus.

Table 1
Number of Articles Citing Word Frequency Norms for the Years 1995 and 1996

Journal	Total Articles	Articles Citing Frequency Norms	Articles Citing Non-English Norms
Journal of Psycholinguistic Research	60	11	1
Journal of Experimental Psychology:			
* Learning, Memory, and Cognition*	216	55	6
Memory & Cognition	127	21	2
Journal of Memory and Language	82	11	6
Total	485	98	15

The size of the corpus would seem to be one important determinant in the effective sampling of word use. This point is substantiated by recent comparison of four corpora and their respective WF counts and correlations with word difficulty (Breland, 1996). Breland found that certain kinds of words are particularly problematic in estimating WF or word difficulty. In the difficulty ratings that he used (those of Dupuy, 1974), only one meaning of a homograph was counted as correct. Responses that may have been related to other meanings were not counted. He also found a number of words for which there were difficulty ratings, but that did not occur in the Brown corpus (e.g., *tricycle*). When the correlations between WF and word difficulty were computed for all four corpora, with and without these problem words, the largest change in the correlation was for the Brown corpus ($-.72$ to $-.81$). In addition, Lovelace (1988) has warned that LF counts may be problematic in the Brown corpus because the samples from each genre of text were relatively small (2,000 words). We will address this issue later in this paper.

The role of WF is not constrained to simple word recognition experiments. Basic memory processes, including WF, have been taking on an increasingly important role in psycholinguistics as higher level models of sentence comprehension become more lexically based. In theories of syntactic processing, the role played by the frequency of occurrence of words in the context of other words has become increasingly important (Burgess & Hollbach, 1988; Burgess & Lund, 1994; Burgess, Tanenhaus, & Hoffman, 1994; MacDonald, 1994; MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Carlson, 1989; Trueswell, Tanenhaus, & Garnsey, 1994). WF has also been implicated in the attempt to distinguish among neuropsychological disorders (Jurado, Junque, Pujol, Oliver, & Vendrell, 1997) and in reading disorders (Hyönä, & Olson, 1995; Troia, Roth, & Yeni-Komshian, 1996).

We recently checked for the incidence of WF norm usage in four journals (*Journal of Experimental Psychology: Learning, Memory, and Cognition, Journal of Memory and Language, Memory & Cognition, and Journal of Psycholinguistic Research*) for 1995–1996. We found that WF norms were used in approximately 20% of all articles (98 of 485 over a 2-year period). The Brown corpus (as either the Kučera and Francis, 1967, or the Francis & Kučera, 1982, word count) was used most often. The Thorndike and Lorge (1944) word count was also used.

(See Table 1.) These results are consistent with Lovelace's (1988) analysis of similar journals for 1985–1986.

THE TALE OF TWO CORPORA[1]

The Brown Corpus

The Kučera and Francis (1967; hereafter KF) WF count is derived from a corpus of American English compiled at Brown University. The published WF count is that published by Kučera and Francis in book form. The corpus (referred to as the Brown corpus) and the WF count are also available from the publisher on computer media. The Brown corpus consists of 1,014,000 words of text in 500 samples of roughly 2,000 words each. A more recent version of this book (Francis & Kučera, 1982; hereafter FK) is also based on the Brown corpus. Both versions contain two frequency lists: an alphabetical list and a rank-ordered list. In the KF version, the alphabetical and rank-ordered lists include all 50,406 word entries. A slightly more limited approach to listing items was taken in the FK version, owing to the additional room required by the listing of the different grammatical classes separately. Approximately 36,000 items are listed in the alphabetical list. The rank-ordered table lists 5,996 words, all of them items that reached a certain frequency threshold. A very useful feature of the more recent version is that the lists are tagged by grammatical class. In the experiments reported in this paper, we used Form C of the machine-readable version of the FK WF count.

The HAL Corpus

The HAL corpus consists of approximately 131 million words gathered during February of 1995 from approximately 3,000 Usenet newsgroups. All Usenet newsgroups that contained text contributed to the corpus. At the time at which the corpus was collected, about 10 million new words of text were available each day. A recent check of the Usenet feed that we use indicated that about 20 million words were available each day. Thus it is an easily available source of text. A major advantage of this source is the extremely broad range of topics that are discussed. With 3,000 newsgroups virtually any topic is covered. The text is very conversational and noisy, much like spoken language. These features overcome what we saw as two limitations to the Brown corpus: small samples of words and more formal language use. One limitation of the

HAL corpus is that it is not yet tagged for parts of speech. Of course, any corpus has limitations and advantages, and these are a function of the question under investigation. The master vocabulary for the HAL corpus contains 3,461,884 lexical entries. This represents a huge number of types, which requires some comment. It is difficult to characterize the complete set of items. Of the 70,000 most frequent items, about half have entries in the standard Unix dictionary and the other half are proper names, slang words, misspellings, and nonword symbols. The remaining 3,391,884 items represent a vast range of LF words, misspellings, hyphenations, nonletter characters, and other nonword symbols. Although for any given user, much of this list would be considered noise, we have found it useful in other projects to have frequency counts for misspellings; emoticons, such as ":)" or ":>("; and slang. As a result, it becomes difficult to clearly state the number of words in the count. For the present purposes, we will claim that there are 97,261 words, but this number is actually larger. The overhead of maintaining the frequency file is not substantial (21.3 MB).

The genesis of the HAL corpus requires a brief introduction. The HAL (Hyperspace Analogue to Language) model of memory encodes meaning by transducing lexical co-occurrence information into semantic and grammatical information. This corpus was collected as the language input for the HAL model. Since the model was to simulate normal human memory, a conversational and widely ranging source of language was desired.

The basic methodological details for HAL are available (Lund & Burgess, 1996). The model has been used to simulate basic word recognition and priming processes (Burgess & Lund, 1997c; Livesay & Burgess, 1997; Lund, Burgess, & Atchley, 1995; Lund, Burgess, & Audet, 1996), as well as neuropsychological memory processing in normals (Burgess & Lund, 1997a) and in deep dyslexics (Buchanan, Burgess, & Lund, 1996). The model has also been used successfully in the simulation of human grammatical and syntactic behavior (Burgess & Lund, 1997b) and developmental data (Burgess, Lund, & Kromsky, 1997; Lund & Burgess, 1997). An important part of every one of these experiments is the notion that WF is related to how a concept develops. Implicit in the success of these experiments is that the incidence of the words in the corpus needs to correspond to ordinary human experience. The experiments in the present paper were performed explicitly to test the hypothesis that the larger, more recent, and conversational corpus will predict human performance on a simple lexical task better than will the older, smaller corpus.

Thus, given the importance of WF as a theoretical construct in models of word recognition as well as other cognitive processes, and given the prevalence of usage of the KF word count, we decided to investigate the correspondence between the KF frequency estimates and a larger, more recent WF count. WF is most often used to segregate stimuli into various frequency ranges (e.g., low,

high), to balance sets of items within or between conditions of an experiment, or to set parameters in computer simulations of word retrieval or reading. Inherent in many of these experiments is some kind of evaluation of the role of LF information. An adequate number of LF items is crucial in a corpus because inadequate sampling is most likely with LF items and thus may introduce considerable unwanted variability into an LF condition.

EXPERIMENT 1
Corpus Correlations

Breland (1996) and Lovelace (1988) have both suggested that a larger corpus would increase the likelihood of more reliable frequency counts particularly for LF words. This experiment was designed to compare the Brown corpus of approximately 1 million words with the training input for the original HAL memory model (see Lund & Burgess, 1996) of approximately 131 million words. If these two corpora correlate well for low, medium, and high frequency ranges, it is unlikely that the larger corpus would actually predict human RT results better.

Method

A set of 8,208 words (all nouns tagged NN) was extracted from Form C of the machine-readable FK word count. Each word had two frequency values associated with it, one based on the Brown corpus and the other based on the HAL corpus. These two sets of frequency values were then correlated. There were four comparisons. The first comparison was over the entire frequency range. The other three comparisons were for low (≤ 10), medium (35–75), and high (≥ 100) frequency ranges, based on frequencies from KF. Unless otherwise specified (i.e., Experiment 2B), WF ranges are based on KF, since this is how most investigators have operationalized WF. These operational definitions of LF, medium frequency (MF) and HF were based on our review of the literature of the common usage of these WF ranges in empirical studies done with the KF frequency counts.

Results and Discussion

The overall correlation between the two corpora was $r = .96$. However, when WF was separated into frequency-based categories, there was a lack of correspondence of the frequency estimates from the two corpora. Although the two corpora correlated well for HF words ($r = .96, p < .0001$; 489 words), the correlations for MF and LF words were considerable weaker (MF, $r = .12, p < .0001$, 713 words; LF, $r = .14, p < .0001$, 3,720 words). The lack of correspondence between the frequency estimates from the two corpora for the LF and MF words motivated Experiment 2.

EXPERIMENT 2
Using WF Counts to Predict Lexical Decision RT

It was clear from Experiment 1 that the Brown corpus and the HAL corpus have differing WF estimates for LF and MF words. Obviously the two corpora would have differing WF counts; one corpus is approximately 131

Table 2
Correlations Between Log Frequency of Words and Reaction Time

	Levels of Frequency	KF Log Frequency Correlated With RT	HAL Log Frequency Correlated With RT
	all	−.52	−.58
Brown corpus categorization	Low (≤10)	−.35	−.43
	Medium (35–75)	−.11	−.38
	High (≥100)	−.42	−.48
HAL corpus categorization	Low (≤1,500)	−.28	−.42
	Medium (5,250–11,250)	.13	−.29
	High (≥15,000)	−.23	−.35

times larger than the other. However, the WFs should have the same relative relationship so that the WF counts can make similar behavioral predictions. Cognitive psychologists are concerned with how WF counts can predict behavior (RTs for lexical decision, naming, word difficulty estimates). The lack of a correlation between the WF estimates of the two corpora for the LF and MF items leaves unclear which corpus is the best predictor of behavior. In this experiment, we evaluate the extent to which WF counts from the Brown and HAL corpora predict RT in a word-naming task.

EXPERIMENT 2A
Brown Corpus Categorization

Method

Subjects. Thirty-two undergraduates participated as part of a class requirement. All subjects were right-handed native speakers of English, with normal or corrected-to-normal vision.

Stimuli. All common nouns shared by the KF corpus and the HAL corpus were extracted for initial consideration. The items were limited to 3 to 10 letters in length with no hyphenations. The experimenters inspected this subset of words and discarded all words that had an obvious verb or adjective use or were considered jargon words (e.g., *frame, fat, byte, bit*). In keeping with how most investigators categorize words by their frequency, the remaining words were grouped into HF (≥100), MF (35–75), and LF (≤10) sets, according to the KF frequency counts. In addition, HAL corpus frequencies were extracted for these words. The experimenters then randomly chose 10 items of each length for each frequency set, resulting in a set of 240 nouns. A 240-item test list was constructed by quasi-randomly placing each word within the list. No semantically related or associated words appeared in close proximity. Twenty practice trials of mixed frequency and word length were presented prior to the experimental list.

Procedure. The stimuli were presented on a computer monitor; the subjects' responses were collected via a microphone connected to the computer by a Digitry CTS system. Each trial began with a central fixation point presented for 500 msec, immediately followed by the word in the same location. The word would remain on the screen until a subject responded or until 2,000 msec had elapsed. The subject's task was to read the word aloud as quickly as possible. If no response was made within 2,000 msec, the subjects heard a 500-msec beep, and the word "timeout" appeared on the screen for 1,000 msec. The intertrial interval was 1,000 msec. The experimenter stayed with the subject during the list presentation to ensure the subject's sustained attention.

Results and Discussion

Overall mean RT was 618 msec. Mean RT was 584 msec for HF words, 602 msec for MF words, and 670 msec for

LF words. Error rates were not recorded, because they tend to be negligible in a simple naming task (usually <1.5%). Table 2 contains correlations between log frequency (for both KF frequencies and HAL frequencies) and RT. The columns in Table 2 correspond to the correlation of either the KF or the HAL frequencies to word naming latency. Recall that the level of frequency (HF, MF, LF) was derived from how other investigators had segregated frequency by using the KF count (HF ≥100, MF 35–75, LF ≤10); these are labeled as "Brown corpus categorization" (top half of Table 2).

Any correlation >.20 (absolute value) was reliable at $\alpha = .05$. Only the log frequencies in Table 2 will be discussed, since log frequency is a better predictor of word recognition RT, although the raw frequency data showed a parallel set of results. Including all frequency ranges, both KF and HAL corpora WF counts produced strong correlations with RT ($rs > -.50$). There were negative correlations between KF log frequency and RT for both LF ($-.35$) and HF ($-.42$) items. In both cases, RT decreased as WF increased. This typical relationship was unreliable with the MF items ($-.11$) using the KF frequencies. HAL corpus log frequencies were better predictors of RT for all frequency levels than were the KF counts (HF, $r = -.48$; MF, $r = -.38$; LF, $r = -.43$), and substantially so with MF items.

Use of the Brown corpus categorization is consistent with the procedure used by most investigators. However, a corpus much larger in size and more recently collected may provide a better measure of categorizing WF, particularly for LF and MF items, and this was the subject of Experiment 2B.

EXPERIMENT 2B
HAL Corpus Categorization

Of course, not all words that are LF via the KF count are LF in the HAL corpus. All correlations using the HAL frequencies predicted RT better than did the correlations from the Brown corpus. This raised the concern that using KF to categorize words by frequency in the first place might be problematic. For example, in KF the words *math* and *den* are LF words. However, in the HAL corpus, they are MF words. As a result, we also analyzed the RT results from the perspective of the HAL corpus (i.e., LF items in the HAL corpus that might not be LF in the KF corpus). In Table 2, this row is labeled "HAL corpus

Table 3
Examples of Words not in the KF Corpus
but Present in the HAL Corpus

ark	aquarium
cashier	cucumber
curfew	diploma
fossil	lettuce
noodle	oboe
reptile	tangerine
wheelchair	yam

categorization" (bottom half of Table 2). We thought it important to evaluate the results both ways.

Method

The RT analysis for this experiment is a reanalysis of the RT data used in Experiment 2A. In this analysis, WF level was defined by roughly scaling the KF frequency ranges to their corresponding ranges in the larger HAL corpus. For example, LF items were defined as having WF in the HAL corpus of $\leq 1,500$ (equivalent to KF ≤ 10). Given that the HAL corpus is larger and had been obtained more recently, we thought that this method of categorization would provide a better division of items into levels of frequency of occurrence for contemporary language use.

Results and Discussion

Although a similar pattern of results was found, this categorization showed a larger difference between the correlations between the KF and HAL corpus frequencies; the HAL corpus counts provided the better predictors of RT. There were reliable correlations between KF log frequency and RT for both LF items ($-.28$) and HF items ($-.23$); RT decreased as WF increased. This typical relationship was unreliable with the MF items ($+.13$) using KF log frequency, and in fact went in the wrong direction. Once again, HAL corpus log frequencies were better predictors of RT for all frequency levels (HF, $r = -.35$; MF, $r = -.29$; LF, $r = -.42$).

GENERAL DISCUSSION

We find that the HAL corpus estimates provide better predictors of RT for all WF ranges. The HAL corpus categorization resulted in a larger performance difference between the two corpora as expected, although the correlations are not as large as in the KF corpus categorization. This is likely due to a decrease in power: Fewer items contributed to these correlations, since the items were originally selected as a function of fitting into the operational definition of KF frequency levels. Even so, it is clear that the larger, more recent corpus provides better predictors for human RT at least in the word naming task.

On the one hand, the smaller KF word count provides reasonable predictors for RT (except for the MF items). However, we think that the degree to which the HAL corpus provides better predictors of human RT is actually underestimated by these results. First, fewer items went into the HAL corpus categorization, and had the power been equal, we suspect that the correlations would have

been larger. Second, the HAL corpus lexicon contains 97,261 lexical items, whereas the KF lexicon contains 50,406 items. These correlations were all computed for items that were contained in both corpora. A corpus that is 131 times larger than the Brown corpus will contain many items that will likely not be in the Brown corpus (a problem discussed by Breland, 1996; Lovelace, 1988; Rudell, 1993). Table 3 shows a few items in the HAL corpus, but not in the Brown corpus. One method of dealing with this issue is to enter a count of one for an item that is not in the KF corpus. However, there has to be an element of noise in this procedure. Our analysis suggests that this noise may be considerable because of the number of familiar words that are not in the Brown corpus.

The results of the present experiments suggest that a larger, more recently collected corpus is preferable for making estimates of WF for predicting human RT. Lower frequency words are more likely to have been representatively sampled. Furthermore, a considerably wider sample of lexical items is represented in the HAL corpus, making the lexicon more useable. The variance accounted for by WF is considerable in RT to a word, whether presented alone, in a priming experiment, or in a natural reading paradigm. Making the best possible estimate of WF would seem to be crucial and is best accomplished with the use of larger, more representative language samples.[2]

REFERENCES

BRELAND, H. M. (1996). Word frequency and word difficulty: A comparison of counts on four corpora. *Psychological Science*, 7, 96-99.

BUCHANAN, L., BURGESS, C., & LUND, K. (1996). Overcrowding in semantic neighborhoods: Modeling deep dyslexia. *Brain & Cognition*, 32, 111-114.

BURGESS, C. (1998). *From simple associations to the building blocks of language: Modeling meaning in memory with the HAL model*. Manuscript submitted for publication.

BURGESS, C., & HOLLBACH, S. C. (1988). A computational model of syntactic ambiguity as a lexical process. In *Proceedings of the Tenth Annual Cognitive Science Society Meeting* (pp. 263-269). Hillsdale, NJ: Erlbaum.

BURGESS, C., & LUND, K. (1994). Multiple constraints in syntactic ambiguity resolution: A connectionist account of psycholinguistic data. In *Proceedings of the Cognitive Science Society* (pp. 90-95). Hillsdale, NJ: Erlbaum.

BURGESS, C., & LUND, K. (1997a). Modeling cerebral asymmetries of semantic memory using high-dimensional semantic space. In M. Beeman & C. Chiarello (Eds.), *Getting it right: The cognitive neuroscience of right hemisphere language comprehension* (pp. 215-244). Hillsdale, NJ: Erlbaum.

BURGESS, C., & LUND, K. (1997b). Modeling parsing constraints with high-dimensional context space. *Language & Cognitive Processes*, 12, 177-210.

BURGESS, C., & LUND, K. (1997c). Representing abstract words and emotional connotation in high-dimensional memory space. In *Proceedings of the Cognitive Science Society* (pp. 61-66). Hillsdale, NJ: Erlbaum.

BURGESS, C., LUND, K., & KROMSKY, A. (1997). Examining issues in developmental psycholinguistics with a high-dimensional memory model. *Abstracts of the Psychonomic Society*, 2, 66.

BURGESS, C., TANENHAUS, M. K., & HOFFMAN, M. (1994). Parafoveal and semantic effects on syntactic ambiguity resolution. In *Proceedings of the Cognitive Science Society* (pp. 96-99). Hillsdale, NJ: Erlbaum.

CATTELL, J. M. (1886). The time it takes to see and name objects. *Mind,* **11,** 63-65.

CHIARELLO, C. (1988). Lateralization of lexical processes in the brain: A review of visual half-field research. In H. A. Whitaker (Ed.), *Contemporary reviews in neuropsychology* (pp. 36-76). New York: Springer-Verlag.

CLARK, S. E., & BURCHETT, R. E. R. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition,* **22,** 55-62.

DOBBS, A. R., FRIEDMAN, A., & LLOYD, J. (1985). Frequency effects in lexical decisions: A test of the verification model. *Journal of Experimental Psychology: Human Perception & Performance,* **11,** 81-92.

DUPUY, H. J. (1974). The rationale, development, and standardization of a basic word vocabulary test. *Vital & Health Statistics,* **2,** 71.

FORSTER, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. C. T. Walker (Eds.), *New approaches to language mechanisms* (pp. 257-287). Amsterdam: North-Holland.

FRANCIS, W. N., & KUČERA, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin.

GERNSBACHER, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General,* **113,** 256-281.

GRAF, P., & WILLIAMS, D. (1987). Completion norms for 40 three-letter word stems. *Behavior Research Methods, Instruments, & Computers,* **19,** 422-445.

GRAINGER, J., O'REGAN, J. K., JACOBS, A. M., & SEGUI, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & Psychophysics,* **45,** 189-195.

HYÖNÄ, J., & OLSON, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **21,** 1430-1440.

JURADO, M. A., JUNQUE, C., PUJOL, J., OLIVER, B., & VENDRELL, P. (1997). Impaired estimation of word occurrence frequency in frontal lobe patients. *Neuropsychologia,* **35,** 635-641.

KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

LIVESAY, K., & BURGESS, C. (1997). Mediated priming: A representational and empirical account using the HAL model. In *Proceedings of the Cognitive Science Society* (pp. 436-441). Hillsdale, NJ: Erlbaum.

LOVELACE, E. A. (1988). On using norms for low-frequency words. *Bulletin of the Psychonomic Society,* **26,** 410-412.

LUND, K., & BURGESS, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers,* **28,** 203-208.

LUND, K., & BURGESS, C. (1997, December). *Recurrent neural networks and global co-occurrence models: Developing contextual representations of word meaning.* Paper presented at the NIPS*97 (Neural Information Processing Systems) Neural Models of Concept Learning post-conference workshop, Breckenridge, CO.

LUND, K., BURGESS, C., & ATCHLEY, R. A. (1995). Semantic and associative priming in high-dimensional semantic space. In *Proceedings of the Cognitive Science Society* (pp. 660-665). Hillsdale, NJ: Erlbaum.

LUND, K., BURGESS, C., & AUDET, C. (1996). Dissociating semantic and associative word relationships using high-dimensional semantic space. In *Proceedings of the Cognitive Science Society* (pp. 603-608). Hillsdale, NJ: Erlbaum.

MACDONALD, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language & Cognitive Processes,* **9,** 157-201.

MACDONALD, M. C., PEARLMUTTER, N. J., & SEIDENBERG, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review,* **101,** 676-703.

MCCLELLAND, J. L., & RUMELHART, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General,* **114,** 159-188.

MONSELL, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148-197). Hillsdale, NJ: Erlbaum.

MORTON, J. (1969). Interaction of information in word recognition. *Psychological Review,* **76,** 165-178.

PLAUT, D. C. (1996). Relearning after damage in connectionist networks: Toward a theory of rehabilitation. *Brain & Language,* **52,** 25-82.

RUDELL, A. P. (1993). Frequency of word usage and perceived word difficulty: Ratings of Kučera and Francis words. *Behavior Research Methods, Instruments, & Computers,* **25,** 455-463.

SCHWANENFLUGEL, P., & SHOBEN, E. (1985). The influence of sentence constraints on the scope of facilitation for upcoming words. *Journal of Memory & Language,* **24,** 232-252.

SEARS, C. R., HINO, Y., & LUPKER, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception & Performance,* **21,** 876-900.

SMITH, E. E., SHOBEN, E. J., & RIPS, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review,* **81,** 214-241.

TANENHAUS, M. K., & CARLSON, G. N. (1989). Lexical structure and language comprehension. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 529-561). Cambridge, MA: MIT Press.

THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words.* New York: Columbia University, Teachers College Press.

TROIA, G. A., ROTH, F. P., & YENI-KOMSHIAN, G. H. (1996). Word frequency and age effects in normally developing children's phonological processing. *Journal of Speech & Hearing Research,* **39,** 1099-1108.

TRUESWELL, J. C., TANENHAUS, M. K., & GARNSEY, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory & Language,* **33,** 285-318.

NOTES

1. There are other large scale corpora available: CELEX (http://www.kun.nl/celex) and many corpora at the Linguistic Data Consortium (http://www.ldc.upenn.edu).

2. We are in the process of making the HAL WF count, the corpus, and software available in CD form (see http://HAL.ucr.edu/WF.html for further information when announced).