Behavior Research Methods
2005, 37 (1), 65-70

N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics

COLIN J. DAVIS
Macquarie University, Sydney, New South Wales, Australia

This article describes a Windows program that enables users to obtain a broad range of statistics concerning the properties of word and nonword stimuli, including measures of word frequency, orthographic similarity, orthographic and phonological structure, age of acquisition, and imageability. It is designed for use by researchers in psycholinguistics, particularly those concerned with recognition of isolated words. The program computes measures of orthographic similarity on line, either with respect to a default vocabulary of 30,605 words or to a vocabulary specified by the user. In addition to providing standard orthographic neighborhood measures, the program can be used to obtain information about other forms of orthographic similarity, such as transposed-letter similarity and embedded-word similarity. It is available, free of charge, from the following Web site: http://www.maccs.mq.edu.au/~colin/N-Watch/.

In the design of stimuli for psycholinguistic experiments, it is standard practice to choose items that differ with respect to the factor(s) of interest but are closely matched with respect to other dimensions that may affect the dependent variable, such as word frequency, length, number of syllables, neighborhood size, spelling–sound regularity, bigram frequency, and so on. The N-Watch program described in this article provides a simple tool for obtaining a broad range of lexical statistics, including measures of word frequency, orthographic similarity, orthographic and phonological structure, age of acquisition (AOA), and imageability. The program runs on Windows PCs (preferably with at least 64 MB of RAM), and the full package (including data files) requires approximately 10 MB of hard disk space. It is available, free of charge, from the following Web site: http://www.maccs.mq.edu.au/~colin/N-Watch/. One noteworthy advantage of the program is that it computes statistics like neighborhood size and related measures of orthographic similarity on line, making it possible to obtain similarity statistics pertaining to a vocabulary specified by the user; these outputs can be obtained for both word and nonword inputs. In addition, the program is useful for obtaining information about other forms of orthographic similarity that affect performance in reading tasks, such as transposed-letter similarity and embedded-word similarity.

Selection of a Default Vocabulary

The size of the orthographic neighborhood of a letter string (N) is determined by counting the number of words that can be formed by substituting a different letter at any of the letter positions within the string (Coltheart, Davelaar, Jonasson, & Besner, 1977). In the computation of N, the number of words deemed to be orthographic neighbors of a given letter string is critically dependent on the reference vocabulary. For example, consider the word *calm*. How many orthographic neighbors does this word have? One obvious neighbor is the high-frequency word *call*, and the neighbors *palm* and *calf* are also not hard to generate; reasonably skilled readers will also be familiar with the word *balm*. So, a sensible value of the N statistic in this case would be 3 or 4. But when compiling measures of N, it is not feasible (or perhaps desirable) to ask large, representative samples of readers to interrogate their mental lexicons. Instead, it is necessary to consult a word list. Consulting a large corpus of written text, such as the COBUILD corpus (which forms the basis of the CELEX English Corpus Types, or ECT, corpus; Baayen, Piepenbrock, & van Rijn, 1995) may not give sensible results. In the present case, the corpus lists 13 orthographic neighbors of the word *calm* (in addition to *call*, *palm*, *calf*, and *balm*, there is *calc*, *cald*, *cale*, *cali*, *caln*, *calo*, *cals*, *cal-*, and *colm*). Some readers of this article may count a number of these items among the letter strings that they know, but it is probably uncontroversial to suggest that $N = 4$ is a better estimate of the neighborhood size of *calm* than $N = 13$. Measures of orthographic similarity that are based on overly comprehensive reference vocabularies are likely to systematically overestimate the latent variable of interest—that is, the number of words in readers' lexicons that are perceptually similar to the

---

Parts of this work were completed while the author was at the Department of Experimental Psychology, University of Bristol. The author thanks HarperCollins Publishers, Dave Balota, Max Coltheart, Simon Dennis, Sue Franklin, and David Howard for permission to include their previously collected lexical information within this software package. Thanks are also due to Sally Andrews, Jeff Bowers, and Hans Stadthagen. Correspondence concerning this article should be addressed to C. J. Davis, Department of Experimental Psychology, University of Bristol, 8 Woodland Rd, Bristol BS8 1TN, England (e-mail:colin.davis@bristol.ac.uk).

Note—This article was accepted by the previous editor, Jonathan Vaughan.

letter string in question. The problem is that many (probably the majority) of the 220,273 separate entries listed in the COBUILD corpus are unlikely to be in the lexicons of skilled readers; indeed, many entries are not even pronounceable (e.g., *0–40 mm*, *zzzzzzzrrrrr*).

A partial solution to this problem is to filter the raw corpus by cross-checking it against a reliable dictionary, eliminating any corpus entries that are not contained in the dictionary, such as misspellings, nonlexical abbreviations, and other linguistic oddities. However, many very-low-frequency items will remain, such as *cwm* and *groat*. If one's aim is to compute a measure of how many words known by a typical reader are orthographically similar to a given letter string, the presence of such rare words in the reference vocabulary is undesirable. This problem can be overcome by further filtering the corpus to exclude items that occur very rarely. This is not a perfect solution, because no particular cutoff will eliminate entries that are not typically known by skilled readers while simultaneously retaining all of the words that they do know. However, choosing a reasonably liberal frequency criterion can provide a satisfactory compromise by eliminating the most obscure words but not too many of the lower frequency words that will be known to individuals with large reading vocabularies. To this end, the default vocabulary for the N-Watch program was selected by cross-checking the raw CELEX ECT corpus against an electronic dictionary containing 65,013 words and then excluding entries that occurred fewer than seven times in the entire corpus of 17.9 million words (for this reason, the program's default vocabulary lists no words with total CELEX frequencies of less than 0.34 per million). In addition, words that were more than 10 letters long (or shorter than 2 letters) were excluded (because words of this length are not typically used in most psycholinguistic experiments). The resulting vocabulary comprises 30,605 words and is reasonably comprehensive, although some very-low-frequency words are not included (e.g., *cusp*), and some of the lowest frequency words that are included are probably not in most people's vocabularies (e.g., *abbe*). Because there is no such thing as an ideal vocabulary, the program is designed to make it easy for users to specify a custom vocabulary (as described below).

Specifying the Stimuli to be Analyzed

The program's main window resembles a spreadsheet, in which each of the stimuli specified by the user occupies a separate row, and the statistics for that stimulus are displayed in separate columns. There are three ways to input stimuli to the program: (1) by typing individual stimuli into the edit line at the top of the screen; (2) by using the File | Open menu option to read in a text file (a list of stimuli, with one stimulus per line); or (3) by pasting a list of stimuli from the clipboard (by using the Edit | Paste menu option, or the right click pop-up menu, or just the shortcut Ctrl-V). The latter option is particularly useful when one has a list of words in another open document (e.g., an Excel spreadsheet or a text file), since the list can be selected, copied onto the clipboard, and then pasted directly into the program.

Available Statistics

When the program starts, the only reported statistic is the total CELEX frequency (per million). Additional output fields can be selected by clicking the Analyse Options button. This brings up a list of available statistics, which can be divided into five broad categories: word-frequency statistics, orthographic statistics, phonological statistics, neighborhood statistics, and assorted other statistics. In the following description, output fields are denoted in italicized capital letters (e.g., *CELEX_W*).

Word-Frequency Statistics

The fact that high-frequency words are responded to more rapidly than low-frequency words are in word-reading tasks is possibly the most reliable finding in experimental psycholinguistics. The relationship between word frequency and reaction time (RT) measures has been studied extensively (e.g., Carroll & White, 1973; Forster & Chambers, 1973; Monsell, Doyle, & Haggard, 1989; Oldfield & Wingfield, 1965; Whaley, 1978). There is a logarithmic relationship between word frequency and RT; that is, frequency differences exert a stronger influence in the lower range of frequencies than in the upper range (e.g., the difference between words with frequencies of 10 and 20 per million is much more salient than the difference between words with frequencies of 100 and 110 per million). In view of the importance of word frequency, one must employ accurate measures of word frequency when designing stimuli for psycholinguistic experiments. Because of the logarithmic relationship between word frequency and RT, it is particularly important to have accurate frequency measures for low-frequency words. However, these words will tend to be associated with the least reliable estimates of frequency, given that word-frequency counts are based on limited samples. This is a particular problem for older word-frequency counts, which were based on relatively small corpora. The most satisfactory solution to this problem is to employ frequency counts that are based on large (and representative) corpora, and, where feasible, to make use of more than one corpus. To this end, the N-Watch program offers a choice of frequency measures, based on five different corpora.

One of the most frequently employed word-frequency counts is the Kučera and Francis (1967) word count, which is based on a corpus of 1,014,000 words of text compiled at Brown University. The KF output field is an integer that indicates how many times the input stimulus occurred in the entire corpus and, as such, can be interpreted as an estimate of frequency per million words. The remaining corpora are considerably larger, but their frequency values are scaled so that they always represent a frequency of occurrence per million words. By default, these values include two decimal places, but the degree of precision can be adjusted by using the View | Options menu.

A more recent source of word-frequency information is the CELEX English linguistic database (Baayen et al., 1995).[1] This frequency count is derived from the COBUILD corpus of 17.9 million words, 16.6 million of which were sampled from written sources (a set of 284 contemporary written texts), the remaining 1.3 million being sampled from spoken English. The full CELEX database includes measures of lemma frequency, which were estimated by a process of disambiguation (for example, determining how often jumper refers to an item of clothing as compared with how often it refers to someone who jumps). For experiments involving isolated word stimuli, however, these disambiguated frequencies are typically irrelevant, because the variable of interest is the frequency with which the given letter string occurs in English text (irrespective of the context in which it occurs). Three different CELEX frequency counts can be obtained from the N-Watch program: (1) a word's total *CELEX* (COBUILD) frequency (which combines the written and spoken frequency counts), (2) its written frequency (*CELEX_W*), and (3) its spoken frequency (*CELEX_S*). In general, the COBUILD corpus is to be preferred over the Kučera–Francis count, since it is based on a sample that is both significantly larger and more representative of everyday reading texts.

Another quite recently compiled corpus is the British National Corpus (BNC), which is based on a sample of 89.7 million words of text derived from 3,261 written texts representing an extremely broad range of modern British English.[2] Finally, the *SMH* corpus was compiled by Dennis (1995) and is based on 38,526 articles published in the *Sydney Morning Herald* in 1994 (23,440,636 words in all).[3] By matching words on a combination of different frequency counts (e.g., BNC, CELEX, and KF), it should be possible to achieve good frequency matches. In some cases, it may be appropriate to match items on log frequency. One of the program's output fields (*LOG10_FRQ*) returns the (base 10) logarithm of a word's frequency (plus 1); it should be noted that the base word frequency in this case depends on the presently selected corpus (the default being the CELEX-derived corpus; see below for further details).

Orthographic Statistics

Most of the statistics in this category are bigram- and trigram-frequency measures, which are both position and length sensitive. For example, the stimulus *edge* contains three bigrams (*ed*, *dg*, and *ge*). For the first of these (*ed*), the corresponding bigram frequency is based on the number (and frequency) of 4-letter words which begin with *ed*. Thus, the type frequency for *ed* is 4 (the four types being *eddy*, *edge*, *edgy*, and *edit*), and the token frequency is the sum of the word frequencies for these four types (= 80.95). By contrast, the bigram frequency for *ed* in Positions 3 and 4 (as in *fled*) is considerably higher: Here, the type frequency is 30, and the token frequency is 1,383.79. These bigram frequencies were computed on the basis of the COBUILD/CELEX word-frequency corpus.[4] The N-Watch program uses these bigram frequencies to compute a variety of summary measures for the entire string. The *BF_TK* field outputs the average bigram token frequency across the entire letter string; for example, for *edge* the *BF_TK* value equals (81 + 79 + 343) / 3 = 167 (rounding all decimal places). The *BF_TP* field is the average bigram type frequency across the entire letter string; for example, for *edge* the *BF_TP* value equals (4 + 2 + 9) / 3 = 5. Summed log bigram frequency (*SLBF*) is the sum of the logarithms of the token frequencies of each of the bigrams contained in the letter string. Mean log bigram frequency (*MLBF*) is simply *SLBF* divided by the number of bigrams in the stimulus (i.e., the number of letters −1). The trigram-frequency measures are computed in a similar fashion. For example, *word* contains the two trigrams *wor* and *ord*, and the summed (position- and length-sensitive) frequency of these two trigrams is (6 + 4) = 10 (for types) and (1,158 + 455) = 1,613 (for tokens), resulting in values of 5 and 807 for *TF_TP* and *TF_TK*, respectively, and of 5.7 and 2.9 for SLTF and MLTF, respectively. Finally, *LEN_L* is the number of letters in the stimulus, and the *CV_O* field provides a simple description of the letter string's orthographic consonant–vowel structure (e.g., *word* has a CVCC structure).

Phonological Statistics

Most of the phonological statistics output by the program are specific to words, or, more correctly, those words for which a pronunciation is listed in the vocabulary file (unlisted stimuli return values of −1). These output fields include the word's (British English) pronunciation (*CLX_PRON*, *DISC_PRON*), its initial phoneme (*P1*), its stress patterns (*STRESS*), the number of phonemes (*LEN_P*) and syllables (*LEN_S*) that it contains, its spelling-to-sound regularity (*REG*), and whether it has any homophones (*HOM*). If the word has a homophone, the spelling of this homophone is output, otherwise a value of −1 is returned. The pronunciation of a word is transcribed in both CELEX phonetic codes (*CLX_PRON*) and DISC phonetic codes (*DISC_PRON*). CELEX codes are closer to IPA than are DISC codes and may be easier to read for those familiar with this system. The advantage of DISC coding, however, is that each phoneme is coded by a single character. For example, the word *charge*, which comprises three phonemes, is coded by the three-character DISC code J#_, whereas the corresponding CELEX code (tSA:dZ) requires six characters. Additional characters are employed as syllable and stress markers. In the *CLX_PRON* field, syllables are delimited by square brackets (e.g., the two-syllable word discharge is coded as [dIs][tSA:dZ]). In the *DISC_PRON* field, syllable boundaries are indicated by hyphens, and the syllable with primary stress is indicated by an initial apostrophe (e.g., *discharge* is coded as dIs-'J#_). The *STRESS* field returns the number of the syllable that is stressed in the word (always 1 for monosyllabic words).

Two additional phonological fields produce outputs based on spelling–sound rules rather than pronunciations listed in the vocabulary file. The first of these

(*GPC*) outputs the regular pronunciation of the letter string that results from application of the grapheme–phoneme correspondence rules specified in the Dual-Route Cascaded (DRC) model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Finally, the *PS_HOM* field can be used to determine whether a letter string is a pseudohomophone; for example, given the input *phocks*, this field returns the spelling *fox*, indicating that the input string is a pseudohomophone that sounds like this word. A value of −1 is returned for inputs that are not pseudohomophones (including inputs that are words). Grapheme-conversion rules do not specify syllable structure, and hence the *GPC*, *REG*, and *PS_HOM* fields are unsuited to the analysis of polysyllabic inputs.

Neighborhood Statistics

There are 12 statistics in this category. *N* is the standard measure of orthographic neighborhood size, determined by counting the number of words that can be formed by substituting a single letter at any of the letter positions within the string (Coltheart et al., 1977). This metric has been found to be related to measures of performance in a variety of reading tasks, including lexical decision, naming, perceptual identification, and semantic categorization (see Andrews, 1997, for a review). It should be noted that the program counts a word as an orthographic neighbor only if that word is included in the presently selected vocabulary (e.g., the default vocabulary does not include the word *chad*, so this word is not counted as a neighbor of *chat*).[5] A list of the neighbors for each stimulus can be seen by switching to a different window (Window | Neighbor List); choose (Window | Main Form) to return to the main window.

Several fields provide information about the distribution of neighbors. The fields *N1* through *N4* display the number of neighbors at Positions 1–4, respectively; for example, *calm* has two neighbors at the first position (*balm* and *palm*), two at the fourth position (*calf* and *call*), and none at Positions 2 and 3. *P* is a count of the number of positions at which legal neighbors can be formed (e.g., $P = 2$ for the stimulus *calm*). Pugh, Rexer, and Katz (1994) found that this metric, which they referred to as "spread," was inversely correlated with lexical decision latency.

Other fields provide information about the frequency of a letter string's neighbors; which of the word-frequency counts is used for this purpose (e.g., *CELEX_W*, *KF*, etc.) depends on the presently selected corpus. The summed frequency of these neighbors (a token-based version of the *N* metric) is measured by the field *NF_SIG*. The average frequency of the letter string's neighbors (i.e., *NF_SIG* / *N*) is measured by *NF_MU*. *NF_MAX* is the frequency of the highest frequency neighbor (e.g., *calm*'s highest frequency neighbor is *call*, which has a CELEX frequency of 239 per million). *NF_MIN* is the frequency of the lowest frequency neighbor (e.g., *calm*'s lowest frequency neighbor is *balm*, which has a CELEX frequency of 1 per million). Finally, it has been suggested that the critical neighbor frequency variable is relative frequency rather than absolute frequency (e.g., Grainger, O'Regan, Jacobs, & Segui, 1989). Two output fields provide measures of relative frequency: *HFN* is the number of neighbors of the input that have higher frequencies than the input string, whereas *LFN* is the number of neighbors of the input that have lower frequencies than the input string. For example, of the four neighbors of *calm*, one is a higher frequency neighbor, and three are lower frequency neighbors. In the case where the input is a nonword, $HFN = N$, and $LFN = 0$.

Other Measures of Orthographic Similarity

Although investigations of orthographic similarity effects have focused mainly on orthographic neighbors, other forms of similarity relationship have been shown to influence performance in standard reading tasks. For example, perception of the word *calm* is affected not only by the presence of orthographic neighbors like *call*, but also by the presence of the orthographically similar word *clam*. This type of similarity relationship, in which two letter strings differ with respect to a single pair of adjacent letters, is known as *transposed letter* (TL) similarity. Empirical work has shown that TL similarity affects performance in a variety of reading tasks, including lexical decision, naming, and semantic categorization (e.g., Andrews, 1996; Chambers, 1979; Taft & van Graan, 1998). Selecting the *TL* field causes the program to check whether the input string is a member of a TL pair—that is, if a word can be formed by transposing an adjacent pair of letters in the input string. If a TL competitor is found, its identity is reported in the *TL* field (e.g., given the input *trial*, the output of the *TL* field is *trail*). If the field *TL_FRQ* is selected, the frequency of the other member of the TL pair is reported. The field *TL_POS* records the (initial) position of the letter transposition (e.g., for the word *trial*, $TL\_POS = 3$).

A further form of orthographic similarity that has recently been shown to influence reading performance is subset/superset similarity. For example, the presence of the embedded word *arm* within the word *army* interferes with both lexical decision (Davis & Taft, in press) and semantic categorization (Bowers, Davis, & Hanley, 2003). Similarly, Davis and Taft found that low-frequency words with higher frequency supersets (e.g., *patent*, which has the higher frequency superset *patient*) take longer to classify in lexical decision than do matched control words. Selecting the fields *SUB* and *SUP* causes the N-Watch program to identify subsets and supersets of the input stimulus, respectively; the frequency of these subsets and supersets can be obtained by selecting the fields *SUB_FRQ* and *SUP_FRQ*.

Specifying a Custom Vocabulary

In some circumstances, it is desirable to calculate measures of orthographic similarity with respect to a specific vocabulary. For example, if a computational model has been trained (or hand-wired) with a particular vocabulary, it may be useful to know how many neighbors a given letter string has in that vocabulary, rather than relying on statistics derived from larger vocabularies. The word *bats* has 17 neighbors in the program's default vocabulary but only 2 neighbors in the vocabulary employed by the IA model (McClelland & Rumelhart, 1981). Thus, it would be inappropriate to classify this item as belonging to a high-density orthographic neighborhood when running simulations of the IA model. Selecting Vocab | IA Vocab configures the program to return orthographic similarity statistics that are based on the 1,179-word vocabulary of the IA model. More generally, selecting Vocab | User Vocab causes the program to output orthographic similarity statistics that are based on a user-specified vocabulary. Vocabulary files should have the extension .vcb and should contain one row for each word, where each row consists of the word's spelling and its frequency (separated by a tab character, and followed by a line break; any additional columns of data will be ignored). It is permissible to use dummy values for the word frequencies in this file, although it should be noted that these values will be employed for the purpose of computing neighbor-frequency measures, number of higher frequency neighbors, and related statistics (e.g., TL frequency).

Subjective Ratings of Familiarity, Frequency, AOA, and Imageability

Other variables that are related to word frequency, but are not themselves objective frequency measures, include subjective familiarity, subjective frequency, and AOA. Subjective familiarity is measured by asking readers to rate the familiarity of a words—for example, on a scale of 1 (*very unfamiliar*) to 7 (*very familiar*). The MRC Psycholinguistic Database (Coltheart, 1981) merged three sets of familiarity norms (Gilhooly & Logie, 1980; Paivio, Yuille, & Madigan, 1968; Toglia & Battig, 1978) to derive measures of printed word familiarity for several thousand words (including 4,412 words in the N-Watch default vocabulary). These norms are returned in the program's *FAM* output field; they are scaled to range between 100 and 700, where higher values indicate greater subjective familiarity. Some researchers have argued that familiarity ratings provide a superior index of the relative frequency of exposure to a word than do objective measures of printed word frequency (e.g., Gernsbacher, 1984; Gilhooly & Logie, 1980). However, Balota, Pilotti, and Cortese (2001) argued that estimates of a word's familiarity are contaminated by semantic variables and that a better way to frame the question is to ask raters to estimate the frequency of exposure. The output of the *EST_FRQ* field is based on the corpus collected by Balota et al., which contains estimates of subjective frequency for 2,938 monosyllabic words. Several hundred participants rated the estimated frequency with which they encountered each word on a scale of 1 (*never encountered*) to 7 (*encountered several times a day*). For comparability with the other rating scales, the original values have been multiplied by 100 to produce integers ranging between 100 and 700.

A number of studies have claimed that effects that have traditionally been attributed to word frequency are better interpreted as reflecting the confounded effects of another variable—the AOA of a word (e.g., Brown & Watson, 1987; Carroll & White, 1973; Morrison & Ellis, 1995). Accordingly, it is appropriate to consider the role of this variable in psycholinguistic experiments, particularly when word frequency is a variable of interest. AOA is typically measured by asking adults to provide ratings of the age at which they acquired words; such ratings tend to be highly reliable (e.g., Carroll & White, 1973) and are highly correlated with objective measures of AOA (Morrison, Chappell, & Ellis, 1997). The N-Watch program returns AOA ratings from two separate databases. The *AOA* output field is based on the norms reported by Gilhooly and Logie (1980); these norms have been multiplied by 100 to produce a range from 100 to 700. Using these norms, AOA values are available for 1,686 words in the default vocabulary. The *AOA2* output field is based on the norms reported by Bird, Franklin, and Howard (2001); these norms also range from 100 to 700.[6] Using these norms, AOA values are available for 1,988 of the words in the default vocabulary.

Another lexical variable that has been found to influence the speed of word identification is imageability—that is, the ease with which a mental image of the word can be formed (e.g., James, 1975; Whaley, 1978). The N-Watch program returns two different imageability norms. The *IMG* output field reports the imageability norms from the MRC Psycholinguistic Database (Coltheart, 1981), which were formed by merging the three smaller databases mentioned earlier (Gilhooly & Logie, 1980; Paivio et al., 1968; Toglia & Battig, 1978); these norms provide measures of imageability for 4,334 words in the default vocabulary. The *IMG2* output field is drawn from the imageability norms collected by Bird et al. (2001); norms for 1,342 words in the default vocabulary are available. Both sets of norms are scaled to produce a range from 100 to 700, where a higher value indicates greater imageability (e.g., *beach* has the highest MRC imageability score of 667; *equanimity*, by contrast, has an imageability score of 145).

Saving the Output

There are two ways to extract the output from the program: (1) by using the File | Save menu option to save the output to a text file (one stimulus per line, with tabs separating the output fields); or (2) by copying selected output to the clipboard (by using the Edit | Copy menu option, or the right click pop-up menu, or just the shortcut Ctrl-C). Once again, the latter option is useful when working with a spreadsheet program like MS Excel: The required fields can be selected, copied to the clipboard and then pasted directly into an open spreadsheet. To select all the input rows and output columns containing data, the option Copy All can be selected from the right click pop-up menu.

REFERENCES

Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory & Language*, **35**, 775-800.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, **4**, 439-461.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX Lexical Database*. *Release 2* [CD-ROM]. Linguistic Data Consortium, University of Pennsylvania, Philadelphia.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, **29**, 639-647.

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, **33**, 73-79.

Bowers, J. S., Davis, C. J., & Hanley, D. (2003). *Automatic semantic activation of embedded words: Is there a "hat" in "that?"* Manuscript submitted for publication.

Brown, G. D., & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and naming latency. *Memory & Cognition*, **15**, 208-216.

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, **25**, 85-95.

Chambers, S. M. (1979). Letter and order information in lexical access. *Journal of Verbal Learning & Verbal Behavior*, **18**, 225-241.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, **33A**, 497-505.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). New York: Academic Press.

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.

Davis, C. J., & Taft, M. (in press). More words in the neighborhood: Interference in lexical decision due to deletion neighbors. *Psychonomic Bulletin & Review.*

Dennis, S. (1995). The Sydney Morning Herald Word Database. *Noetica: Open Forum*, **1**. Available at http://psy.uq.edu.au/cogpsych/noetica/.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behavior*, **12**, 627-635.

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.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, **12**, 395-427.

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.

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception & Performance*, **1**, 130-136.

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

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, **88**, 375-407.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, **118**, 43-71.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, **50A**, 528-559.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 116-153.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, **17**, 273-281.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs*, **76**(Suppl. 2), 1-25.

Pugh, K. R., Rexer, K., & Katz, L. (1994). Evidence of flexible coding in visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 807-825.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology*, **55A**, 1339-1362.

Taft, M., & van Graan, F. (1998). Lack of phonological mediation in a semantic categorization task. *Journal of Memory & Language*, **38**, 203-224.

Toglia, M. P., & Battig, W. R. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.

Whaley, C. P. (1978). Word–nonword classification time. *Journal of Verbal Learning & Verbal Behavior*, **17**, 143-154.

NOTES

1. The CELEX database is available on CD-ROM from the Linguistic Data consortium: http://www.ldc.upenn.edu/.

2. The frequencies from this corpus are available on line at the following ftp address: ftp://ftp.itri.bton.ac.uk/bnc/.

3. The SMH frequencies are available online at the following address: http://www2.psy.uq.edu.au/CogPsych/Noetica/OpenForumIssue4/SMH_Statistics.Filtered.

4. The program for generating these values can be obtained from the author.

5. Users of the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002) should note that the values of *N* computed by the N-Watch program for some nonwords will differ slightly from those given in the ARC Nonword Database, because the former count includes polysyllabic neighbors (e.g., *jury* is counted as a neighbor of *jurz* by the N-Watch program but not by the ARC Nonword Database).

6. The Bird et al. (2001) database includes separate AOA and imageability ratings for homographs (e.g., for the noun and verb senses of *wonder*). In these cases, the program reports the smaller AOA value—that is, that corresponding to the earlier acquired word and the imageability rating associated with this entry.

(Manuscript received May 22, 2003;
revision accepted for publication February 14, 2004.)