

## The English Lexicon Project

David A. Balota<sup>1</sup>, Melvin J. Yap<sup>1</sup>, Michael J. Cortese<sup>2</sup>,  
Keith I. Hutchison<sup>3</sup>, Brett Kessler<sup>1</sup>, Bjorn Loftis<sup>7</sup>, James H. Neely<sup>4</sup>,  
Douglas L. Nelson<sup>5</sup>, Greg B. Simpson<sup>6</sup>, and Rebecca Treiman<sup>1</sup>

**(manuscript in press at Behavior Research Methods)**

<sup>1</sup>Department of Psychology, Washington University

<sup>2</sup>Department of Psychology, College of Charleston

<sup>3</sup>Department of Psychology, Montana State University

<sup>4</sup>Department of Psychology, SUNY at Albany

<sup>5</sup>Department of Psychology, University of South Florida

<sup>6</sup>Department of Psychology, University of Kansas

<sup>7</sup>Southwestern Bell Corporation

Correspondence should be sent to: David Balota  
Department of Psychology  
Washington University  
St. Louis, MO 63130

Tel: 1-314-935-6549

Fax: 1-314-935-7588

Email: [dbalota@artsci.wustl.edu](mailto:dbalota@artsci.wustl.edu)

Abstract

The English Lexicon Project is a multi-university effort to provide a standardized behavioral and descriptive dataset for 40,481 words and 40,481 nonwords. It is available via the Internet at <http://lexicon.wustl.edu>. Data from 816 participants across six universities were collected in a lexical decision task (approximately 3400 responses per participant), and data from 444 participants were collected in a speeded naming task (approximately 2500 responses per participant). The present paper describes the motivation for this project, the methods used to collect the data, and the search engine that affords access to the behavioral measures and descriptive lexical statistics for these stimuli.

## The English Lexicon Project

One could easily argue that the most commonly studied stimulus set in experimental psychology involves English words. The study of the memory and reading of words has been central to research since Cattell (1886). Words are well-described units that provide the link between perception and meaning, and so have been critical to developments in computational modeling (e.g., McClelland & Rumelhart, 1981), neuroimaging (e.g., Petersen, Fox, Posner, Mintun, & Raichle, 1989), and conceptions of attention and automaticity (e.g., Neely, 1977; Stroop, 1935), among many other research areas.

Given the importance of words as a stimulus set, one might assume that there are relatively straightforward ways to study lexical processing, and that there is a well-constrained set of findings to which one can appeal in building models of word processing. Although there has been considerable progress in understanding how people process words, there are some clear gaps in the available evidence. This paper describes the English Lexicon Project (ELP), which provides a behavioral database of over 40,000 words and nonwords that will help fill some of these gaps. The present description will focus on visual word recognition, although, as described below, the current database has relevance for other aspects of word processing, such as memory and production of words during speech. Before describing the ELP, we will briefly describe the behavioral measures in the database, the limitations in our current knowledge, and how this database will help address these limitations.

### Lexical Decisions and Naming as the Behavioral Targets

Although there are multiple ways to measure lexical processing (e.g., eye-fixation data, probability of identifying degraded stimuli, semantic verification), most work on isolated

visual word recognition has employed two relatively simple tasks, lexical decision and speeded naming. In the lexical decision task (LDT), participants are presented with a string of letters (either a word or a nonword, e.g., FLIRP), and are asked to press one button if the string is a word and another button if the string is a nonword. In the speeded naming task, participants see a visual word (or sometimes a nonword), and are asked to name the word aloud as quickly and as accurately as possible. Although researchers have clearly demonstrated task-specific effects in the LDT and naming task (e.g., Balota & Chumbley, 1984, vs. Balota & Chumbley, 1985; Neely, Keefe, & Ross, 1989, vs. Keefe & Neely, 1990), these two tasks are still the major driving force in research on isolated word recognition and have been the gold standard in developing computational models of lexical processing (see Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1996; Plaut, McClelland, Seidenberg, & Patterson, 1996, Seidenberg & McClelland, 1989; Zorzi, Houghton, & Butterworth, 1998). The ELP includes measures of both lexical decision and naming performance.

#### What are the Limitations in Current Knowledge?

Researchers investigating the processes involved in visual word recognition typically employ factorial designs in which item variables (e.g., word frequency, spelling-to-sound regularity, neighborhood density, syntactic class) are “manipulated” on a relatively small set of items (typically 10 to 20 items per condition) and the effects of these variables on the speed and accuracy of lexical decision and/or naming performance are measured. In the vast majority of these studies, a mean is calculated for each participant across items (or for each item across participants in some studies) and then entered into an analysis of variance, and the effects of “factors” are measured. Although this approach has been fruitful in identifying

a number of important variables that modulate speeded lexical decision and naming performance, it is also clear that this approach has some potential limitations.

First, it is difficult to select a set of items that only vary on one dimension, while controlling the myriad of variables that have been identified in the literature (see Cutler, 1981). For simplicity, consider the influence of spelling-to-sound correspondences, such as the fact that the word PINT is not pronounced according to common spelling-to-sound principles, whereas HINT is consistent with such principles. The influence of irregular spelling-to-sound correspondence will depend on a number of factors such as the frequency of the target word, the number and frequency of neighbors with similar spelling-to-sound correspondences (friends), the number and frequency of neighbors with different spelling-to-sound correspondences (enemies), and probably a host of other variables (see, e.g., Jared, McRae, & Seidenberg, 1990). A single set of 10 to 20 stimuli used in a given experiment may indeed show idiosyncratic influences of the multitude of uncontrolled variables that limits the generalizability of the targeted variable to other sets of items.

Second, one may be concerned that implicit knowledge by the researcher could bias the selection of words for an experiment. For example, Forster (2000) has shown that researchers in word recognition can predict very well which of two relatively well-matched stimuli would yield faster lexical decision latencies. Hence, it appears that researchers possess knowledge of lexical processing that could implicitly or explicitly play a role in the selection of stimuli to test hypotheses in a given experiment.

Third, list contexts (e.g., the other items in a list) can influence how a variable affects performance in a word recognition experiment. For example, Andrews (1997), in a review of the orthographic neighborhood size literature, has argued that the inconsistencies in the

lexical decision task data are largely driven by differences in stimulus list environments (e.g., legal, FLIRP, vs. illegal, FLPRI, nonword foils) across experiments. Thus, list contexts can modulate the presence, magnitude, and even direction of experimental effects in the literature.

Fourth, the vast majority of lexical processing research has been dedicated to single-syllable words, and hence, many experiments have exclusively used lists that only include such words. The major computational models that have been developed by Coltheart et al. (2001), Plaut et al. (1996), and Seidenberg and McClelland (1989) have all exclusively dealt with monosyllabic words (but see Ans, Carbonnel, & Valdois, 1998, for an example of a multisyllabic computational model). Many interesting issues arise with multisyllabic words, such as the role of syllable structure, stress patterning, morphological structure, type of length effects (e.g., orthographic vs. phonological vs. syllabic length), among many others. It is even possible that single-syllable lists may produce a list context effect that may directly modulate the results.

### The English Lexicon Project

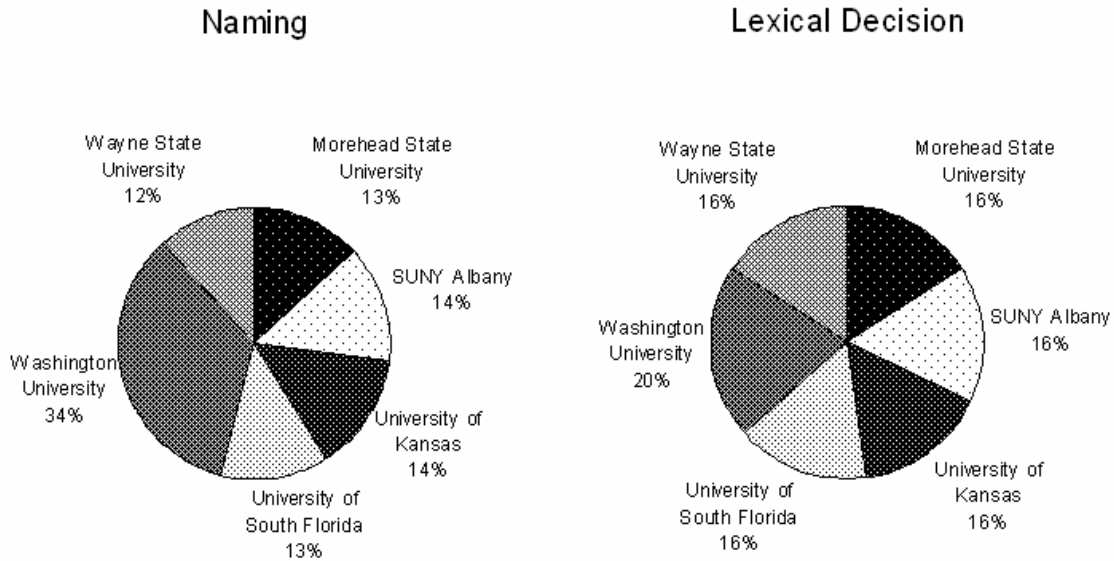
Given the richness of the English lexicon, along with the problems noted above, it is appropriate to expand the horizons of experimental work addressing lexical processing beyond the standard factorial experiments with a relatively limited set of monosyllabic stimuli. The ELP is such an attempt. The ELP is a large database of descriptive and behavioral data, along with a search engine that affords access to this database. The ELP involves a multi-university collaborative effort to provide a large database of naming and lexical decision latencies across 1260 participants for 40,481 words and nonwords, along with speeded naming latencies for each of the words. The goal of this project is to provide a

database that is easily accessible for researchers from diverse backgrounds and interests that can at least minimize the concerns raised above with current approaches to lexical processing.

A few researchers have gone beyond traditional factorial studies to provide data on large databases. Such *megastudies* provide estimates of word recognition performance for a large set of words (Balota & Spieler, 1998; Besner & Bourassa, 1995; Kessler, Treiman, & Mullennix, 2002; Spieler & Balota, 1997; Treiman, Mullennix, Bijeljac-Babic, and Richmond-Welty, 1995). For example, Treiman et al., using multiple regression and speeded naming latencies as a dependent variable, examined the effects of different sub-syllabic units on naming performance. Similarly, by examining the lexical decision and naming latencies to 2,428 words (Balota & Spieler, 1998; Spieler & Balota, 1997), Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004) demonstrated that standard psycholinguistic variables (e.g., length, frequency, orthographic neighborhood size) were able to account for nearly half the variance in the behavioral data. However, such large-scale studies are relatively rare, tend to focus mostly on speeded naming performance, the raw data are not easily accessible to the research community, and, are almost always based on single syllable words (see Chateau & Jared, 2003, for an exception). The ELP is the logical extension to this body of work, and provides normative word recognition performance for 40,000 mono- and multisyllabic words, along with a search engine that affords access to a rich set of descriptive lexical characteristics. We will now turn to a presentation of the ELP, and later in the General Discussion section will present some of the many possible uses of this database.

## Method

*Participants.* Participants were native English speakers recruited from research participant pools in the six testing institutions (see Figure 1 for a breakdown of the proportion of participants from each institution and Table 1 for descriptive statistics of participant demographics for the two tasks). The universities included private and public institutions across the Midwest, Northeast and Southeast regions of the United States, and at each institution a word recognition researcher directed the collection of data. Participants were either paid \$25 for participation in the two sessions or received credit for course requirements. Each participant participated in either the naming task or the lexical decision task. Participants took part in two different sessions that were conducted on different days, separated by no more than one week. At the end of the second testing sessions, participants also received the Morningness-Eveningness Questionnaire (a circadian rhythm questionnaire) (Horne & Ostberg, 1976), the Shipley (1940) vocabulary test, and a general health questionnaire. Based on the consistency of the previous datasets that have been developed (e.g., Spieler & Balota, 1998), we decided that an individual participant could produce stable data for approximately 2,500 (naming) to 3,500 stimuli (lexical decision).



**Figure 1. Proportion of participants from each institution.**

*Apparatus.* The experimental software and testing apparatus were identical at each of the six testing sites. Stimuli were presented via a 19 inch Trinitron monitor (P991), which was controlled by a Pentium III PC. Stimuli were presented in uppercase in the standard QBASIC font in the 80 (column) x 30 (row) mode. For the naming task, there were two microphones. One microphone (Sony condenser microphone, ECM-MS907) was connected to a Sony DAT recorder (PCM-M1) for future off-line analyses of the naming data. The second was a head-mounted Shure microphone (WH20QRT), which was connected to an in-house developed electronic voice relay. This voice-key was modeled after the Gerbrands G1341T microphone (no longer commercially available) and had similar response characteristics as the Gerbrands voice-key.

*Stimuli.* Because of the enormous potential size of the candidate set of words for this project, we were compelled to make decisions to limit the size of the master list. The master list of 40,481 words was based initially on the Kučera and Francis (1967) norms, and then was supplemented by the CELEX (Baayen, Piepenbrock, & Gulikers, 1993)

norms. The Kučera and Francis stimuli were initially screened so that stimuli that were clearly nonwords (probably due to typographical errors) or had non-alphabetic characters (e.g., A\*\*\$RE) were eliminated. Because the target set of participants in this study were undergraduates, we had three undergraduate research assistants at Washington University, under the direction of a graduate student and a faculty member, assist in the selection of the items. Words that were eliminated included obscenities, slurs, misspellings, and alternate (e.g., British English) spellings. Common multimorphemic words were included, but because of space constraints, we could not include all possible derivatives for a given stem. For example, JOYFUL and JOYFULLY were kept but UNJOYFULLY was removed. Proper names and contractions were included. After the master list was constructed, nonwords were created for the lexical decision task. Pronounceable nonwords were generated by changing one or two letters in a corresponding target word. The location of the letter change alternated across different words to include early, middle, and late positions.

*Creation of Sublists for Individual Participants.*

*Naming.* Participants either received 2530 or 2531 items. For all 40,481 words to be sampled once, 16 participants (15 with 2530 and 1 with 2531) were necessary. To obtain 25 observations per word, 400 participants were necessary ( $16 \times 25 = 400$ ). In order to assign words to sublists, words in the master list (40,481 items) were first sorted alphabetically. To minimize the likelihood that an individual participant would name words with the same morphemic stem, the first 16 adjacent words in the master list were randomly assigned to 16 sublists. This process was repeated until there were no unassigned words in the master list. This process of assigning words from the master list

to the 16 sublists was repeated 25 times, resulting in a sublist for each of the 400 participants to be tested. This process also insured that, for each group of 16 participants, no two participants named the same word. Practice and buffer items (see below) were taken from other participants in the group of 16 (i.e., participant 1 received buffer items from participant 2's test lists).

*Lexical Decision.* A similar procedure was used to create sublists for the lexical decision task portion. For LDT, since there was a nonword for each word, the total number of items in the master list was 80,962. For the LDT, 24 participants (17 with 3374 and 7 with 3372) were necessary for a single observation for all 80,962 items. Hence, 816 ( $24 \times 34 = 816$ ) participants afforded 34 observations per item in lexical decision performance. For the LDT, words were sorted alphabetically, and when words were assigned to the sublists, the corresponding nonword was also assigned to the same sublists. To minimize the number of times a participant would see words with the same root, 24 adjacent words in the master list were randomly assigned to 24 sublists.

*Procedures.*

*Naming Task.* Participants received 40 practice trials before each experimental session. The practice trials for a given participant were selected from a different counterbalanced list. Each session contained several blocks of 250 trials (the last block of the last session was either 280 or 281 trials). The first session consisted of 1500 trials; the second session was either 1030 or 1031 trials. Each block of 250 trials was followed by a rest break. The break after practice trials was 15 seconds long. At 750 trials, the break was 3 minutes long. All other breaks (after each 250 trials) were 1 minute long. During 3-minute long breaks, the computer screen instructed participants to "Please use this time to

get a drink, stretch, or walk around”. A countdown timer was shown at the bottom of the screen to inform participants how much time was remaining in the break.

During each break, the proportion of accurate trials and mean response latencies were presented for the 250 trials preceding the break. If the accuracy was 80% or greater, a message “Please maintain this level of accuracy” was presented. If the accuracy was less than 80%, a message “Please increase your level of accuracy” was presented. If the mean response latency was less than 1000 ms, a message “Please maintain this reaction time” was presented. If the mean response latency was greater than 1000 ms, a message “Please decrease your response time” was presented. After the break, three auditory tones were presented to alert the participant. The participant was instructed to press B to begin the next block. A blank screen was presented 2.5 second after the keypress and before two buffer trials initiated the next test block.

The sequence of events during each naming trial was as follows: (a) three asterisks were presented at the center of the screen for 250 ms; (b) a 50 ms tone was then presented indicating the onset of the next trial; (c) a 250 ms dark interval was presented; (d) the target word was presented centered at the same location the asterisks were presented; (e) the participant named the word; (f) the computer detected the voice onset; (g) the word remained on the screen for an additional 250 ms after voice onset; (h) the word was erased from the screen. If the response latency for the vocal response was less than 4000 ms, a screen appeared asking the participant to manually code the accuracy of their response. The four choices were 1) correct pronunciation, 2) uncertain of pronunciation, 3) mispronunciation, and 4) microphone error. Participants were instructed beforehand regarding the importance and use of these coding options.

If the voice key detected a sound while the word was initially presented (0 ms response latency), an error message and a warning tone was presented. The trial was reinitiated after a 2500 ms delay. If no response was made within 4000 ms, the item was removed from the screen and the words “Too slow” were presented for 500 ms. At the end of each trial, a blank screen was displayed during the 1000 ms inter-trial interval.

*Lexical Decision Task.* There were 40 practice trials before each experimental session. As in naming, each session contained several blocks of 250 trials (the last block of the last session was only 122 or 124 trials). The first session included 2000 trials, whereas the second session included either 1372 or 1374 trials. The breaks were identical to the naming task, including the feedback concerning accuracy and response latency.

The sequence of events during each trial was the same as in the naming task, except that the participant pressed the “/” key for a word response and the “z” key for a nonword response. If the participant responded before the letter string was presented, an error message on the screen and tone were presented and the trial was reinitiated after a 2500 ms delay. If no response was made within 4000 ms, the item was removed from the screen and the words “Too slow” were presented for 500 ms. If the response was inaccurate, the words “Incorrect!” appeared for 500 ms. If the participant pushed an invalid key, instructions about the key mapping appeared for 2000 ms. At the end of the trial, a blank screen was displayed during the 1000 ms inter-trial interval.

### **Description of Lexical Characteristics and Behavioral Measures**

In this section, we describe the information that is available at the website at <http://elexicon.wustl.edu>. Descriptive statistics for words (see Table 2) and nonwords

(Table 3) are also included. Interested users are encouraged to access the website to explore the range of variables available.

In order to ensure that extreme response latencies did not disproportionately influence the item mean correct reaction time for the naming and lexical decision tasks, we used a two-step outlier identification procedure: First, any response latency faster than 200 ms or slower than 3000 ms were identified as outliers. Second, for the remaining correct RTs, a mean and standard deviation were calculated, and any RT that was less 3 SDs below the mean for that participant or greater than 3 SDs above the mean for that participant were also identified as outliers. The mean item latencies for naming and lexical decision performance, along with the z-score estimates, described below, are based on the remaining correct observations across the participants who received that particular item. Also, as described below, one can also access the individual participant level data if one wishes to utilize an alternative screening procedure.

### **Characteristics Available from Submitting and Generating Lists of Items**

As described further below, the user may submit a list of items to generate characteristics for that list of items or have the website generate a list of items with specific constraints. Within both of these commands, the following set of variables can easily be accessed.

#### **General Fields**

*Length* is the number of letters in the word.

*Freq\_KF* refers to the Kučera and Francis (1967) frequency norms. These norms are derived from a corpus of 1,014,000 words drawn from a variety of American English texts. Although these norms are available in the dataset, we strongly encourage users to

use the HAL norms (described below) or some other estimate of word frequency since it has been repeatedly demonstrated that the Kučera and Francis norms are relatively poor estimates of raw frequency, most likely because of its small corpus (see, for e.g., Balota et al. 2004; Zevin & Seidenberg, 2002).

*Freq\_HAL* refers to the Hyperspace Analogue to Language (HAL) frequency norms (Lund & Burgess, 1996), based on the HAL corpus, which consists of approximately 131 million words gathered across 3,000 Usenet newsgroups during February 1995.

*Log\_Freq\_HAL* refers to log-transformed HAL frequency norms.

### **Orthographic Neighborhood Characteristics**

*Ortho\_N* is the number of words that can be obtained by changing one letter while preserving the identity and positions of the other letters (i.e., Coltheart's N; Coltheart, Davelaar, Jonasson, & Besner, 1977). For example, the ELP returns the following orthographic neighbors of CAT: CAB, CAM, CAD, TAN, PAN, BAN, CAP, RAN, FAN, CAT, CON, VAN, DAN, CAR, and MAN.

*Freq\_N* is the mean log of HAL frequencies of the orthographic neighbors of a particular word.

*Freq\_Greater* is the number of orthographic neighbors of an item that are more frequent than that item based on the HAL frequencies.

*Freq\_G\_Mean* is the mean log of HAL frequencies of the orthographic neighbors for an item that are more frequent than that item.

*Freq\_Less* is the number of orthographic neighbors for an item that are less frequent than that item based on the HAL frequencies.

*Freq\_L\_Mean* is the mean log of HAL frequencies of orthographic neighbors for an item that are less frequent than that item.

*Freq\_Rel* is the log HAL frequency of an item divided by the mean log HAL frequency of its orthographic neighbors.

*Neighborhood Button* permits a user to display the actual orthographic neighbors of a submitted word or non-word.

### **Bigram Frequency Characteristics**

*BG\_Sum* refers to summed bigram frequency, where bigram is defined as a sequence of two letters. The summed bigram frequency of a letter string (e.g., DOG) is the sum of the frequencies of its successive bigrams (i.e., DO & OG).

*BG\_Mean* refers to average bigram frequency, which is the summed bigram frequency divided by the number of successive bigrams.

*BG\_Freq\_By\_Pos* refers to summed bigram frequencies by position, and is based on bigram frequencies that are sensitive to positions within words. Position-sensitive bigram counts take into account the letter positions where a bigram occurs. For example, the bigram frequency for DO in DOG counts DO bigrams only when they appear in the first two positions of a word.

### *Phonological Characteristics*

*Pronunciation* provides phonological transcriptions for words, and is largely based on the Unisyn Lexicon developed by the Centre for Speech Technology Research at the University of Edinburgh (Centre for Speech Technology Research, University of Edinburgh, n.d.).

*Number of phonemes* provides the phoneme count for a word. For example, CAT /kæt/ has three phonemes.

*Number of syllables* provides the syllable count for a word. For example, CAT /kæt/ has one syllable.

### **Morphological Characteristics**

*Morpheme Parse - Letters* parses words into morphemes (units of meaning), using letter segments (e.g., CATS is parsed as CAT-S).

*Morpheme Parse - Phonemes* parses words into morphemes, using phoneme segments (e.g., CATS is parsed as kæt-s).

*Number of Morphemes* provides the morpheme count for a word. For example, CATS has two morphemes, /CAT/ + /S/.

### **Part of Speech**

*Part of Speech* provides the word's part of speech, and shows how the word is used. Parts of speech include verbs, nouns, adjectives, and prepositions.

### **Behavioral Results: Lexical Decision**

*I\_Mean\_RT* is the mean lexical decision latency (in ms) for a particular word across participants.

*I\_Zscore* provides the standardized mean lexical decision latency for each word. Each participant's raw lexical decision latencies are standardized using a Z-score transformation, and the mean Z-score for all participants presented with a particular word is then computed. This metric allows the lexical decision performance for different words to be directly compared, with more negative Z-scores denoting shorter latencies. Because there is considerable variability across participants in overall response latency and each

participant only receives a subset of the stimuli, the standardized item score is the most reliable measure, minimizing the influence of a participant's processing speed and variability (see Faust, Balota, Spieler, & Ferraro, 1999).

*I\_SD* is the standard deviation of lexical decision latencies for a word across participants.

*Obs* is the number of observations available for a word.

*I\_Mean\_Accuracy* is the proportion of accurate responses for a particular word, excluding errors and outliers.

### **Behavioral Results: Naming**

*I\_NMG\_Mean\_RT* is the mean naming latency (in ms) for a particular word across participants.

*I\_NMG\_Zscore* provides the standardized mean naming latency for a word. This metric allows the naming performance for different words to be directly compared, with more negative Z-scores denoting shorter latencies (see above).

*I\_NMG\_SD* is the standard deviation of naming latencies for a word across participants.

*I\_NMG\_Obs* is the number of observations available for a word.

*I\_NMG\_Mean\_Accuracy* is the proportion of accurate responses for a particular word, excluding errors and outliers.

### **Behavioral Results: Lexical Decision to Nonwords**

*NWI\_Mean\_RT* is the mean lexical decision latency (in ms) for a particular nonword across participants.

*NWI\_Zscore* is the standardized lexical decision latency for a nonword. This metric allows the lexical decision performance for different nonwords to be directly compared, with more negative *Z*-scores denoting shorter latencies.

*NWI\_SD* is the standard deviation of lexical decision latencies for a nonword across participants.

*Obs* is the number of observations available for a nonword.

*NWI\_Mean\_Accuracy* is the proportion of accurate responses for a particular nonword, excluding errors and outliers.

### **Individual participant data**

In addition to accessing item information, the database also affords the opportunity to access individual participant data. The following describes the list of parameters used for such selection that are available for the lexical decision and naming data.

*Sub\_ID* refers to participants' unique ID.

*Trial* refers to the position a trial appears within an experimental block.

*Type* refers to the lexicality of a trial, where word=1 and nonword=0, only applicable for lexical decision data.

*D\_Accuracy* refers to trial accuracy, where accurate=1 and error=0.

*D\_RT* refers to trial response time (in ms).

*D\_word* shows the specific item presented on a trial.

*Outlier* indicates whether the response time for a trial is extreme or not, where outlier = 1 and non-outlier = 0 (see above description of outlier screening procedure).

*D\_Zscore* refers to z-standardized response times, with more negative z-scores denoting shorter latencies.

### **Using the ELP queries**

In this section we will describe how to use the search engine available at the ELP website. There are of course multiple configurations of the requests. Here we provide specific examples, but direct exploration of the website and search engine is encouraged.

As noted, the ELP has four basic functions:

- 1) Generate a list of stimuli with specific characteristics
- 2) Submit a list of stimuli to obtain lexical characteristics and behavioral performance measures (i.e., naming and lexical decision) for these stimuli
- 3) Submit a list of arbitrary nonword stimuli to obtain lexical characteristics
- 4) Provide access to individual participants' data

#### *1) Generate a list of stimuli with specific characteristics*

This function allows users to generate stimulus sets with pre-specified parameters. Users can select the characteristics they are interested in, as well as place constraints on the range of values for these characteristics. For example, a user might be interested in obtaining a pool of words which are four to five letters in length, low frequency (e.g., log HAL frequency  $\leq 4$ ), and monomorphemic. The instructions for doing this are provided below.

- a. Go to the ELP homepage (<http://elexicon.wustl.edu>), and click on 'Generate Lists of Words with Specific Lexical Characteristics'.

Washington University in St. Louis English Lexicon Project Web Site

**Welcome to the English Lexicon Project Homepage.**

**Overview**

The English Lexicon Project (supported by the National Science Foundation) affords access to a large set of lexical characteristics, along with behavioral data from visual lexical decision and naming studies of 40,481 words and 40,481 nonwords.

The naming and lexical decision data are currently being collected from six testing Universities. To date, we have collected 2,752,698 reaction time measurements from 816 subjects in the lexical decision experiment. We have also collected 1,125,880 experimental measurements from 444 subjects in the naming experiment.

**Database News and Update ~ [Click Here!](#)**

**There are four main functions of this Web Site:**

- ♦ [Generate Lists of Words with Specific Lexical Characteristics](#)
- ♦ [Generate Lists of Nonwords with Specific Lexical Characteristics](#)
- ♦ [Access Trial-Level Lexical Decision Data](#)
- ♦ [Access Trial-Level Naming Data](#)

**Figure 2. The English Lexicon Project Homepage.**



b. Click on ‘Generate Lists of Items with Specific Lexical Characteristics’.



c. This brings the user to the page shown below, where the different lexical and behavioral categories are displayed. Clicking on the ‘+’ icon expands a particular field.



For example, clicking on ‘General Fields’ reveals the length and frequency measures. For our query, we will check ‘Length’, ‘Log\_Freq\_HAL’, and ‘NMorph’.



Please select the fields and any constraints to be used in the query:

General Fields

Length  

Freq\_KF  

Freq\_HAL  


Log\_Freq\_HAL  


Orthographic Neighborhood



Bigram Frequency

Phonological Characteristics

Morphological Characteristics

MorphSp (Morpheme Parse - Letters) 

MorphPr (Morpheme Parse - Phonemes) 

NMorph (Number of Morphemes)  

Part of Speech

Behavioral Results: Lexical Decision

Behavioral Results: Naming

Figure 3. Interface for generating word lists with specific characteristics.

d. The icon to the right of the question mark is the constraint button, and allows users to set the range of values for their measure of interest. For example, to select words that are four to five letters long, click on the constraint button for ‘Length’. This brings up a pop-up box, where 4 and 5 can be respectively typed into the ‘Minimum Value’ and ‘Maximum Value’ boxes. Repeat this step for ‘Log\_Freq\_HAL’ (Minimum Value=0, Maximum Value=4), and ‘NMorph’ (Minimum Value=1, Maximum Value=1).

Please enter any desired constraints upon the length or leave blank if non are desired:

Attribute	Minimum Value	Maximum Value
Length	<input type="text" value="4"/>	<input type="text" value="5"/>

Figure 4. Placing constraints on the length variable.

e. At this point, the user has to respond to the two query options, ‘Query Scope’ and ‘Method of Distribution’. ‘Query Scope’ is, by default, limited to the restricted ELP

lexicon. Selecting the complete ELP lexicon yields orthographic neighborhood and bigram frequency measures that are computed using a larger 79,672 word corpus based on the HAL norms. In contrast, the restricted ELP lexicon computes these measures using the 40,481 word lexicon presented to the ELP participants.

**Query Options:**

<b>Query Scope:</b>	<b>Method of Distribution:</b>
<input checked="" type="radio"/> Restricted ELP Lexicon	<input type="radio"/> Browser
<input type="radio"/> The complete ELP Lexicon	<input checked="" type="radio"/> Email

**Figure 5. Options for method of distribution.**

Turning to ‘Method of Distribution’, the user can choose to view the query output directly on the browser or to receive it as an email attachment. It is usually better to ask for the output as an attachment via the Email option, given the large amount of information that is generated by the typical query.

f. Once these options are selected, click on the ‘Execute Query’ button (see above), and the website will ask for an email address if the Method of Distribution is via Email.

Output will be sent to the email address provided, in a comma-delimited file that can be opened in any text editor or spreadsheet program. If the Method of Distribution is via the Browser, the results will be displayed immediately on the screen.

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D
1	Word	Length	Log_Freq	NMorph
2	abash	5	0	1
3	wrack	5	3.76	1
4	yawl	4	3.4	1
5	yokel	5	3.97	1
6	whorl	5	3.37	1
7	ulna	4	2.77	1
8	true	4	0	1
9	thrum	5	3.18	1
10	tilth	5	2.89	1
11	tepee	5	2.77	1
12	tamp	4	3.4	1
13	swash	5	3.56	1
14	swum	4	3.83	1
15	strew	5	2.56	1

**Figure 6. List of words with specific lexical characteristics.**

g. The procedure for generating a list of nonwords with specific lexical characteristics is very similar. From the homepage, click on ‘Generate Lists of Nonwords with Specific Lexical Characteristics’, followed by ‘Search for nonwords matching your criteria’. The remaining steps are conceptually very similar to the instructions described above. Note that only a subset of lexical characteristics (length, orthographic neighborhood, and bigram frequency) are available for nonwords.

*2) Submit a list of stimuli to obtain lexical characteristics and behavioral performance measures for these stimuli*

This function allows users to submit lists of words and to obtain lexical characteristics and behavioral performance measures for these items. For example, a user may wish to obtain length, frequency counts, and naming latencies for a list of words.

a. Starting from the ELP homepage, click on ‘Generate Lists of Words with Specific Lexical Characteristics’.

b. Click on ‘Submit Lists of Items to generate Lexical Characteristics’.

c. The list of words can be either submitted as an ASCII text file (one word on each line), or be pasted or typed directly into a box on the browser. To submit a file, select ‘Filename Containing List of Words’. To do the latter, select ‘Free Text List of Words’.

### List Options:





#### Method of Submission:

- Free Text List of Words
- Filename Containing List of Words

**Figure 7. Options for submitting items.**

d. Next, select the fields of interest. To obtain length and frequency counts for the submitted stimuli, check all four fields under ‘General Fields’. In addition, check ‘I\_NMG\_Mean\_RT’ under ‘Behavioral Results: Naming’.

General Fields

- Length 
- Freq\_KF 
- Freq\_HAL 
- Log\_Freq\_HAL 

**Figure 8. Selecting all the General Fields.**

e. For ‘Query Scope’ and ‘Method of Distribution’, follow the instructions described in Step 1e.

f. Click on the ‘Execute Query’ button, which will either present a blank box to enter stimuli, or will ask for the path of the to-be-submitted text file, depending on the method of submission chosen in Step 2c.

g. The output can then be viewed directly on the browser or be received as an email attachment, depending on the method of distribution selected in Step 2e. For the latter option, note that *two* comma-delimited text files will be sent as attachments. One file provides lexical and behavioral characteristics for individual items, while the other file provides summary statistics of the selected characteristics for the submitted items.

	A	B	C	D	E	F	G	H
1	Total_Count	Distinct_Count	Invalid_Count	Mean_Length	Mean_Freq_KF	Mean_Freq_HAL	Mean_Log_Freq_HAL	Mean_I_NMG_Mean_RT
2	14	14	0	4.64	34	25.29	2.82	695.98

**Figure 9. Summary statistics of submitted words.**

	A	B	C	D	E	F	G
1	Occurrences	Word	Length	Freq_KF	Freq_HAL	Log_Freq_HAL	I_NMG_Mean_RT
2	1	abash	5	NULL	1	0	594.43
3	1	strew	5	NULL	13	2.56	777.81
4	1	swash	5	NULL	35	3.56	654.88
5	1	swum	4	1	46	3.83	752.48
6	1	tamp	4	1	30	3.4	665.24
7	1	tepee	5	NULL	16	2.77	768.2
8	1	thrum	5	NULL	24	3.18	690.53
9	1	tilth	5	1	18	2.89	725
10	1	true	4	231	0	0	699.79
11	1	ulna	4	NULL	16	2.77	740
12	1	whorl	5	NULL	29	3.37	788.33
13	1	wrack	5	1	43	3.76	626.92
14	1	yawl	4	1	30	3.4	651.68
15	1	yokel	5	2	53	3.97	608.4

**Figure 10. Lexical and behavioral characteristics of submitted words.**

h. The procedure for submitting a list of nonwords to obtain specific lexical characteristics is very similar. From the homepage, click on ‘Generate Lists of Nonwords with Specific Lexical Characteristics’, followed by ‘Submit lists of non-words in the database and obtain lexical characteristics’. The remaining steps are conceptually very similar to the instructions described above. For this query, note that only a subset of lexical characteristics are available for nonwords, and that submitted nonwords have to be in the ELP nonword database in order to generate an output. For nonwords not represented in the database, see Function 3 (described below).

### *3) Submit a list of arbitrary nonword stimuli to obtain lexical characteristics*

This function allows users to obtain lexical characteristics for lists of nonwords, whether these words are represented in the database or not. For example, a user may wish to obtain orthographic neighborhood and bigram frequency characteristics for a set of nonwords.

- a. Starting from the ELP homepage, click on ‘Generate Lists of Nonwords with Specific Lexical Characteristics’.
- b. Click on ‘Submit arbitrary lists of non-words in and generate lexical characteristics’.
- c. The list of nonwords can be either submitted as an ASCII text file (one word on each line), or be pasted directly into a box on the browser. To submit a file, select ‘Filename Containing List of Words’. To do the latter, select ‘Free Text List of Words’. For more details, follow the instructions described in Step 2c.
- d. Next, select the fields of interest. For our example, we will select all the available lexical characteristics.

**Submit Arbitrary Lists of Non-Words to generate Lexical Characteristics**

Generates lexical characteristics for a user submitted list. This list can either be a text file, or can be pasted within a text block on this form.

---

**List Options:**


**Method of Submission:**

Free Text List of Words  
 Filename Containing List of Words


---


**Please select which fields you would like returned:**

General Fields


Length 


Orthographic Neighborhood


Ortho\_N 

Neighborhood Button 

Bigram Frequency

BG\_Sum 

BG\_Mean 

BG\_Freq\_By\_Pos 

**Figure 11. Interface for submitting novel nonwords.**

- e. For ‘Query Scope’ and ‘Method of Distribution’, follow the instructions described in Step 1e.

f. Click on the ‘Execute Query’ button, which will either present a blank box to enter stimuli, or will ask for the path of the to-be-submitted text file, depending on the method of submission chosen in Step 3c. Once these options are selected, click on the ‘Execute Query’ button, and the website will ask for an email address, if Method of Distribution is Email. Output will be sent to the email address provided, in a comma-delimited file that can be opened in any text editor or spreadsheet program. Again, if Method of Distribution is browser, the data will be immediately displayed on screen.

	A	B	C	D	E	F	G
1	Total_Count	Distinct_Count	Mean_Length	Mean_Ortho_N	Mean_BG_Sum	Mean_BG_Mean	Mean_BG_Freq_By_Pos
2	26	26	4.96	3.58	5774.69	1436.96	1171.04

**Figure 12. Summary statistics of submitted nonwords.**

	A	B	C	D	E	F	G
1	Occurrences	Word	Length	Ortho_N	BG_Sum	BG_Mean	BG_Freq_By_Pos
2	1	LEMM	4	0	4809	1603	731
3	1	BOUD	4	4	2796	932	947
4	1	DRAS	4	5	4958	1652.67	1308
5	1	SOID	4	6	1978	659.33	518
6	1	FLOD	4	8	2477	825.67	748
7	1	DALT	4	9	3969	1323	853
8	1	LARN	4	9	5457	1819	1433
9	1	SOIN	4	9	8027	2675.67	846
10	1	SOUT	4	14	3399	1133	958
11	1	SPRUD	5	0	3194	798.5	581
12	1	HOOVE	5	0	4562	1140.5	1170
13	1	THRID	5	0	5293	1323.25	1000
14	1	POPLE	5	0	6019	1504.75	1208
15	1	SHOIR	5	1	3648	912	928

**Figure 13. Characteristics of submitted nonwords.**

*4) Provide access to individual participants’ data.*

This function provides access to the raw lexical decision and naming latencies.

a. Starting from the ELP homepage, click on ‘Access Lexical Decision and Naming Data’.

b. Depending on whether the user is interested in retrieving lexical decision data or naming data, click on ‘Access Trial-Level Lexical Decision Data’ or ‘Access Trial-Level Naming Data’, respectively. In the following example, lexical decision trial data for the first five participants will be retrieved, with only accurate word trials selected.

c. The previous step brings the user to the page depicted below. Clicking on the ‘+’ icon expands General Fields. For our purposes, check all the options (descriptions provided below) available.

### Access Lexical Decision Data

The following query allows one to display Lexical Decision data collected to date. You will enter any desired constraints, and all data meeting those criteria will be returned.

---

**Please select the fields and any constraints to be used in the query:**

General Fields

- Sub\_ID
- Trial
- Type
- D\_Accuracy
- D\_RT
- D\_word
- Outlier
- D\_Zscore

**Figure 14. Interface for accessing individual trial-level data.**

- c. A number of constraints can be set by the user.
- i. *Sub\_ID* specifies the participants whose data are to be downloaded.
  - ii. *Trial* specifies the trials (e.g., first 100 trials) of interest
  - iii. *Type* specifies *lexicality* (words=1; nonwords=0); only applicable to lexical decision
  - iv. *D\_Accuracy* specifies trial accuracy (correct=1; error=0)
  - v. *D\_RT* specifies trial response times (in ms)
  - vi. *D\_word* refers to the specific items presented on a trial
  - vii. *Outlier* specifies whether a trial is an outlier (outlier=True; non-outlier=False)
  - viii. *D\_Zscore* specifies Z-transformed response times

d. Click on the constraint buttons for ‘Sub\_ID’, enter 1 as the ‘Minimum value’ and 5 as the ‘Maximum value’. This selects the first five participants.

Please enter any desired constraints upon the SUB\_ID or leave blank if non are desired:

Attribute	Minimum Value	Maximum Value
SUB_ID	<input type="text" value="1"/>	<input type="text" value="5"/>

Figure 15. Placing constraints on the subject ID variable.

e. Following this, for ‘Type’, enter 1 for both minimum and maximum values (selects only words). For ‘D\_Accuracy’, enter 1 for both minimum and maximum values (selects only accurate responses).

f. Finally, to have the data sent to you, under query options, go to ‘Method of Distribution’ in the ‘Query Options’ box. The user selects whether the data are presented on the browser or sent as an email attachment. Given the typically large size of the data files, the Email option is recommended.

**Query Options:** Method of Distribution:

Browser  
 Email

Figure 16. Options for method of distribution.

	A	B	C	D	E	F	G	H
1	Sub_ID	Trial	Type	D_Accuracy	D_RT	D_word	Outlier	D_Zscore
2	3	5	1	1	721	butting	0	-0.31935
3	3	10	1	1	797	abbreviated	0	-0.11191
4	3	11	1	1	734	crackle	0	-0.28387
5	3	12	1	1	954	confiding	0	0.316624
6	3	13	1	1	810	director's	0	-0.07643
7	3	14	1	1	681	appeals	0	-0.42853
8	3	15	1	1	860	postmaster's	0	0.060049
9	3	16	1	1	598	videotape	0	-0.65508
10	3	17	1	1	594	plus	0	-0.666
11	3	19	1	1	647	incise	0	-0.52134
12	3	20	1	1	752	godchild	0	-0.23474
13	3	23	1	1	804	inspected	0	-0.0928
14	3	25	1	1	666	giggles	0	-0.46948
15	3	28	1	1	1875	katapuzus	0	0.694000

Figure 17. Trial-level data for a typical participant.

## Discussion

In this paper, we have described the English Lexicon Project (ELP), a web-based repository encompassing descriptive and behavioral measures for over 80,000 items, based on data collected from nearly 1300 participants at six different universities. We have presented the methods used to collect the data, defined the variables included in the database, and illustrated how the search engine facilitates access to the ELP's many resources.

One reservation researchers might have about the ELP is that naming or making lexical decisions to 3,000 items would produce variability due to fatigue or boredom. If this were the case, then one might worry that these data sets will be relatively noisy and have limited utility. Hence, traditional word recognition studies involving 100 to 200 words may produce a better reflection of normal lexical performance. There are a number of responses to this potential concern. First, Spieler and Balota (1997) demonstrated that four simple predictor variables (log frequency, neighborhood density, orthographic length, and phonological onset) accounted for 42% of the item level variance in a large-scale behavioral dataset of naming performance. Thus, there is a considerable amount of predictable variance in these data sets. Second, although the estimates for proportion of variance accounted for fluctuate slightly across different large data sets, one observes the same pattern of predictive power in the data sets obtained by Seidenberg and Waters (1989), Spieler and Balota (1997), and Treiman et al. (1995). Finally, Balota and Spieler (1998) selected the same items from a large-scale study that were used in a standard factorial study conducted by Taraban and McClelland (1987, Experiment 1a). The results were remarkably similar to the pattern reported by Taraban and McClelland (see Balota & Spieler, 1998, Figure 1).

*Potential Uses for the Website.* We envision this database being used in a number of different types of research endeavors. First, as noted, most computational models are based on single-syllable words, and the vast majority of studies of lexical processing have only included single-syllable words, and so this literature has produced a relatively restricted account of lexical processing. The present database will afford a database to extend these models to multisyllabic (and multimorphemic) stimuli.

Second, these data will be important for researchers interested in targeting particular variables. In some cases, experiments might be replaced by accessing the database. As described above, the database can be easily accessed to provide naming and lexical decision data for any set of words and/or nonwords in the database. Because of the size of the database, one can resample different sets of items with particular characteristics from the database. In other situations, researchers may use the database as a preliminary step in developing further follow-up experiments to test the influence of a given variable. The advantage of the present database is that it will minimize the idiosyncratic effects that could arise because of specific selection of items or list contexts, described earlier. We believe that the standardization of this dataset will be helpful in minimizing the inconsistencies that have arisen in the word recognition literature. One might be concerned that there may be a bias for researchers to search for items that actually produce an effect in the data. Of course, this is also a problem in the current word recognition literature. Ultimately, the peer review process will be responsible for evaluating the potential for this problem in any given paper.

Third, researchers from other disciplines outside of word recognition (e.g., memory, perception, neuroimaging, and neuropsychology) will be able to use this database to select items that are equated along a number of relevant dimensions. The ELP can generate word

and nonword lists based on behavioral data from the naming and lexical decision results, or a myriad of descriptive characteristics. As an example, one could easily generate a list of words that produce lexical decision latencies within the range of 800-850 ms, are 12 letters in length, and have 5 orthographic neighbors. This database will also be important in the area of memory and text comprehension, wherein, researchers want to look at effects of variables above and beyond differences in isolated lexical processing. For example, one could equate the lexical processing characteristics across different texts at the word level, as reflected by naming and lexical decision performance obtained from the ELP, to investigate the influence of higher-level integration processes. Finally, this data set could be useful in obtaining baselines across counterbalanced lists, which ultimately will reduce the error variance in standard factorial designs that counterbalance targets across conditions.

In summary, the ELP is potentially useful for equating items across a wide variety of research enterprises that use words as the primary stimulus. Obviously, for a number of reasons, the lexical characteristics included in the ELP are not, and were not meant to be, exhaustive. For example, metrics that reflect spelling-sound relationships and semantic characteristics are not included, and these measures, along with others, have been shown to be relevant in word recognition performance (see Balota et al., 2004, for a review).

Nevertheless, we believe the variables available represent a good start.

*Examples of ELP applications in the recent literature.* Although the ELP has been on-line for a relatively short time, it has already been fruitfully employed in published papers by different researchers for disparate purposes. For example, in Balota et al.'s (2004) megastudy, the ability of different variables to predict naming and lexical decision latencies for 2,428 monosyllabic words was investigated. In order to test the replicability from this

study, the same 2,428 words were extracted from the ELP. The results provided a remarkably clear replication of the predictive power of the same set of variables (see Balota et al., 2004, Figures 10 and 11). Of course, this also is relevant to the discussion above regarding the reliability of data from studies wherein participants produce responses to 2,000 to 3,000 words and/or nonwords. In addition, New, Ferrand, Pallier, and Brysbaert (2005) examined length effects using lexical decision latencies for 33,006 items from the ELP. They observed an interesting quadratic relationship between length and lexical decision performance, with facilitatory effects (faster latencies for longer words) for shorter words, null effects for medium-length words, and inhibitory effects (slower latencies for longer words) for longer words. This may explain why null length effects are typically observed in factorial lexical decision experiments, where medium-length stimuli are most often used. Finally, Kello (2006) has recently developed a junction model of word reading, which employs different modes (lexical and sublexical) of processing. The most interesting aspect of this model is that it can simulate response times for multisyllabic words. Kello has so far tested the model against 30,000 response times from the ELP. Intriguingly, the model was able to account for a substantial proportion of variance in both lexical decision and naming performance, after articulatory characteristics and length were partialled out. We anticipate that these initial published demonstrations of the utility of the ELP will markedly increase in the future, and we hope that the database will be helpful for researchers across diverse areas in cognitive science.

## References

- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, 4, 439-461.
- Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multi-trace memory model of polysyllabic word reading. *Psychological review*, 105, 678-723.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Balota, D. A. & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception & Performance*, 10, 340-357.
- Balota, D. A. & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory & Language*, 24, 89-106.
- 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.
- Besner, D. & Bourassa, D.C. (1995, June). *Localist and parallel processing models of visual word recognition: A few more words*. Paper presented at the Brain Behavior, and Cognitive Science Society, Halifax CA.
- Cattell, J.M. (1886). The time it takes to see and name objects. *Mind*, 11, 63-65.
- Centre for Speech Technology Research, University of Edinburgh. (n.d.). *Unisyn lexicon* [Data file]. Retrieved July 19, 2004, from <http://www.cstr.ed.ac.uk/projects/unisyn/>

- Chateau, D., & Jared, D. (2003). Spelling-sound consistency effects in disyllabic word naming. *Journal of Memory & Language, 48*, 255-280.
- Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). Hillsdale, NJ: Erlbaum.
- 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.
- Cutler, A. (1981). Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition, 10*, 65-70.
- Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125*, 777-799.
- Forster, K.I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition, 28*, 1109-1115.
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103*, 518-565.
- Horne, J. A. & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology, 4*, 97- 110.
- Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory & Language, 29*, 687-715.
- Keefe, D. E. & Neely, J. H. (1990). Semantic priming in the pronunciation task: The role of prospective prime-generated expectancies. *Memory & Cognition, 18*, 289-298.

- Kello, C. T. (2006). Considering the junction model of lexical processing. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing*, Psychology Press.
- Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory & Language*, 47, 145-171.
- Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Lund, K., & Burgess, C. (1996) Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments & Computers*, 28, 203-208.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375-407.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, 106, 226-254.
- Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 15, 1003-1019.
- New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2005). Re-examining word length effects in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review* (in press).

- Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1989). Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience, 1*, 153-170.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.
- Seidenberg, M. S., & Waters, G. S. (1989, November). *Naming words aloud: A mega-study*. Paper presented at the meeting of the Psychonomic Society, Atlanta, GA.
- Shipley, W.C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology, 9*, 371-377.
- Spieler, D.H., & Balota, D.A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8*, 411-416.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.
- Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory & Language, 26*, 608-631.
- Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General, 124*, 107-136.
- Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory & Language, 47*, 1-29.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud?

A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception & Performance*, 24, 1131-1161.

### Acknowledgments

This research was supported by Grant AGO3991 from the National Institute on Aging and Grant BCS 0001801 from the National Science Foundation. Thanks are extended to Greg Burgess and Mark Law for programming and help in the early planning stages of the project, and to Christine Abadire, Melissa Davis, April Fugett, Leilani Goodman, Jeff Johnson, Lawrence Locker, Jr., Susan Marshall, Tameka Roberson, Jeff Templeton, and Mark Yates for help in various aspects of collecting the data. Thanks are also extended to Curt Burgess for providing the word frequency estimates from the HAL database, and Stephen Isard and Susan Fitt for allowing access to the Unisyn Lexicon for the phonological information available on the ELP website.

Table 1: Means and standard deviations of participant demographics for the speeded naming and lexical decision tasks.

	Speeded Naming (n = 444)		Lexical Decision (n = 816)	
	Mean	SD	Mean	SD
<b>Age</b>	23.51	9.31	22.86	6.85
<b>Education (no. of years)</b>	14.73	1.75	14.76	1.72
<b>Shipley Vocabulary Age</b>	17.59	1.28	17.39	1.35
<b>Mean RT</b>	735.06	106.69	787.86	164.83
<b>Percent Errors</b>	8.41	4.77	14.40	6.90
<b>*a. Uncertain of pronunciation</b>	3.81	2.90		
<b>*b. Mispronunciation</b>	2.91	2.12		
<b>*c. Voice key problem</b>	1.09	1.18		
<b>*d. Time out</b>	0.60	2.20		
<b>Percent Outliers</b>	4.90	2.96	3.57	3.74

\* These coding categories are only applicable to the speeded naming task, where participants coded their own response.

Table 2: Descriptive and Behavioral Data for the Words used in the ELP.

<b>General Fields</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>SD</b>
Length	1	21	8.00	2.46
K & F frequency	1	69,971	29.73	557.64
HAL frequency	0	23,099,033	10,778.67	192,226.37
Log HAL frequency	0	17	6.16	2.40
<b>Orthographic Neighborhood</b>				
Type orthographic neighborhood size	0	25	1.29	2.73
Token orthographic neighborhood size	0	16	7.02	1.88
Type neighborhood frequency (Greater)	0	21	1.69	2.32
Token neighborhood frequency (Greater)	1	16	8.42	1.93
Type neighborhood frequency (Less)	0	23	1.69	2.41
Token neighborhood frequency (Less)	0	16	5.89	1.74
<b>Bigram Frequency</b>				
Summed bigram frequency	11	59,803	14,313.09	7,848.76
Average bigram frequency	5.5	6,910	1,974.56	729.99
Summed bigram frequency by position	1	6,603	2,149.21	1,070.14
<b>Phonological Characteristics</b>				
Pronunciation	NA			
Number of Phonemes	1	17	6.70	2.28
Number of Syllables	1	8	2.54	1.10
<b>Morphological Characteristics</b>				
Morpheme Parse - Letters	NA			
Morpheme Parse -Phonemes	NA			
Number of Morphemes	1	7	2.11	-.86
<b>Part of Speech</b>				
Part of Speech	NA			
<b>Behavioral Results: Lexical Decision</b>				
Mean LDT latency (Raw)			784.07	134.93
Mean LDT latency (Standardized)			.00	.46
Standard Deviation of LDT latencies			277.51	92.39
Number of Observations			27	6.92
Mean Accuracy			.84	.20
<b>Behavioral Results: Naming</b>				
Mean Naming latency (Raw)			722.58	109
Mean Naming latency (Standardized)			-.04	.46
Standard Deviation of Naming latencies			178.47	69.30
Number of Observations			24.00	4.04
Mean Accuracy			.93	.11

Table 3. Descriptive and Behavioral Data for the Nonwords used in the ELP.

<b>General Fields</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>SD</b>
Length	1	21	8.00	2.46
<b>Orthographic Neighborhood</b>				
Type orthographic neighborhood size	0	24	1.78	2.22
<b>Bigram Frequency</b>				
Summed bigram frequency	11	56,657	13,562.11	7,537.63
Average bigram frequency	10.5	5,197	1,867.73	714.78
Summed bigram frequency by position	0	6,985	1,936.96	952.82
<b>Behavioral Results: Lexical Decision</b>				
Mean LDT latency (Raw)			855.78	113.75
Mean LDT latency (Standardized)			-.24	.36
Standard Deviation of LDT latencies			299.02	85.43
Number of Observations			28.72	4.87
Mean Accuracy			.88	.13