Journal of Experimental Psychology: General
2004, Vol. 133, No. 2, 283–316

0096-3445/04/$12.00 DOI: 10.1037/0096-3445.133.2.283

Visual Word Recognition of Single-Syllable Words

David A. Balota
Washington University

Michael J. Cortese
College of Charleston

Susan D. Sergent-Marshall
Washington University

Daniel H. Spieler
Georgia Institute of Technology

Melvin J. Yap
Washington University

Speeded visual word naming and lexical decision performance are reported for 2,428 words for young adults and healthy older adults. Hierarchical regression techniques were used to investigate the unique predictive variance of phonological features in the onsets, lexical variables (e.g., measures of consistency, frequency, familiarity, neighborhood size, and length), and semantic variables (e.g., imageability and semantic connectivity). The influence of most variables was highly task dependent, with the results shedding light on recent empirical controversies in the available word recognition literature. Semantic-level variables accounted for unique variance in both speeded naming and lexical decision performance, with the latter task producing the largest semantic-level effects. Discussion focuses on the utility of large-scale regression studies in providing a complementary approach to the standard factorial designs to investigate visual word recognition.

The study of the processes involved in isolated word recognition has been central to developments in experimental psychology since the days of Cattell (1886). Researchers have accumulated a vast amount of information regarding the statistical properties of words, including word frequency, subjective familiarity, meaningfulness, letter frequency, bigram frequency, trigram frequency, spelling-to-sound consistency, syntactic class, and concreteness (see Balota, 1994, and Henderson, 1982, for reviews). Word recognition research has been critical in developing computational models (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; McClelland & Rumelhart, 1981; Plaut, McClelland, Seidenberg, & Patterson, 1996), distinguishing between automatic and attentional processes (e.g., Fodor, 1983; Neely, 1977), providing insights into reading acquisition (e.g., Perfetti, 1994), and understanding neural substrates of language processing (e.g., Coltheart, Patterson, & Marshall, 1980; Petersen, Fox, Posner, Mintun, & Raichle, 1988). One might argue that the word has been as central to developments in cognitive psychology and psycholinguistics as the cell has been to biology.

Given the importance of word recognition research, one might assume that there are well-accepted methods for studying lexical processing. For example, probably the best way to study the integration of lexical information within reading is to analyze people's eye movements (e.g., eye fixation and gaze durations) as they are reading text (see Rayner, 1998; Rayner & Pollatsek, 1989). However, during reading, there are multiple sources of information available (e.g., syntactic information, semantic constraints, parafoveal visual information), and so there are limits to this approach for models of isolated word recognition. Another procedure is to study how subjects identify words that are visually degraded by brief presentations and pattern masking. Unfortunately, there are also limitations with this procedure. Specifically, when subjects receive a degraded stimulus, they may rely on general knowledge about frequency and spelling patterns of words to make sophisticated guesses about the target stimulus (e.g., Broadbent, 1967; Catlin, 1973).

Because of the above concerns, researchers have continued to rely heavily on two measures: speeded lexical decision and naming performance. In the lexical decision task (LDT), subjects are presented with a visual string (either a word or a nonword, e.g., *flirp*), with their task being to decide as quickly as possible

David A. Balota, Susan D. Sergent-Marshall, and Melvin J. Yap, Department of Psychology, Washington University; Michael J. Cortese, Department of Psychology, College of Charleston; Daniel H. Spieler, School of Psychology, Georgia Institute of Technology.

This work was supported by Grants AGO3991 and RO1 AG17024 from the National Institute on Aging and Grant BCS 0001801 from the National Science Foundation.

We thank Keith Hutchison, Mark Law, Maura Pilotti, Martha Storandt, Michael Strube, and Jeff Templeton for their assistance in various stages of this project. In addition, we thank Brett Kessler, Rebecca Treiman, Barbara Juhasz, and Keith Rayner for helpful comments on an earlier version of this article. Finally, we thank Mark Steyvers for providing the connectivity estimates, Doug Nelson for providing the semantic set size estimates, Curt Burgess for providing the frequency estimates from the Hyperspace Analogue to Language database, and Brett Kessler and Rebecca Treiman for providing both their consistency estimates and their naming data.

All item-level data are available at http://www.artsci.wustl.edu/~dbalota/labpub.html.

Correspondence concerning this article should be addressed to David A. Balota, Department of Psychology, Box 1125, Washington University, One Brookings Drive, St. Louis, MO 63130. E-mail: dbalota@artsci.wustl.edu

whether the string is a word or nonword. In the speeded naming task, subjects are presented with a visual word (or sometimes a nonword) and are asked to name the word aloud as quickly and as accurately as possible. These two tasks are clearly the major driving force in isolated word recognition research and have been the gold standard in developing computational models of lexical processing (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1996; Seidenberg & McClelland, 1989; Zorzi, Houghton, & Butterworth, 1998).

In speeded naming and lexical decision studies, researchers typically have used 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 fewer than 20 items per cell). Mean latency and accuracy are calculated for each subject across items (or for each item across subjects in some studies) and then entered into an analysis of variance (ANOVA), and the effects of factors are measured. A reliable influence of a factor is typically interpreted as being consistent or inconsistent with a given model. Although this approach has been fruitful in identifying important variables that modulate speeded lexical decision and naming performance, it has some potential difficulties. We believe that these difficulties may diminish the rate of accumulation of knowledge in the field and may lead to counterproductive controversies regarding the presence or absence of an effect of a targeted variable. We now turn to some of these difficulties.

First, it is quite difficult to select a set of items that only vary on one categorical dimension. Cutler (1981) argued that because so many factors have been identified in word recognition research, it is virtually impossible to select a sufficient number of items in all cells of a factorial design. Cutler also suggested that the literature contains a number of such failures to control for relevant factors, and these failures have led to a number of false starts in theoretical developments. Consider the influence of spelling-to-sound correspondences, for example, the fact that *pint* is not produced according to common spelling-to-sound principles, whereas *hint* is consistent with such principles. The influence of spelling-to-sound correspondence depends on a number of factors, such as the frequency of the target word, the number and frequency of words with similar spelling-to-sound correspondences (friends), the number and frequency of words with different spelling-to-sound correspondences (enemies), and probably a host of other variables (Jared, McRae, & Seidenberg, 1990; Plaut et al., 1996; Stone, Vanhoy, & Van Orden, 1997). Plaut et al. have argued that it is best to consider a variable such as consistency as a continuous factor as opposed to a categorical variable as in the standard ANOVA design. The ultimate problem here is that it has been difficult to reach definitive answers regarding the influence of factors from categorical studies in word recognition with a relatively small set of items without introducing potentially contaminating factors (e.g., consider the recent controversy regarding backward consistency effects in studies by Peereman, Content, & Bonin, 1998, and Ziegler & Ferrand, 1998). Hence, one might argue that it is time to go beyond arguing about the presence or absence of a given effect of a categorical variable on the basis of a relatively limited sample of items that could potentially vary on a number of continuous dimensions.

Second, Forster (2000) has recently pointed out that word recognition researchers may have implicit knowledge regarding lexical variables and that this knowledge could influence the inferences drawn from experiments. Forster has demonstrated this by asking expert researchers in word recognition to make lexical processing predictions for pairs of words. Specifically, on each trial, these researchers were asked to predict which of two words would produce faster lexical decision performance. The expert word recognition researchers could make such predictions above and beyond standard predictor variables, such as word frequency. If researchers can make such predictions implicitly or explicitly, it is possible that when they select items for their categorical manipulations and have a hypothesis in mind, this could influence the results (see Rosenthal, 1995). Thus, Forster suggested that a better approach would be to randomly select words from a much larger set of items that have the targeted characteristics.

A third concern about the standard factorial experiments is that list contexts (i.e., the characteristics of words within a list) often vary across experiments reported in the literature. This is likely due to the fact that researchers naturally load their lists with items that have extreme values along the targeted factor dimensions; for example, half of the words may have irregular spelling-to-sound correspondences. Hence, subjects may become either implicitly primed or even explicitly sensitive to the factor being manipulated. There are many demonstrations of list-context effects in the literature. For example, Seidenberg, Waters, Sanders, and Langer (1984) demonstrated that the influence of spelling-to-sound correspondence was sensitive to the presence of other similarly spelled words within the list (also see Lupker, Brown, & Colombo, 1997; Monsell, Patterson, Graham, Hughes, & Milroy, 1992; Zevin & Balota, 2000). Glanzer and Ehrenreich (1979) and Gordon (1983) have demonstrated that simple word-frequency effects can be modulated by the relative proportion of high-frequency and low-frequency words within the lexical decision experiment. Andrews (1997) has suggested that the inconsistencies across studies of orthographic neighborhood size effects in lexical decision could be due to differences in lexical decision strategies induced by unusual stimulus list environments. Although list-context effects can be of interest, unwanted list-context effects could be minimized if subjects were exposed to a sample of items that were not selected on the basis of fitting factorial designs.

A fourth potential problem with standard factorial designs involves a concern about categorizing continuous variables. Consider word frequency. Typically, researchers investigate high- versus low-frequency words as opposed to using frequency as a continuous variable in a regression model. Of course, this problem extends to virtually all variables that researchers have investigated as categorical variables. Moreover, this concern extends to other areas of cognitive psychology, such as memory and attention, wherein continuous variables are treated as categorical variables. Statisticians have historically pointed out that categorizing continuous variables can lead to a decrease in statistical power and reliability (see, e.g., Cohen, 1983; Humphreys, 1978; Maxwell & Delaney, 1993). This work has typically focused on between-subject variability, where researchers often categorize individual characteristics (e.g., age might be categorized as young vs. old). MacCallum, Zhang, Preacher, and Rucker (2002) have recently reported a review of the literature, along with a series of simulations, which nicely demonstrated that with a relatively small number of observations, such categorization can decrease reliability and lead to the inappropriate rejection of the null hypothesis. These concerns naturally extend to between-items manipulations in word recognition studies.

A fifth potential problem is that the field has emphasized the search for significant effects for a specific set of stimuli without taking into account the more general implications for the lexical processing system. For example, if one obtains a reliable interaction among three factors in a 2 × 2 × 2 design, does one want to argue that this is a general reflection of lexical processing, or is it possible that this interaction is limited to the selected set of 80 words used in such a design (assuming 10 words per each of the 8 cells)? The search for a significant effect does not typically motivate researchers to report the amount of unique variance that a given factor accounts for in a design. This latter information may ultimately be more important than the more complex effects that reach the magical significance level. This was demonstrated by Spieler and Balota (1997), who found a surprisingly large influence of length in letters (4.4% unique variance, compared with 6.3% for log frequency and 2.2% for orthographic neighborhood size) on speeded naming performance in their study of 2,870 single-syllable words. Although the theoretical interpretation of this effect is still being discussed (see Balota & Spieler, 1998; Seidenberg & Plaut, 1998), these results may be more supportive of a serial analysis (see Coltheart et al., 1993) than a parallel analysis in speeded word naming. The point here is that the driving force in this literature should no longer be if a variable has an impact on lexical processing: It should also include consideration of how much of a contribution that variable makes toward lexical processing.

There have recently been some initial examinations of speeded naming performance on large sets of English words (e.g., Balota & Spieler, 1998; Besner & Bourassa, 1995; Kessler, Treiman, & Mullennix, 2002; Spieler & Balota, 1997; Trieman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). For example, Spieler and Balota had 31 subjects name all 2,820 single-syllable words that both the Seidenberg and McClelland (1989) and the Plaut et al. (1996) models were trained on. The results were very informative: Although the computational models did an excellent job of accommodating aspects of the data obtained from standard factorial experiments, these models appeared to have some limitations when it came to accounting for individual item-level variance. For example, log frequency alone accounted for 7.3% of the variance, whereas the error scores from the Seidenberg and McClelland (1989) model and the settling times from the Plaut et al. (1996) model accounted for 10.1% and 3.3% of the variance, respectively. This same general pattern of results was replicated with a group of healthy older adults (see Balota & Spieler, 1998).

The approach taken in the present study was to compare naming and lexical decision latencies on a large corpus of stimuli (all monosyllabic English words in the Kučera & Francis, 1967, norms) in order to obtain estimates of the unique variance predicted by an extended set of targeted variables. This set of items was the focus of our study because these words have consistently been the target of computational models of speeded word naming and lexical decision performance (e.g., Seidenberg & McClelland, 1989). We used regression techniques to control for the influence of contaminating variables and allowed the language, instead of the experimenter, to define the stimulus set. We selected the following targeted variables from the extant literature to investigate: phonological onsets, length in letters, orthographic density, objective frequency, subjective frequency, feedforward onset consistency, feedforward rime consistency, feedback onset consistency, feedback rime consistency, imageability, meaningfulness, number of associates, and estimates of semantic connectivity. We focused on these variables because of their theoretical importance in available models and because of the controversies that these variables have produced in the available literature.[1] In addition, we decided to consider a set of limited variables to avoid problems associated with suppressor variables. We discuss additional variables in the General Discussion section.

We have a number of goals for the present research. First, the large database of naming and lexical decision latencies obtained in this study affords a comparison of the predictive power of five different measures of word frequency (see the *Comparison of Word-Frequency Estimates* section for a description of the five measures of word frequency). An initial set of analyses will identify the best word-frequency measure, and then this measure will be used in subsequent regression analyses. As described below, there is considerable difference in the predictive power of different word-frequency measures (see also Burgess & Livesay, 1998; Zevin & Seidenberg, 2002).

A second goal of the present work is to test predictions regarding the differential effects of specific variables on lexical decision versus naming. For example, we anticipate that word frequency should have a greater influence on lexical decision than on naming performance. Such a prediction follows from the simple observation that the LDT places more of an emphasis on frequency-based information in making the word–nonword discrimination (e.g., Balota & Chumbley, 1984; Besner & Swan, 1982), whereas the naming task emphasizes the onset of the appropriate articulation. However, we expect that effects of spelling-to-sound consistency should be greater in naming than in lexical decision because naming requires the use of phonological information, whereas the LDT does not place the same premium on this information (e.g., Cortese, 1998). We also expect semantic variables to have a greater influence on lexical decision than on naming. A number of lexical decision studies have shown an influence of meaning-based

[1] The selection of variables to enter into the regression analyses was based on (a) a variable's unique status in the available literature, (b) the lack of redundancy with variables that were included, and (c) availability of norms for a large set of items. For example, we did not include in the regression analyses age of acquisition as a predictor variable because age-of-acquisition norms were available for only about 25% of the items. Moreover, there has been some recent controversy regarding the status of this variable in predicting performance above and beyond cumulative frequency (see Zevin & Seidenberg, 2002). We also excluded variables such as bigram frequency and orthographic neighborhood frequency because initial analyses indicated that these variables were not related to any of the dependent measures and, in the case of bigram frequency, there have been repeated failures to demonstrate an influence of this variable (see, e.g., Andrews, 1992; Treiman et al., 1995). Although the final analysis included consistency measures that were based on token estimates (based on frequency-weighted counts of friends and enemies) instead of type estimates (based on simple counts of friends and enemies), it is noteworthy that the same pattern of significant effects of consistency were observed with type counts. Finally, as noted in the General Discussion section, to explore alternative accounts of the consistency effects, we included the spelling frequency of the onset and rime units in Step 2 of the regression analyses, and the inclusion of these variables did not alter the results.

variables (e.g., Chumbley & Balota, 1984; James, 1975), whereas semantic effects appear to be restricted to the naming of low-frequency irregular words (e.g., Cortese, Simpson, & Woolsey, 1997; Strain, Patterson, & Seidenberg, 1995). As discussed below, an intriguing issue is whether one can detect semantic effects in a large-sample study of speeded naming performance after other factors have been controlled.

A third goal is to compare the performance of young and older adults. The question here is how the lexical processing system changes with an additional 50 years, on average, of practice with words, along with the accompanying cognitive changes that occur in older adults. Regarding word naming, Spieler and Balota (2000) have shown that word frequency has more predictive power for older adults than for young adults, whereas orthographic neighborhood size has more predictive power for young adults than for older adults. As discussed later, this pattern could be due to cohort biases in the standard word-frequency norms. It is also possible that semantic variables will have differential predictive power for young and older adults. The finding of larger frequency effects for older adults in naming suggests the possibility that connections between orthography and semantics (i.e., a direct route to meaning) may become stronger with age. However, the novel task demands of lexical decision may shift the focus from phonological conversion to familiarity-based information. If semantic information is incorporated into a word's perceived familiarity and older adults are less likely to engage the specific task demands of the LDT (see Balota & Faust, 2001), then young adults may be more likely to tap into this source of information than are older adults. This would result in stronger semantic effects for young adults than for older adults in lexical decision performance.

Finally, the present study affords a database for researchers to evaluate models and constrain their development. In addition, researchers interested in areas such as memory, perception, and neuropsychology will be able to use this database to select items that are equated along a number of descriptive dimensions, such as frequency, familiarity, orthographic neighborhood size, and bigram frequency, and also on behavioral measures of mean naming and/or lexical decision latencies. This is the first step in making available even larger databases (see the English Lexicon Project [ELP] Web site at http://elexicon.wustl.edu/ for a database for over 40,000 words and nonwords).

Method

Subjects

Thirty young adults (mean age = 20.5 years, $SD = 2.0$) and 30 older adults (mean age = 73.6 years, $SD = 5.1$) participated in the lexical decision study. The young adults averaged 14.9 years of education ($SD = 1.6$) and scored an average of 34.5 ($SD = 2.5$) on the Shipley vocabulary subtest (Shipley, 1940). The Shipley vocabulary subtest is a four-alternative multiple-choice vocabulary test with a maximum score of 40. The older adults averaged 15.1 years of education ($SD = 2.4$) and scored an average of 35.8 ($SD = 2.6$) on the Shipley vocabulary subtest. As described in Spieler and Balota (2000), 31 young adults (mean age = 22.6 years, $SD = 5.0$) and 29 older adults (mean age = 73.4 years, $SD = 3.0$) performed the naming task. The young adults averaged 14.8 years of education ($SD = 2.0$) and scored an average of 35.1 ($SD = 2.7$) on the Shipley vocabulary subtest. Older subjects averaged 15.7 years of education ($SD = 2.8$) and scored an average of 37.1 ($SD = 3.0$) on the Shipley vocabulary subtest. There were no reliable age differences in education (both ts < 1.44), but older adults did have higher vocabulary scores, $t(58) = 1.97$, $p = .05$, and $t(58) = 2.72$, $p < .05$, in the LDT and naming task, respectively. Young adults were recruited from the undergraduate population of Washington University, whereas the older adults were recruited from the Aging and Development Subject Pool at Washington University. Subjects were paid $40 for participation in the lexical decision study and $20 for participation in the naming study. The difference in payment was due to the fact that the LDT was nearly twice as long as the naming task.

Stimuli

The stimuli for the LDT consisted of 2,906 monosyllabic words and 2,906 length-matched pronounceable nonwords. Each nonword for the LDT was constructed by changing from 1 to 3 letters in a corresponding word. The words and nonwords were matched in length and ranged between 2 and 8 letters in length. The words for the naming task consisted of 2,870 monosyllabic words used as the training corpora for the connectionist models of Seidenberg and McClelland (1989) and Plaut et al. (1996). The words ranged in frequency from 0 to 69,971 per million (Kučera & Francis, 1967).

Apparatus

An IBM-compatible computer was used to control the display of the stimuli and to collect subjects' responses. Display of all stimuli was synchronized with the vertical retrace of the monitor to measure response latencies to the nearest millisecond. The stimuli were displayed on a 14-in. VGA monitor. A Gerbrands Model G1341T voice-operated relay interfaced with the computer served to collect naming latencies.

Procedure

LDT

Each individual participated in two sessions of equal length on separate days within a period of 7 days. Subjects, seated in front of a computer, were told that a single letter string would appear in the center of the computer screen and that their task was to silently read each string, decide whether it was a word or nonword, and indicate their decision by a keyboard button press. Subjects were instructed to be as fast as possible while minimizing errors.

Each trial consisted of the following sequence of events: (a) A fixation point was presented at the center of the monitor for 400 ms, (b) a blank screen appeared for 200 ms, and (c) a stimulus was presented at the position of the fixation point. The stimulus remained visible until a keyboard response was made. Subjects pressed the slash key for words and the *Z* key for nonwords. The fixation point appeared 1,200 ms after a correct response. After an incorrect response, a message stating that the response was incorrect was presented slightly below the fixation point for 1,500 ms, after which the screen was cleared. The subject pressed the space bar to begin the 1,200-ms delay period.

Stimuli were organized in 10 blocks of trials (Blocks 1–9 = 600 stimuli per block; Block 10 = 412 stimuli). Blocks were counterbalanced across subjects in a Latin square design, and trials within each block were randomly presented with the constraint that there would be an equal number of words and nonwords of comparable length.[2] Stimuli were rerandomized and assigned to lists anew for each group of 10 subjects. Breaks occurred after every 150 trials within a block and between blocks. Two filler trials consisting of short two-syllable stimuli were presented at the beginning of the experiment and after every break. Twenty practice trials preceded the experiment.

[2] The t tests between words and nonwords were performed with length as a dependent measure for each list, and all ps $> .20$.

Naming Task

The naming task was similar to the LDT with the exception that subjects read aloud the words, and their responses triggered the computer via a voice key. After the computer detected the response, the stimulus word was erased from the screen, and the subject coded the response by pressing a button on the mouse to move on to the next trial. If there was a pronunciation error or if an extraneous sound triggered the voice key, subjects pressed the right button on the mouse. If the subject believed their correct pronunciation triggered the voice key, then they pressed the left button on the mouse. Pressing either mouse button initiated a 1,200-ms intertrial interval.

Results and Discussion

The present analyses included only those words ($N = 2{,}726$) for which naming and lexical decision latencies as well as subjective frequency values (Balota, Pilotti, & Cortese, 2001) were available. To directly compare lexical decision and naming performance across both age groups, we decided to ensure that there was clear evidence that both groups were likely to know the stimulus words. Thus, we took the conservative approach of only including words that achieved at least a 67% level of accuracy (i.e., 20 out of 30 subjects responded correctly) in the LDT for both the young and the older adults. These criteria preserved 2,428 words.

Any response that was coded as an error in the naming task (0.7% for the young adults and 0.4% for the older adults) or any trial that produced an incorrect response in the LDT (6.1% for the young adults and 2.4% for the older adults) was excluded from the response latency analyses. In addition, any response faster than 200 ms or slower than 3,000 ms (1,500 ms for the naming task) was identified as an extreme score. After excluding these extreme scores, a mean and a standard deviation were calculated for each subject. Response latencies above or below 2.5 standard deviations from each subject's mean latency were removed. The percentage of latencies removed for naming was 3.3% for the young adults and 4.3% for the older adults, whereas the percentage of latencies removed for lexical decision was 2.1% for the young adults and 2.4% for the older adults.

Before addressing the predictive power of the different variables, we first report some overall global analyses, which provide information about the consistency in response latencies across tasks and across age groups at the individual item level.

Item-Specific Consistencies Across Tasks

The first question addressed is the extent to which there is consistency across the naming task and the LDT. Figures 1 and 2 provide the scatter plots for the same set of items across naming and lexical decision for the young and older adults, respectively. As shown, there is relatively little consistency across tasks, suggesting that either (a) there is simply too much variability at this level of analysis and/or (b) there are considerable task-specific operations that are modulating performance at the item level. As we discuss below, it is clear that the latter is more critical. Naming and lexical decision performance are more related in the older adults ($R^2 = .170$) than in young adults ($R^2 = .079$). This is interesting because older adults are more variable than young adults are and, as noted below, the predictive power of the targeted variables is smaller in the older adults than in the young adults. It is possible that this difference in cross-task correlations may reflect that the young adults, as compared with the older adults, are more likely to engage in task-specific operations, thereby decreasing the cross-task correlations.

Age and General Slowing

Because of the large number of observations for each subject, one question that can be powerfully addressed is whether there are task-specific changes that are sensitive to age. According to a simple general slowing perspective, one should be able to predict the individual item mean reaction times (RTs) for the older adults by multiplying the mean RTs obtained from the young adults by

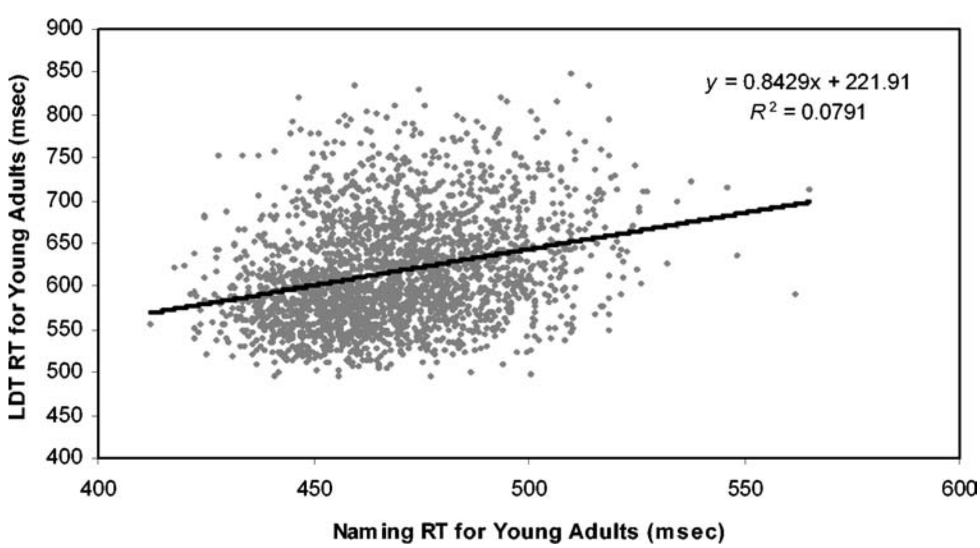


Figure 1. Mean item naming latencies as a function of mean item lexical decision latencies for the young adult subjects. LDT = lexical decision task; RT = reaction time.

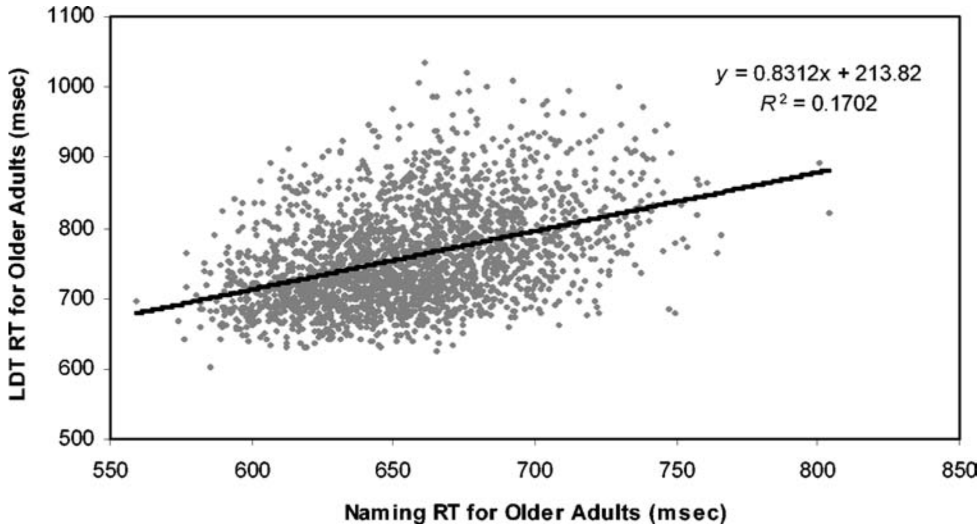


Figure 2. Mean item naming latencies as a function of mean item lexical decision latencies for the older adult subjects. LDT = lexical decision task; RT = reaction time.

some constant, and adding some constant. A priori, one might expect a different slowing function for older adults in lexical decision performance, which involves a more attention-demanding decision process than does naming performance, which one might argue is more stimulus driven. In fact, Cerella and Fozard (1984) even failed to find a reliable effect of age on speeded naming performance, but others have reported age differences in this task (e.g., Balota & Duchek, 1988). Alternatively, one might predict a consistent general slowing function across the tasks, once one corrects for differences in the variance associated with the two tasks (see Faust, Balota, Spieler, & Ferraro, 1999).

One way of looking at general slowing functions is to plot the young adults' means for a set of conditions as a function of the older adults' means. This is called a Brinley plot (Brinley, 1965). Figures 3 and 4 provide the Brinley plots for the naming and lexical decision item-level performance, respectively. Note first that there appears to be remarkable consistency in the size of the between-group reliability estimates in naming (R^2 = .428) and in lexical decision (R^2 = .430), even though this involved two different groups of young and older adults. Of course, the considerable increase in the amount of variance (a three- to fourfold increase) accounted for within tasks, compared with between tasks (see previous section), suggests that powerful task-specific operations modulate naming and lexical decision performance. Moreover, as shown in Figures 3 and 4, there appears to be relatively little change in the slope of the Brinley functions across tasks, with both slopes being relatively close to the identity function of 1, that is, for lexical decision performance, older adult RT = (0.73 ×

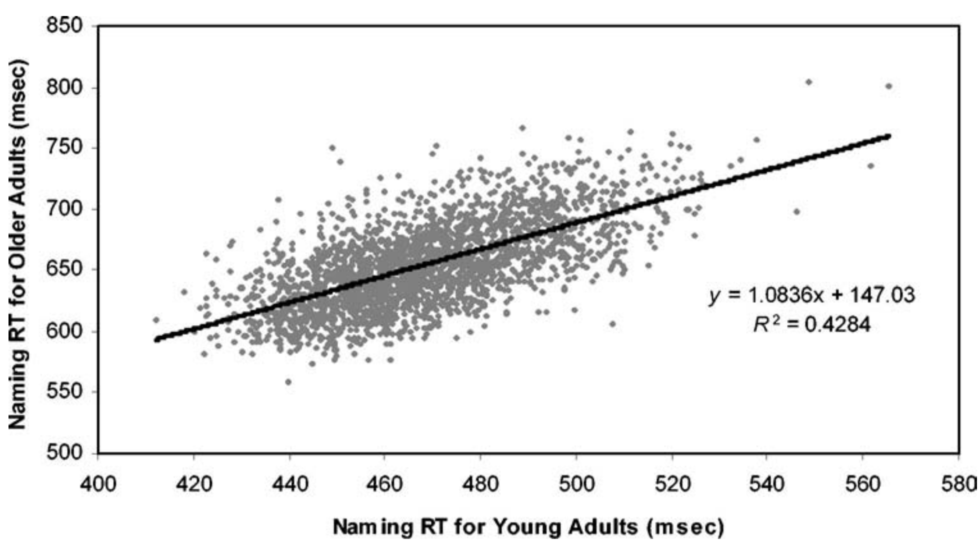


Figure 3. Mean item naming latencies for the young adults as a function of mean item naming latencies for the older adults (Brinley plot). RT = reaction time.

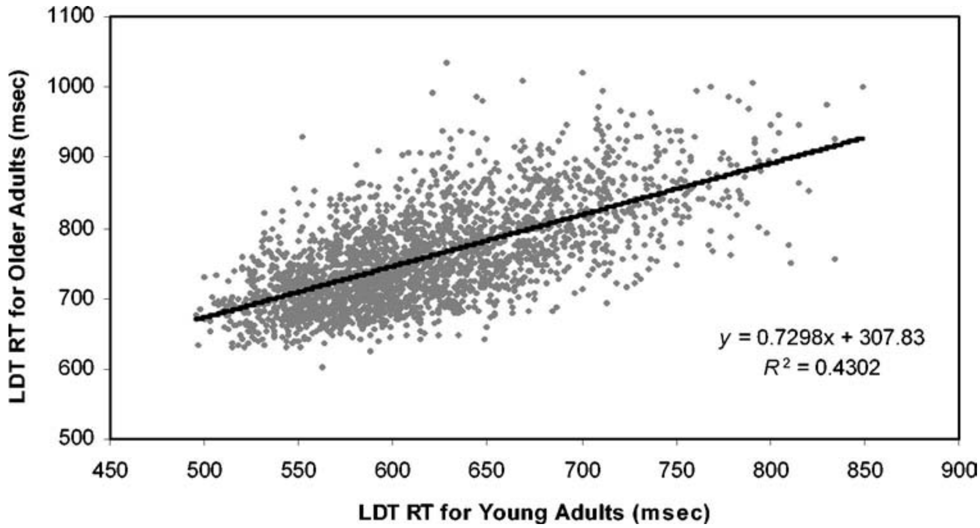


Figure 4. Mean item lexical decision latencies for the young adults as a function of mean item lexical decision latencies for the older adults (Brinley plot). LDT = lexical decision task; RT = reaction time.

young adult RT) + 308, and for naming performance, older adult RT = (1.08 × young adult RT) + 147. This suggests that the underlying lexical processing system is relatively stable across the two age groups. However, the intercepts change across these tasks, with the Brinley function producing a larger intercept for the LDT (308) than for the naming task (147). This may reflect the relatively larger differences in input, output, and decision processes in healthy older adults compared with young adults. For example, Bashore (1994) has argued from evoked-response data that a large portion of age-related slowing is due to output processes. To test the reliability of these observations, we regressed each older adult against the mean of the young adults for lexical decision and naming performance and then submitted the standardized regression coefficients and intercepts to *t* tests to determine if there were task-specific changes in these Brinley functions. The results of these *t* tests yielded a reliable age-related difference in intercepts, $t(57) = 3.87$, $p < .001$, but not in slopes, $t(57) < 1.00$. Thus, at this global level, there is evidence of a main effect of age on overall response latency but relatively little evidence of an age-related change in the processes associated across items in naming and lexical decision performance. The larger intercept in lexical decision compared with naming in these slowing functions may be viewed as consistent with age sensitivity to the decision processes tied to the LDT.

RT Distribution Analyses

We also considered the data at the individual-subject level to determine the nature of the RT distributions via the ex-Gaussian function. The ex-Gaussian function is the convolution of an exponential function and a Gaussian function (see Luce, 1986, for details). Although there are clearly other procedures for describing RT distributions (Van Zandt, 2000), the ex-Gaussian is useful as a first-level description and has the nice property that the mean response latency of an empirical distribution is approximated by the sum of the mean of the Gaussian component and the mean of the exponential component. Balota and Spieler (1999) have provided evidence that the influences of specific variables (e.g., frequency and repetition) have differential effects in naming and lexical decision performance on the parameters of the ex-Gaussian (also see Andrews & Heathcote, 2001). Each subject's empirical RT distribution was fit to the ex-Gaussian function to obtain maximum likelihood estimates of mu, which reflects the mean of the Gaussian component of the distribution; sigma, which reflects the standard deviation associated with the Gaussian component; and tau, which reflects the mean and standard deviation associated with the exponential component of the distribution. Table 1 presents the means of each of the three parameters across subjects as a function of task and age group.

To compare the components across subjects, we submitted each of the parameters to a 2 (age group) × 2 (task) ANOVA. Estimates of mu were larger for the older adults than for the young adults, $F(1, 116) = 183.78$, $MSE = 3{,}178.17$, $p < .001$, $\eta^2 = .61$. The effect of task on mu was much smaller and only marginally reliable, $F(1, 116) = 5.42$, $MSE = 3{,}178.17$, $p < .05$, $\eta^2 = .05$. The Group × Task interaction did not approach significance, $p > .15$. Turning to sigma, there was again an effect of group, $F(1, 116) = 23.22$, $MSE = 275.58$, $p < .001$, $\eta^2 = .17$, but no effect of task. However, sigma produced a reliable Group × Task interaction, $F(1, 116) = 7.63$, $MSE = 275.58$, $p < .01$, $\eta^2 = .06$, which reflected the fact that older adults produced more variance in the

Table 1
Mean Ex-Gaussian Estimates for Young and Older Adults for Both Naming and Lexical Decision Performance

Task	Mu	Sigma	Tau
Lexical decision			
Young adults	464	43	147
Older adults	590	50	168
Naming			
Young adults	426	40	42
Older adults	579	63	76

Gaussian component in the naming task compared with the LDT, whereas the young adults produced similar levels in the two tasks. Turning to the exponential component, tau, there were main effects of group, $F(1, 116) = 9.33$, $MSE = 2{,}425.82$, $p < .005$, $\eta^2 = .07$, reflecting larger estimates of tau in older adults than in young adults, and task, $F(1, 116) = 119.26$, $MSE = 2{,}425.82$, $p < .001$, $\eta^2 = .51$, reflecting larger estimates of tau in lexical decision compared with naming performance. However, there was no evidence of an interaction between the two factors, $F < 1.00$. Overall, these results suggest that group influences all three components of the RT distribution, with the largest influence being on the mean of the Gaussian component. In contrast, task has a dramatic influence on the exponential component, reflecting the fact that the lexical decision data are much more skewed than the naming data are. It has been argued that this increased skewing in lexical decision may, in part, reflect the binary decision component in this task, compared with speeded naming performance (see Balota & Spieler, 1999).

Figures 5 and 6 display the ex-Gaussian functions for the Vincentized (grouped in percentiles across subjects) data based on the mean maximum likelihood estimates obtained from the individual subject analyses for both naming and lexical decision performance, respectively. Consistent with the results from the ANOVAs, there is considerably more skewing of the lexical decision distributions in Figure 6 than of the naming distributions in Figure 5. In addition, as shown within each figure, the shapes of the RT distributions are relatively similar (although some differences are described above) for older adults and young adults, with the major difference being a shift in the distributions for the older adults compared with the younger adults.

Comparison of Word-Frequency Estimates

We now turn to a comparison of measures of word frequency to determine which measure will ultimately be used in the subsequent regression analyses. If there are differences among the word-frequency measures and one uses a weak measure, a considerable amount of frequency-based information could be lost in an analysis. Hence, we compared the predictive power of the following five objective word-frequency measures. The Kučera and Francis (1967) frequency norms are derived from a corpus of 1,014,000 words drawn from a wide variety of American English texts. The Center for Lexical Information (CELEX) word-form frequency norms are derived from a 17.9-million-word corpus built from a mixture of written texts (Baayen, Piepenbrock, & van Rijn, 1993). The Zeno frequency norms (Zeno, Ivens, Millard, & Duvvuri, 1995) are based on more than 17 million words culled from approximately 6,300 textbooks, works of literature, and popular works of fiction and nonfiction. The Hyperspace Analogue to Language (HAL) frequency norms (Lund & Burgess, 1996) are based on the HAL corpus, which consists of approximately 131 million words gathered across 3,000 Usenet newsgroups in February 1995. The MetaMetrics frequency norms are a recently developed corpus of 350 million words that span 21,000 computer text files containing fiction, nonfiction, and kindergarten–12th-grade textbooks (MetaMetrics, Inc., 2003). For each of the above norms, we took the log of the sum of the frequency of the item plus 1. For comparison purposes, we also used the subjective frequency norms (Balota et al., 2001), which are based on college students' subjective ratings of how frequently they have encountered a word in their lifetime.

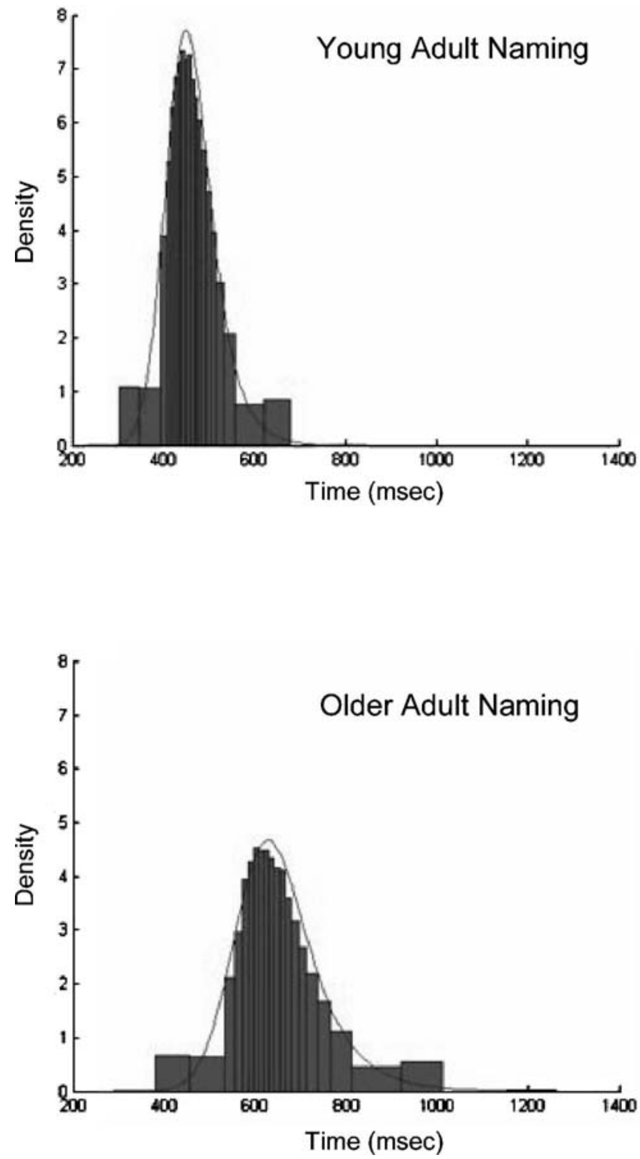
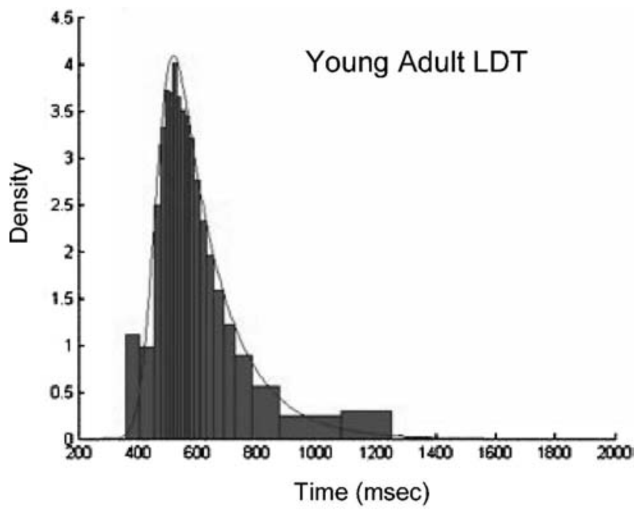


Figure 5. Ex-Gaussian functions for Vincentized young (top half) and older (bottom half) adult naming latencies. Note that *y*-axis values should be multiplied by .001.

Figure 7 displays the R^2 estimates from the different frequency counts as a function of age and task. There are three points to note from Figure 7. First, as expected, the predictive power of word frequency is consistently larger in lexical decision than in naming. Second, older adults tend to produce larger word-frequency effects in naming than do young adults, whereas the opposite pattern is found in lexical decision; that is, young adults produce larger word-frequency effects than do older adults. Third, and more important, there is considerable variability in the amount of variance accounted for by the word-frequency estimates. Specifically, in both naming and lexical decision performance, the Kučera and Francis (1967) norms account for the least amount of variance, followed by CELEX (Baayen et al., 1993) norms. These two



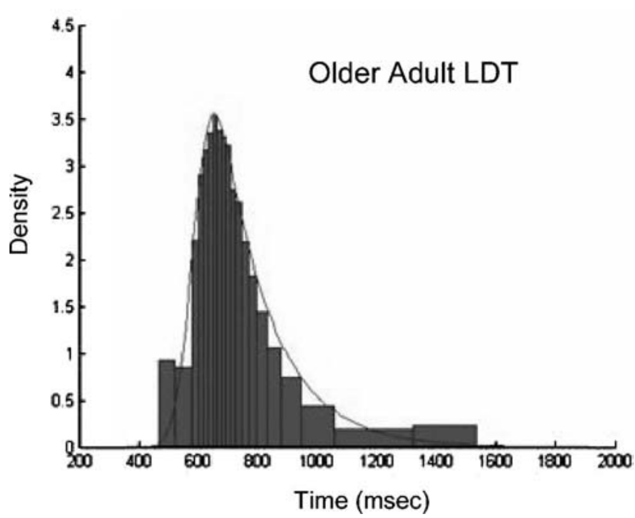


Figure 6. Ex-Gaussian functions for Vincentized young (top half) and older (bottom half) adult lexical decision performance. LDT = lexical decision task. Note that *y*-axis values should be multiplied by .001.

norms are probably most commonly used to control for and investigate the influence of word frequency. These differences are exaggerated in the LDT, where one finds, for example, that a 10% difference in variance is accounted for between the Zeno et al. (1995) norms and the Kučera and Francis norms (see Burgess & Livesay, 1998, for further discussion of word-frequency norms). Because of the consistently large influence of the Zeno et al. norms across groups and tasks, we decided to use these norms as our objective word-frequency norms in the regression analyses described below. Zevin and Seidenberg (2002) reached the same conclusion regarding the relative quality of the Zeno et al. norms compared with the CELEX and the Kučera and Francis norms.

Figures 8A through 8D display the scatter plots of the item means and the linear and quadratic functions for the log of the Zeno et al. (1995) norms and the Balota et al. (2001) subjective frequency norms. As shown, there is considerable scatter in both the naming task and the LDT. More interesting is that there is a distinct quadratic component in lexical decision that does not occur in naming performance. Specifically, for naming performance, the R^2 for the linear and quadratic components are virtually identical, whereas for lexical decision performance, the quadratic component adds as much as 7% of the variance. Clearly, even taking the log of word-frequency estimates does not capture the nonlinear relationship between word frequency and response latency in lexical decision performance (also see Murray & Forster, in press, for further evidence that log frequency has a nonlinear relationship to lexical decision response latencies). To apply the same predictor variables across the naming task and the LDT, we have used the log of the Zeno et al. norms in the present analyses. We return to the quadratic component in lexical decision performance in the General Discussion section.

Predictor Variables for the Regression Analyses

We now turn to regression analyses to assess the predictive power of a set of targeted predictor variables that have been identified from the literature. We group these variables into the following three sets: surface level, lexical level, and semantic level.

Surface Level

The first step in each analysis involved coding the initial phoneme of the words. We call this the surface-level coding because this will, in part, capture sensitivity to voice key biases (see Rastle & Davis, 2003). However, it is also the case that this surface-level coding may be sensitive to the ease of the implementation of the different phonological codes during articulation. Each word in the data set was coded dichotomously (1 or 0) according to the following 13 categories (see Spieler & Balota, 1997; Treiman et al., 1995), where 1 denotes the presence of the feature and 0 denotes the absence of a feature: affricative, alveolar, bilabial, dental, fricative, glottal, labiodental, liquid, nasal, palatal, stop, velar, and voiced. As shown below, this first step in coding onsets is quite powerful in predicting naming response latencies (see Kessler, Treiman, & Mullennix, 2003, for alternative procedures for coding onsets).

Lexical Level

At this level in the analyses, we entered variables that involve characteristics above the individual phoneme or letter but that are not traditionally considered semantic-level variables.

Word length. Word length refers to the number of letters in each word.

Neighborhood size. Neighborhood size refers to the number of orthographic neighbors 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). These neighborhood size values were based on 40,481 words available at the ELP Web site (http://elexicon.wustl.edu/).

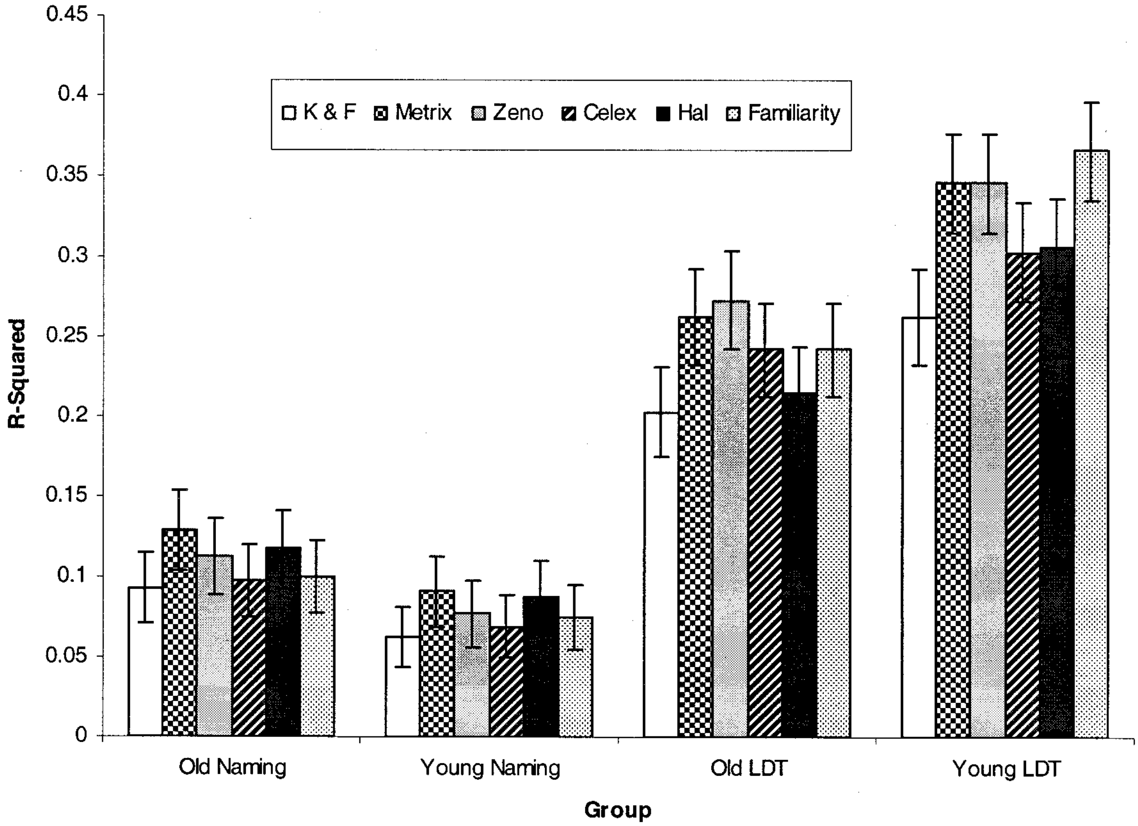


Figure 7. R^2 estimates from the six different frequency counts as a function of age and task. Error bars represent 95% confidence intervals. K & F = Kučera and Francis (1967) frequency norms; Metrix = the MetaMetrics frequency norms (MetaMetrics, Inc., 2003); Zeno = the Zeno et al. (1995) frequency norms; Celex = the Center for Lexical Information word-form frequency norms (Baayen, Piepenbrock, & van Rijn, 1993); HAL = the Hyperspace Analogue to Language (HAL) frequency norms (Lund & Burgess, 1996); Familiarity = subjective frequency norms (Balota et al., 2001); Old Naming = performance of older adults on the naming task; Young Naming = the performance of young adults on the naming task; Old LDT = the performance of older adults on the lexical decision task; Young LDT = the performance of young adults on the lexical decision task.

Objective frequency. As noted above, we have selected the log of (frequency + 1) taken from the Zeno et al. (1995) norms as our objective frequency index.

Subjective frequency. Subjective frequency, as described above, was taken from Balota et al. (2001).

Consistency measures. As shown in Figure 9, we used linguistic principles to decompose syllables into their onsets and rime components. In our example, we refer to the figure to illustrate how a word can vary along four continuous consistency dimensions: feedforward onset consistency, feedforward rime consistency, feedback onset consistency, and feedback rime consistency. We first describe the different measures of consistency conceptually and then explain how we operationalized them. These consistency measures were based on a pool of 4,444 monosyllablic words available from the ELP (Balota et al., 2002), which included a large set of single-syllable words that were known to at least two out of three undergraduate raters; see http://elexicon.wustl.edu/ for details. These estimates are more comprehensive than those based only on the single-syllable words in the Kučera and Francis (1967) norms and, hence, the consistency measures that are available from Ziegler, Stone, and Jacobs (1997).

Feedforward onset consistency of a word is computed with reference to its spelling onset neighbors, that is, words that share the same orthographic onset. For example, because the orthographic onset of *cad* is *c*–, its spelling onset neighbors include, among others, *car*, *can*, *card*, and *cite*. *Cad* is high on feedforward onset consistency because most of its spelling onset neighbors are friends, that is, they share the same pronunciation (/k/) for the orthographic onset, and only a few are enemies, that is, they have a different pronunciation (/s/) for the onset. Conversely, *cite* is low on feedforward onset consistency because most of its onset spelling neighbors are enemies. The vast majority of *c*– onset words have the onset pronounced as /k/ rather than /s/.

Feedforward rime consistency reflects the spelling rime neighbors of a word, that is, words that share the same orthographic rime. The rime neighbors of *cad* include *sad*, *mad*, *lad*, and *squad*. *Cad* is high on feedforward rime consistency because most of its rime neighbors are friends and have *–ad* pronounced as /æd/.

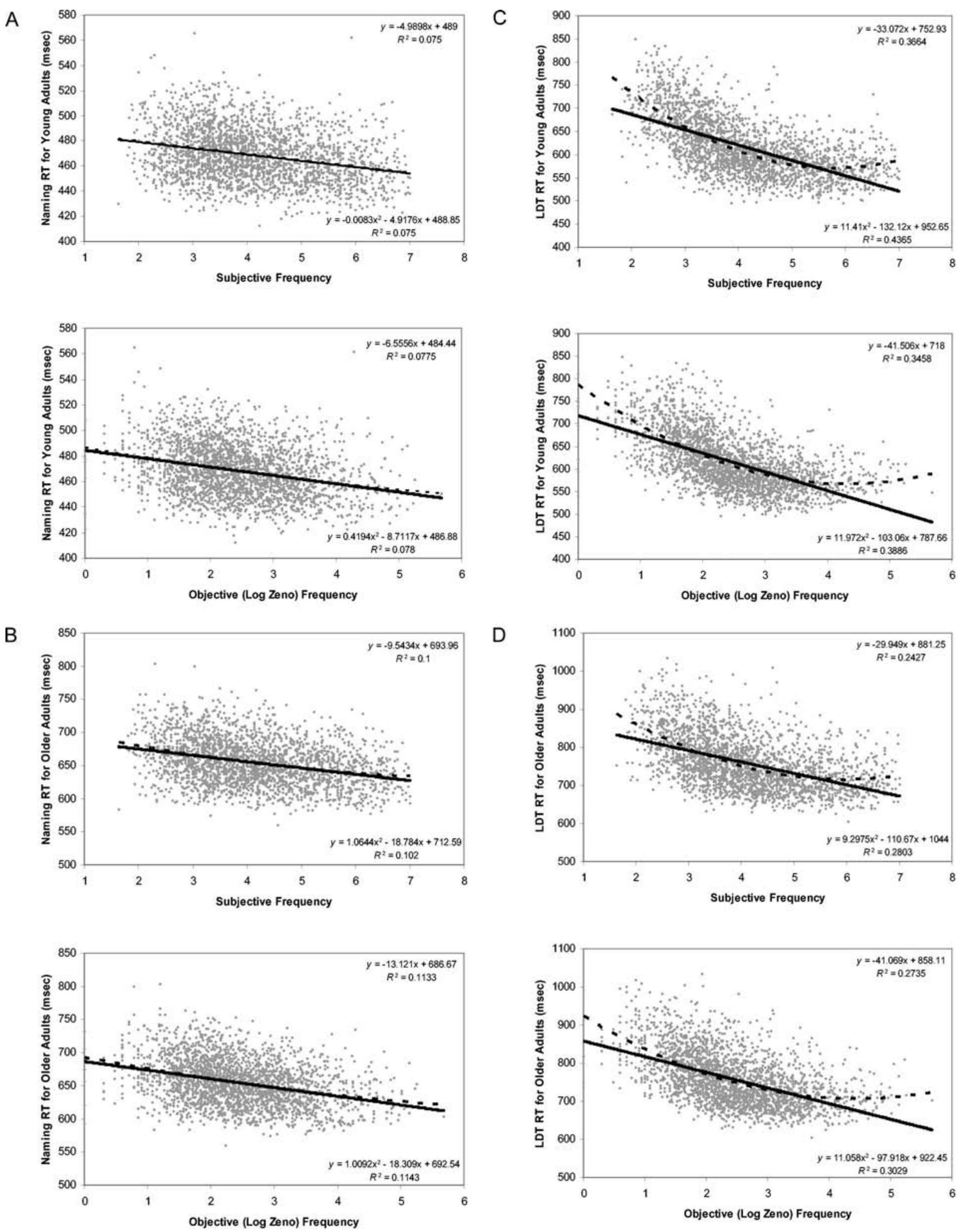


Figure 8. A: Scatter plots depicting the linear and quadratic trends for objective and subjective frequency measures in young adults' naming. B: Scatter plots depicting the linear and quadratic trends for objective and subjective frequency measures in older adults' naming. C: Scatter plots depicting the linear and quadratic trends for objective and subjective frequency measures in young adults' lexical decision. D: Scatter plots depicting the linear and quadratic trends for objective and subjective frequency measures in older adults' lexical decision. Scatter plots on the left represent performance on the naming task; scatter plots on the right represent performance on the lexical decision task. RT = reaction time; LDT = lexical decision task.

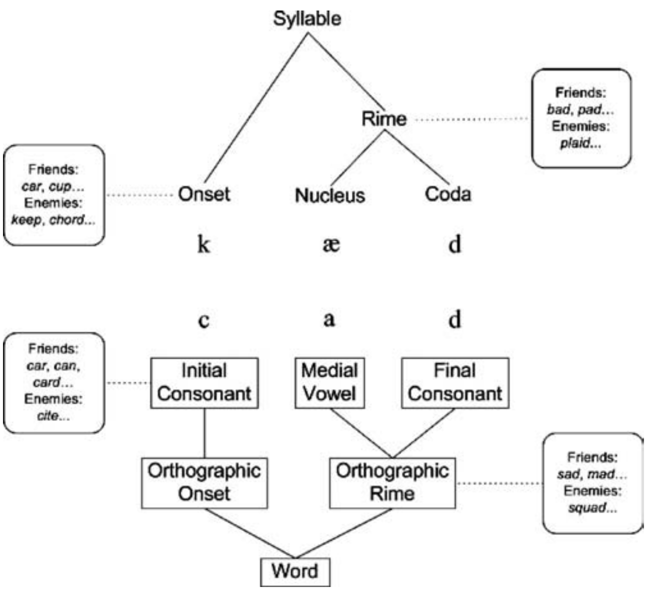


Figure 9. Onset and rime organization for syllabic structure.

Squad is low on feedforward rime consistency because most of its neighbors are enemies; that is, the rime *–ad* is typically not pronounced the way it is in *squad*.[3]

The feedback onset consistency of *cad* is computed with reference to its phonological onset neighbors, that is, words that begin with the phonological onset /k/. Examples include *cup*, *chord*, and *keep*. *Cad* is high on feedback onset consistency because the majority of words beginning with /k/ are also spelled with *c–*. Conversely, *chord* is low on feedback onset consistency because very few /k/ onset words are spelled with *ch–*.

The feedback rime consistency of *cad* is determined by the number of words that have rimes pronounced as /æd/, which include *bad*, *pad*, and *plaid*. *Cad* is very feedback rime consistent because most /æd/ neighbors are also spelled with *–ad*. In contrast, *plaid* is very feedback rime inconsistent. It is the only word with the /æd/ rime spelled as *–aid*.

Each of these consistency measures was operationalized using the following definition (a variant of Luce's choice axiom, see Luce, 1977), where *f* is the number of friends (including the target word itself) and *e* is the number of enemies. This is a token definition, because the consistency of a word is weighted by both the number and the log word frequencies of its neighbors, with values ranging between 0 (least consistent) to 1.00 (most consistent). An alternative way of computing consistency would be to calculate consistency at the type level, which does not consider the word frequencies of the friends and enemies. We used the frequency of the friends and enemies because Jared et al. (1990) demonstrated the importance of the frequency of friends and enemies in calculating consistency.

$$Consistency = \frac{\sum_{i=1}^{f} \lg \text{freq}(\textit{friends})}{\sum_{i=1}^{f} \lg \text{freq}(\textit{friends}) + \sum_{i=1}^{e} \lg \text{freq}(\textit{enemies})}. \quad (1)$$

For example, to calculate the feedforward rime consistency of *branch*, we need to determine the log frequencies of its friends and enemies. *Branch* has two friends (*blanch* and *ranch*) and one enemy (*stanch*).

$$Consistency(\text{branch}) = \frac{\lg \text{freq(branch)} + \lg \text{freq(blanch)} + \lg \text{freq(ranch)}}{\lg \text{freq(branch)} + \lg \text{freq(blanch)} + \lg \text{freq(ranch)} + \lg \text{freq(stanch)}}. \quad (2)$$

Semantic Level

As noted earlier, there has been some debate regarding the unique role of meaning in naming and lexical decision performance above and beyond other confounding variables (e.g., Gernsbacher, 1984). The important theoretical issue here is whether meaning provides a top-down influence during word recognition or whether word recognition must precede access to meaning. To explore this issue, we entered three different sets of semantic variables in the third step of the regression analyses after the onset (surface-level) and lexical-level predictor variables were partialed out. The first set involved standard measures of semantic information obtained for the set of items available from the Toglia and Battig (1978) norms and the Nelson, McEvoy, and Schreiber (1998) norms. This set of analyses was based on 997 words. The second set of analyses included a new set of imageability norms for all words (Cortese & Fugett, 2003). Finally, the third set of analyses involved more recent connectivity measures that are available on 1,625 words (Steyvers & Tenenbaum, 2004). We entered these three sets of semantic predictors separately, because the sets may tap different qualities of semantic representation and, more important, are based on different subsets of items.

Nelson's set size. Nelson's set size is the number of associates produced by 2 or more subjects in free association (Nelson et al., 1998). These norms were collected on 5,000 words across 6,000 subjects, with each subject providing free associations to a subset of 100 to 120 words.

Imageability. These values were obtained from the Toglia and Battig norms (1978) and are ratings of the ease with which an image can be generated when a given word is presented. In addition, as noted above, we also included a more recent measure of imageability obtained by Cortese and Fugett (2003).

Meaningfulness. These values were obtained from the Toglia and Battig norms (1978). They are ratings of how strongly other words come to mind and the number of associates that come to mind when subjects are presented with target words.

[3] For strange words (e.g., *aisle*, see Seidenberg, Waters, Barnes, & Tanenhaus, 1984), one might expect that these consistency measures are inappropriate, because the degree of consistency is actually complete, that is, these words do not have any rime neighbors. We identified 44 such strange words and conducted a second set of analyses without these items. The pattern of consistency effects did not change when these items were excluded; therefore, these items remained in all analyses.

Table 2
Means and Standard Deviations for the Predictor Variables and the Dependent Variables Used in the Regression Analyses for Words

Variable	*M*	*SD*
Predictor variables		
Entered in Step 1		
Voiced	0.46	0.50
Bilabial	0.24	0.43
Labiodental	0.08	0.26
Dental	0.02	0.14
Alveolar	0.34	0.47
Palatal	0.11	0.31
Velar	0.14	0.34
Glottal	0.05	0.22
Stop	0.39	0.49
Fricative	0.34	0.47
Affricative	0.03	0.18
Nasal	0.07	0.25
Liquid glide	0.14	0.35
Entered in Step 2		
Feedback onset token	0.92	0.19
Feedforward onset token	0.97	0.12
Feedback rime token	0.74	0.30
Feedforward rime token	0.90	0.21
Objective frequency	2.44	0.88
Neighborhood count	6.92	5.16
Length	4.36	0.86
Subjective frequency	4.12	1.14
Entered in Step 3		
Set size	14.10	5.10
T&B imageability	4.79	0.97
T&B meaningfulness	4.30	0.63
C&F imageability	4.33	1.37
WordNet	1.63	0.86
Connectivity	3.10	0.70
Dependent variables		
Reaction time (in ms)		
Lexical decision—older	757.96	69.00
Lexical decision—young	616.79	62.01
Naming—older	654.67	34.25
Naming—young	468.46	20.69
Accuracy (proportion correct)		
Lexical decision—older	0.95	0.06
Lexical decision—young	0.92	0.08
Naming—older	0.95	0.05
Naming—young	0.96	0.05

Note. T&B = Toglia & Battig (1978); C&F = Cortese & Fugett (2003).

The third set of semantic measures was based on recent work by Steyvers and Tenenbaum (2004) regarding small-world semantic networks. These measures tap into the degree of interconnectivity between words (and presumably concepts) in semantic memory. Steyvers and Tenenbaum analyzed the connectivity among words from three large-scale databases: word association norms from Nelson et al. (1998), Miller's (1990) WordNet, and *Roget's Thesaurus of English Words and Phrases* (Roget, 1911). The primary metric here is how many connections does a given word have to other words in the network and how many words are connected to that given word, that is, bidirectional connections. Steyvers and Tenenbaum found that the large-scale structure of these semantic networks follows a nonrandom structure that is found in other domains, such as the neural networks in the worm *Cænorhabditis elegans* and the power grid of the western United States (see Watts & Strogatz, 1998). Specifically, there is sparse connectivity (given the size of the network and number of connections, most words are connected to relatively few other words), short average path lengths between words (i.e., one can connect two words in an approximately 5,000-word network via a relatively small number of paths, i.e., 5), and strong local clustering (a few words are highly interconnected to other words). The strong local clustering is important because these highly interconnected hubs allow access to a considerable amount of information via relatively few connections. Because connectivity follows a power function, we have taken the log of both of these predictor variables.

Log of Nelson's connectivity measure. These values were taken from the Nelson et al. (1998) norms and reflect the number of unique words produced by 2 or more subjects in free association to a given word plus the number of times the given word was produced by 2 or more subjects in the free-association norms to other words. For example, if 2 or more subjects produced 5 different words associated with the word *dog*, and *dog* was produced in the response sets of other words 7 times, then the connectivity measure would be log of 12.

Log of WordNet connectivity. This database was developed by Miller (1990) and has some similarity to *Roget's Thesaurus of English Words and Phrases* (Roget, 1911), but Miller's database also takes into consideration aspects of psycholinguistic theory. WordNet is based on the distinction between word forms (WordNet has over 120,000 word forms) and word meanings (it has over 99,000 word meanings). Words are connected if they share the same meaning (i.e., if they are synonyms) or when the same form is connected to multiple meanings.

Descriptive Statistics

Tables 2 (words) and 3 (nonwords) provide the means and standard deviations for each of the predictor variables, along with each of the dependent measures, as a function of age group. Table 4 presents the intercorrelation matrix among the predictor variables, as well as among the dependent measures.

Table 3
Means and Standard Deviations for the Predictor Variables and the Dependent Variables Used in the Regression Analyses for Nonwords

Variable	*M*	*SD*
Predictor variables		
Neighborhood count	5.50	4.28
Length	4.38	0.81
Dependent variables		
Reaction time (in ms)		
Lexical decision—older	856.84	74.53
Lexical decision—young	679.96	55.01
Accuracy (proportion correct)		
Lexical decision—older	0.92	0.07
Lexical decision—young	0.92	0.07

Table 4
Correlation Matrix for the Dependent Measures and the Predictor Variables From Steps 2 and 3 of the Regression Analyses

	Dependent variables							
Variable	1	2	3	4	5	6	7	8
1. LDT-O-Acc	—	.45***	.23***	.15***	−.53***	−.43***	−.25***	−.16***
2. LDT-Y-Acc		—	.18***	.17***	−.56***	−.66***	−.24***	−.17***
3. Name-O-Acc			—	.35***	−.25***	−.19***	−.25***	−.22***
4. Name-Y-Acc				—	−.20***	−.18***	−.32***	−.31***
5. LDT-O-RT					—	.66***	.41***	.29***
6. LDT-Y-RT						—	.35***	.28***
7. Name-O-RT							—	.66***
8. Name-Y-RT								—
9. Length								
10. Obj. freq.								
11. Sub. freq.								
12. Neigh. size								
13. FF onset consist.								
14. FB onset consist.								
15. FF rime consist.								
16. FB rime consist.								
17. T&B meaningful.								
18. T&B image.								
19. C&F image.								
20. Nelson set size								
21. WordNet								
22. Connectivity								

Note. LDT-O-Acc = lexical decision task, older adults, accuracy; LDT-Y-Acc = lexical decision task, young adults, accuracy; Name-O-Acc = naming task, older adults, accuracy; Name-Y-Acc = naming task, young adults, accuracy; LDT-O-RT = lexical decision task, older adults, reaction time; LDT-Y-RT = lexical decision task, young adults, reaction time; Name-O-RT = naming task, older adults, reaction time; Name-Y-RT = naming task, young adults, reaction time; Obj. freq. = objective frequency; Sub. freq. = subjective frequency; Neigh. size = neighborhood size; FF onset consist. = feedforward onset consistency; FB onset consist. = feedback onset consistency; FF rime consist. = feedforward rime consistency; FB rime consist. = feedback rime consistency; T&B meaningful. = Toglia & Battig (1978) meaningfulness measure; T&B image. = Toglia & Battig imageability measure; C&F image. = Cortese & Fugett (2003) imageability measure.
* $p < .05$. ** $p < .01$. *** $p < .001$.

A few correlations in Table 4 are noteworthy. First, length and orthographic neighborhood size are negatively correlated (−.65), reflecting the smaller neighborhoods for longer words. Second, as expected, there is a strong positive (.78) relationship between the Zeno et al. (1995) objective frequency norms and the subjective frequency estimates obtained by Balota et al. (2001). Of course, it is also noteworthy that considerable variance (39%) is not shared across these estimates, and so the subjective frequency estimates appear to be tapping into useful unique frequency-based information. Finally, imageability and meaningfulness are related, and imageability is related to both subjective and objective frequency. This latter relationship is precisely why we first partialed out subjective and objective frequency before addressing the influence of semantic variables. Although a few other clusters of correlations reached significance because of the large number of observations, this is not surprising.[4]

Regression Analyses

Two classes of regression analyses were performed. First, regression analyses were performed on the mean item latencies and accuracies across subjects within each age group (young and older) and each task (LDT and naming). These analyses are the more standard procedure for investigating the predictive influence of variables on naming latencies (e.g., Treiman et al., 1995). Second, for the subject-level analyses (see Balota & Chumbley, 1984; Lorch & Myers, 1990), regression analyses were performed on each subject's response latencies and accuracies, allowing us to obtain standardized regression coefficients (betas) for each predictor variable for each subject. Betas standardize predictors using different measurement scales so that their effects can be directly compared. Separate 2 (age group) × 2 (task) ANOVAs were then performed for each predictor (with the standardized regression coefficient as the dependent variable) to determine if there are reliable changes in the influence of a predictor variable as a function of age group and/or task.

For both the item- and subject-level regression analyses, we used a three-step hierarchical approach. The first step included the 13 phonological onset variables, the second step included the lexical variables, and the third step involved the semantic variables.

[4] To explore the possible influence of suppressor variables, we entered only one of the correlated variables (e.g., either length or orthographic neighborhood size) into the hierarchical regression analyses to determine if such correlated variables influenced the remaining predictor variables. The pattern of reliable effects did not change, compared with when both variables were added into the regression equation. Hence, the combined influence of these correlated predictor variables did not modulate the influence of the remaining variables.

Predictor variables													
9	10	11	12	13	14	15	16	17	18	19	20	21	22
.00	.29***	.30***	.00	.03	.04	.02	−.01	.18***	.13***	.21***	.10***	.26***	.22***
.02	.43***	.45***	.01	.05*	.04	.02	.01	.32***	.16***	.30***	.12***	.32***	.27***
−.11***	.10***	.12***	.08***	.08***	.10***	.15***	.11***	.08**	.03	.09***	.03	.11***	.05*
−.14***	.11***	.10***	.13***	.17***	.14***	.16***	.12***	.09**	.07*	.09***	.03	.12***	.10***
.12***	−.52***	−.49***	−.08***	−.01	−.02	.01	.00	−.26***	−.09**	−.25***	−.06*	−.38***	−.34***
.09***	−.59***	−.61***	−.09***	−.03	−.02	.04	.00	−.37***	−.13***	−.28***	−.12***	−.41***	−.47***
.37***	−.34***	−.32***	−.31***	−.09***	−.07***	−.06**	−.08***	−.12***	−.01	−.09***	.01	−.24***	−.22***
.40***	−.28***	−.27***	−.36***	−.10***	−.02	−.07**	−.10***	−.07*	.01	−.07**	.02	−.18***	−.19***
—	−.16***	−.16***	−.65***	.00	.02	−.03	−.01	.03	.04	−.07**	.08**	−.03	−.07**
	—	.78***	.13***	−.06**	−.01	−.13***	−.07**	.11***	−.30***	.02	.10***	.46***	.59***
		—	.12***	−.07***	−.02	−.11***	−.05*	.18***	−.30***	−.02	.05*	.41***	.59***
			—	.12***	.09***	.02	.13***	−.01	.00	.06**	−.01	.12***	.10***
				—	.22***	.04	.07**	.08**	.13***	.07**	.06*	.06**	.02
					—	.08***	.05**	.02	.06	.04	.05*	.07**	.02
						—	.23***	.03	.10***	.04	.01	.01	−.06*
							—	.10***	.13***	.08***	.01	.06**	−.02
								—	.53***	.40***	−.01	.25***	.50***
									—	.89***	−.05	.00	.08*
										—	−.06*	.07**	.03
											—	.16***	.42***
												—	.45***
													—

Item-Level Regression Analyses

Response Latencies

The results of the mean item-level response latency regression analyses are displayed in Table 5. To control for the number of predictors, we report adjusted R^2 estimates for the item-level regressions. Adjusted R^2 estimates are unbiased estimates of the squared population correlation coefficient that take into account the sample size and the number of predictors in the model (Cohen, Cohen, West, & Aiken, 2003). First, consider the results from Step 1. As expected, the phonological onset variables predicted considerably more variance in naming than in lexical decision for both young adults (.35 vs. 01) and older adults (.22 vs. 04). The large amount of variance accounted for at the onset level clearly indicates that coding at the feature level is a powerful predictor of performance and some control of this simple level is necessary in naming studies. Interestingly, albeit to a much lesser extent, the phonological onsets (in particular voicing) do account for a reliable portion of the variance in the LDT, with slightly larger effects in older adults. Although the size of the individual regression coefficients for the onset variables is much smaller in lexical decision than in naming, the direction of the coefficients is identical across tasks and across young and older adults. This suggests that the articulatory and/or phonological processes that are involved in generating the onsets in naming performance also contribute, albeit to a much smaller extent, to lexical decision performance. Finally, it is noteworthy that the phonological onset variables entered in Step 1 accounted for considerably more variance in the young adults' naming performance (.35) than in the older adults' naming performance (.22). It is possible that the articulation may be more variable in older adults, thereby decreasing the predictive power of these phonological onset variables in the older adult group.

Turning to the lexical-level variables in Step 2, there are a number of observations. First, as expected, the LDT appears to be more dependent on the frequency-based information than the naming task is, and subjective and objective frequency account for comparable amounts of performance in lexical decision. Second, the correspondence between spelling and sound (as reflected by both the feedforward and the feedback consistency estimates for onset and rime) appears to predict both naming and lexical decision performance. Specifically, in naming, feedforward rime consistency, feedback rime consistency, and feedback onset consistency predict performance for both older and young adults. In lexical decision, the effects are somewhat smaller and localized to feedforward onset and rime consistency. Third, the influence of length in letters is much greater in naming than in lexical decision for both groups of subjects. Finally, orthographic neighborhood size is a predictor for young adult naming performance but not lexical decision performance. Moreover, there appears to be some tendency for an inhibitory effect of neighborhood size for older adults in the LDT.

Table 5
Standardized Reaction Time Regression Coefficients From Steps 1 and 2 of the Item-Level Regression Analyses for Young and Older Adult Lexical Decision Task (LDT) and Naming Performance

	Young		Older	
Predictor variable	LDT	Naming	LDT	Naming
Step 1				
Affricative	−0.18	−0.39***	−0.24†	−0.27*
Alveolar	0.44	1.18***	0.43	1.06***
Bilabial	0.39	1.03***	0.38	0.90**
Dental	0.13	0.37***	0.10	0.31**
Fricative	−0.26	−0.76**	−0.29	−0.41
Glottal	0.18	0.27*	0.10	0.18
Labiodental	0.25	0.63***	0.23	0.49**
Liquid	−0.32	−0.89***	−0.44†	−0.64**
Nasal	−0.20	−0.72***	−0.21	−0.44**
Palatal	0.35	0.57**	0.35	0.56**
Stop	−0.36	−1.12***	−0.52	−0.82**
Velar	0.39	1.03***	0.45†	0.92***
Voiced	0.10***	−0.12***	0.13***	−0.01
R^2	.01	.35	.04	.22
Step 2				
Length	−0.00	0.16***	0.07**	0.18***
Objective frequency	−0.32***	−0.13***	−0.37***	−0.20***
Subjective frequency	−0.38***	−0.13***	−0.21***	−0.13***
Neighborhood size	0.02	−0.10***	0.08***	−0.05*
Feedforward onset consistency	−0.07***	−0.03†	−0.04†	−0.03†
Feedback onset consistency	0.00	−0.08***	−0.02	−0.10***
Feedforward rime consistency	−0.05**	−0.08***	−0.05**	−0.08***
Feedback rime consistency	−0.02	−0.08***	−0.03†	−0.07***
R^2	.42	.49	.34	.39

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Turning to the semantic variables, one can see from the data in the top third of Table 6 that the standard semantic predictor variables (based on a subset of the full set of items) did in fact produce a reliable effect at the item level, which was larger in lexical decision than in naming. In addition, it appears that the older adults are somewhat less influenced by the standard semantic variables than are the young adults. The effect of imageability on naming and lexical decision extends to the full set of items available from the Cortese and Fugett (2003) imageability norms, as shown in the middle third of Table 6. Turning to the semantic connectivity estimates in the bottom third of Table 6, one can see that the WordNet (Miller, 1990) connectivity measure predicts naming and lexical decision performance, and the connectivity measure also predicts lexical decision. Specifically, the more connectivity a given word entails, the faster the response latency. The semantic connectivity measures are again larger in lexical decision than in naming.

Accuracy Analyses

The accuracy measures are based on the same items that went into the response latency measures. As noted above, to ensure that the observed effects across tasks and across groups were due to items that the subjects actually knew, we eliminated any items that did not have at least 20 observations in both young and older adults, based on lexical decision performance. Thus, these accuracy measures already excluded a set of items that did not reach this threshold.

The results of the item accuracy regression analyses are displayed in Table 7. As shown, the predictive power of the onset variables on accuracy in naming is dramatically reduced, compared with the predictive power on response latencies (shown in Table 5). This is expected because the coding of onsets should primarily influence response latencies due to voice key sensitivity and articulation instead of accuracy. There is again considerable consistency in the sign of the regression coefficients across naming and lexical decision, and older adults again appear to be less influenced by onsets in naming than are the young adults. The voicing variable again is reliable in lexical decision performance.

Turning to Steps 2 and 3, one finds a similar pattern as observed in the response latency data. In particular, for lexical decision, the frequency measures predict the majority of the variance, whereas in naming, the consistency measures account for the most variance. Finally, as shown in Table 8, the semantic measures included in Step 3 indicate that there is again evidence of semantic variables influencing performance more in lexical decision than in naming, and again there is a consistent influence of the Cortese and Fugett (2003) imageability estimates.

Subject-Level Regression Analyses

Response Latencies

The mean standardized regression coefficients based on the individual subjects' regression analyses are displayed in Table 9. (For simplicity, we do not include the phonological onset variables here, although these were partialed out for each subject.) As shown here, these regression coefficients are quite consistent with the item-level regression analyses. To directly compare the effects of task and/or group on each of the predictor variables, we present below the results of 2 (age group) × 2 (task) ANOVAs for each predictor (using the standardized regression coefficient as the dependent variable). For each variable, the main effect of task will be examined first, then the main effect of age, and finally the interaction between task and age.

Looking at the effects of task, we observed that there were larger effects of objective frequency, $F(1, 116) = 118.54$, $MSE = 0.002$, $\eta^2 = .505$; subjective frequency, $F(1, 116) = 97.09$, $MSE = 0.001$, $\eta^2 = .456$; and feedforward onset consistency, $F(1, 116) = 9.16$, $MSE = 0.0007$, $\eta^2 = .073$, in lexical decision than in naming. In contrast, there were larger effects of feedback onset consistency, $F(1, 116) = 30.74$, $MSE = 0.0006$, $\eta^2 = .209$; feedback rime consistency, $F(1, 116) = 18.34$, $MSE = 0.0004$, $\eta^2 = .137$; neighborhood size, $F(1, 116) = 57.73$, $MSE = 0.001$, $\eta^2 = .332$; and length, $F(1, 116) = 25.07$, $MSE = 0.002$, $\eta^2 = .178$, in naming than in lexical decision. Thus, nearly every variable produced a main effect of task, emphasizing the different constellation of processes engaged by the two tasks.

Turning to the effects of age, compared with young adults, older adults produced a smaller effect of orthographic neighborhood, $F(1, 116) = 8.31$, $MSE = 0.001$, $\eta^2 = .067$. Furthermore, older adults (compared with young adults) produced larger influences of

Table 6
Results From Step 3 Item-Level Reaction Time Regression Analyses With Standard Predictor Variables and Connectivity Measures

	Young		Older	
Order of entry into regression model	LDT	Naming	LDT	Naming
Standard semantic variables (n = 997)				
Step 1: Phonological onsets, R^2	.022**	.394***	.081***	.251***
Step 2: Lexical characteristics, R^2	.176***	.496***	.200***	.356***
Step 3: Semantic variables, R^2	.238***	.496***	.219***	.356***
Nelson set size, β	−0.08**	−0.01	−0.01	0.00
T&B imageability, β	−0.16***	−0.06*	−0.13***	−0.05
T&B meaningfulness, β	−0.13***	0.03	−0.04	−0.01
Cortese & Fugett (2003) imageability (n = 2,342)				
Step 1: Phonological onsets, R^2	.011***	.348***	.047***	.221***
Step 2: Lexical characteristics, R^2	.414***	.495***	.337***	.390***
Step 3: Semantic variables, R^2	.486***	.500***	.390***	.392***
Imageability, β	−0.27***	−0.04*	−0.23***	−0.05**
Semantic connectivity measures (n = 1,625)				
Step 1: Phonological onsets, R^2	.012**	.370***	.053***	.245***
Step 2: Lexical characteristics, R^2	.278***	.477***	.244***	.365***
Step 3: Semantic variables, R^2	.310***	.479***	.254***	.371***
WordNet, β	−0.07**	−0.04†	−0.09***	−0.09***
Connectivity, β	−0.21***	−0.04	−0.08*	−0.04

Note. LDT = lexical decision task; T&B = Toglia & Battig (1978).
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

objective frequency, $F(1, 116) = 10.20$, $MSE = 0.002$, $\eta^2 = .081$, but smaller influences of subjective frequency, $F(1, 116) = 23.98$, $MSE = 0.001$, $\eta^2 = .171$.

There was a reliable Age × Task interaction for the subjective frequency variable, $F(1, 116) = 19.15$, $MSE = 0.001$, $\eta^2 = .142$, such that young adults produced larger effects of subjective frequency than did older adults, but this was localized in lexical decision ($p < .001$) and did not occur in naming ($p > .05$).

Turning to the semantic variables, all three standard variables (Nelson's set size, imageability, and meaningfulness) produced larger coefficients in lexical decision than in naming (all ps $< .02$). However, it is interesting that there was no main effect of age for any of these three variables. There was an Age × Task interaction for the Toglia and Battig (1978) meaningfulness measure, $F(1, 116) = 8.03$, $MSE = 0.001$, $\eta^2 = .065$, which reflected the larger influence of task for young adults relative to older adults. The Cortese and Fugett (2003) imageability measure produced a main effect of task, $F(1, 116) = 276.34$, $MSE = 0.0008$, $\eta^2 = .704$, with larger influences of this variable in lexical decision than in naming. Finally, for the semantic connectivity measures, the Nelson et al. (1998) connectivity estimates resulted in main effects of task, $F(1, 116) = 39.39$, $MSE = 0.001$, $\eta^2 = .253$; age, $F(1, 116) = 9.54$, $MSE = 0.001$, $\eta^2 = .076$; and a Task × Age interaction, $F(1, 116) = 7.75$, $MSE = 0.001$, $\eta^2 = .063$. The interaction reflected the relatively large effect of connectivity for the young adult lexical decision performance. For the WordNet (Miller, 1990) connectivity estimates, there was a significant main effect of age (larger effects for older adults), $F(1, 116) = 4.14$, $MSE = 0.0008$, $\eta^2 = .034$.

Accuracy

Because Step 1 of the regression analyses involved a dichotomous dependent measure at the subject level, we used logistic regression for these analyses. The mean standardized regression coefficients for the subject-level accuracy analyses are displayed in Table 10. The coefficient used in Table 10 was the odds ratio. An odds ratio of 1.0 is associated with no relationship between the predictor and the dependent variable, whereas odds ratios greater than 1.0 correspond to positive regression coefficients and odds ratios less than 1.0 correspond to negative regression coefficients. The results indicated that lexical decision is more influenced than naming by objective frequency, $F(1, 116) = 40.21$, $MSE = 0.06$, $\eta^2 = .257$; subjective frequency, $F(1, 116) = 38.73$, $MSE = 0.05$, $\eta^2 = .250$; and orthographic neighborhood, $F(1, 116) = 4.73$, $MSE = 0.001$, $\eta^2 = .039$. In contrast, naming is more influenced than lexical decision by length, $F(1, 116) = 16.59$, $MSE = 0.05$, $\eta^2 = .125$; feedback onset consistency, $F(1, 116) = 9.48$, $MSE = 0.919$, $\eta^2 = .076$; feedforward rime consistency, $F(1, 116) = 17.25$, $MSE = 1.16$, $\eta^2 = .129$; and feedback rime consistency, $F(1, 116) = 19.68$, $MSE = 0.288$, $\eta^2 = .145$.

Turning to the effects of age, there was a Task × Age interaction, $F(1, 116) = 10.63$, $MSE = 0.05$, $\eta^2 = .084$, for the subjective frequency measure, which indicated that the age difference (young > old) was primarily localized in the LDT. Compared with

Table 7
Standardized Accuracy Regression Coefficients From Steps 1 and 2 of the Item-Level Regression Analyses for Young and Older Adult Lexical Decision Task (LDT) and Naming Performance

Predictor variable	Young LDT	Young Naming	Older LDT	Older Naming
Step 1				
Affricative	0.21	0.95***	0.21	0.12
Alveolar	−0.37	−2.57***	−0.38	−0.34
Bilabial	−0.32	−2.30***	−0.33	−0.27
Dental	−0.13	−0.77***	−0.08	−0.16
Fricative	0.40	2.38***	0.40	0.28
Glottal	−0.17	−1.10***	−0.13	−0.10
Labiodental	−0.21	−1.32***	−0.24	−0.12
Liquid	0.37	1.88***	0.38	0.26
Nasal	0.22	1.28***	0.22	0.14
Palatal	−0.31	−1.68***	−0.24	−0.17
Stop	0.43	2.58***	0.50	0.36
Velar	−0.35	−1.90***	−0.34	−0.30
Voiced	−0.09**	0.00	−0.08**	−0.01
R^2	.01	.04	.01	.01
Step 2				
Length	0.08**	−0.08**	0.00	−0.12***
Objective frequency	0.21***	0.08**	0.19***	0.04
Subjective frequency	0.32***	0.05	0.17***	0.11***
Neighborhood size	−0.02	0.01	−0.07**	−0.07*
Feedforward onset consistency	0.07**	0.16***	0.06*	0.04†
Feedback onset consistency	0.02	0.09***	0.03	0.10***
Feedforward rime consistency	0.08***	0.15***	0.06**	0.14***
Feedback rime consistency	0.02	0.09***	0.00	0.08***
R^2	.25	.13	.12	.08

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

young adults, older adults produced smaller effects of length and orthographic neighborhood size, $F(1, 116) = 6.75$, $MSE = 0.05$, $\eta^2 = .055$, and $F(1, 116) = 14.94$, $MSE = 0.001$, $\eta^2 = .114$, respectively. Generally, the pattern mirrors the analyses of response latencies, with the most salient finding being that the frequency measures better predict lexical decision, whereas the consistency measures better predict naming.

Turning to the analyses of the semantic variables, the only effects that emerged were a main effect of task for the Cortese and Fugett (2003) imageability estimates, $F(1, 116) = 58.44$, $MSE = 0.020$, $\eta^2 = .335$, and Nelson et al.'s (1998) estimate of connectivity, $F(1, 116) = 15.56$, $MSE = 0.113$, $\eta^2 = .118$. Both of these reflect larger effects in lexical decision than in naming.

Nonword Performance

Table 11 displays the regression coefficients for the item-level analyses (top half) and subject-level analyses (bottom half) for both accuracy and response latencies. (Nonwords were only included in the LDT.) The nonwords were only coded for length and neighborhood density. The results from both the item-level regression analyses and the subject-level regression analyses indicated that both predictor variables were highly reliable in these analyses. Thus, in contrast to lexical decision for words, both length and neighborhood density are strong predictors of nonword performance. Specifically, nonwords that are long and nonwords that have many orthographic neighbors produce relatively slow and less accurate lexical decision performance. The effects of these variables are slightly smaller in older adults than in young adults, but these differences did not reach significance in the subject-level analyses.

General Discussion

The present study provides evidence regarding the predictive power of standard lexical processing variables for virtually all single-syllable monomorphemic words in both naming and lexical decision performance and in both young adults and older adults. Although there are a number of intriguing aspects of these results concerning the standard predictor variables (discussed in detail below), we first discuss some concerns about the utility of large-scale studies of isolated word processing (for additional discussions of these issues, see Balota & Spieler, 1998; Seidenberg & Plaut, 1998). For example, one might argue that naming (or making lexical decisions to) nearly 3,000 words may produce variability due to fatigue or boredom. Hence, such data sets might be too noisy to usefully constrain word recognition models. However, we were able to account for, on average, 50% of the young adult variance and 40% of the older adult variance in these two lexical processing tasks via the regression equation on the full data set. Given the fact that there was considerable overlap in the response latencies for some of the words (as shown in Figures 1 and 2) and hence little variance to predict for these items, one might consider these to be relatively large amounts of variance accounted for by the predictor variables. In fact, the parameter estimates from the Seidenberg and McClelland (1989) and Plaut et al. (1996) models account for only 10.1% and 3.3% of the item-level variance for the present young adult naming data, respectively. We also found considerable consistency across subjects in the pattern of regression coefficients as reflected by the subject-level regression analyses. A third way of assessing the utility of such large databases is to select a subset of items from a published study to determine if one can replicate the obtained results at the mean level. Thus, we (Balota & Spieler, 1998) selected a subset of items from the present young adult naming data set that was used in the factorial study conducted by Taraban and McClelland (1987, Experiment 1A). Taraban and McClelland found a Frequency × Regularity interaction such that spelling-to-sound regularity produced a larger effect for low-frequency words compared with high-frequency words. This pattern was replicated in the same set of items taken from the present naming data set.

Although the above approaches provide support for the utility of large-scale databases, there is now a data set that affords a replication for both the lexical decision and the naming data. In particular, we were able to access lexical decision and naming data for the single-syllable words from a large data set of over 40,481 words from the ELP (Balota et al., 2002). The ELP involves a collaborative effort among six universities to provide behavioral and descriptive lexical processing information along with a search engine available on the Web. The lexical decision data are based on 30 to 35 observations per item and the naming data are based

Table 8
Results From Step 3 Item-Level Accuracy Regression Analyses With Standard Predictor Variables and Connectivity Measures

	Young		Old	
Order of entry into regression model	LDT	Naming	LDT	Naming
Standard semantic variables (n = 997)				
Step 1: Phonological onsets, R^2	.011*	.138***	.021**	.036***
Step 2: Lexical characteristics, R^2	.043***	.217***	.053***	.081***
Step 3: Semantic variables, R^2	.079***	.220***	.078***	.078***
Nelson set size, β	0.07*	0.05†	0.08**	0.02
T&B imageability, β	0.12**	0.04	0.17***	−0.01
T&B meaningfulness, β	0.10**	0.03	−0.03	0.02
Cortese & Fugett (2003) imageability (n = 2,342)				
Step 1: Phonological onsets, R^2	.010**	.043***	.013***	.013***
Step 2: Lexical characteristics, R^2	.238***	.135***	.113***	.079***
Step 3: Semantic variables, R^2	.331***	.138***	.152***	.083***
Imageability, β	0.31***	0.06**	0.20***	0.07**
Semantic connectivity measures (n = 1,625)				
Step 1: Phonological onsets, R^2	.006*	.064***	.012**	.010**
Step 2: Lexical characteristics, R^2	.109***	.127***	.060***	.057***
Step 3: Semantic variables, R^2	.118***	.134***	.073***	.057***
WordNet, β	0.05†	0.07**	0.04	0.04
Connectivity, β	0.11***	0.06†	0.15***	0.00

Note. LDT = lexical decision task; T&B = Toglia & Battig (1978).
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

on 25 to 30 observations per item (details of the methods are available at the ELP Web site, http://elexicon.wustl.edu/). Currently, 816 individuals have provided data for the LDT and 423 have provided data for the naming task. Each subject provides data for a subset of approximately 3,000 of the 40,481 tested words. To take into consideration overall differences across individuals in response latencies and the fact that each subject did not contribute to the data for all single-syllable words, we used the mean z score

Table 9
Mean Standardized Reaction Time Regression Coefficients From Steps 2 and 3 of the Subject-Level Regression Analyses for Young and Older Adult Lexical Decision Task (LDT) and Naming Performance

	Young		Older	
Predictor variable	LDT	Naming	LDT	Naming
Step 2				
Length	0.001	0.059***	0.030*	0.063***
Objective frequency	−0.125***	−0.048***	−0.149***	−0.070***
Subjective frequency	−0.149***	−0.049***	−0.084***	−0.045***
Neighborhood size	0.009	−0.040***	0.030***	−0.023**
Feedforward onset consistency	−0.026***	−0.007	−0.020***	−0.010†
Feedback onset consistency	0.001	−0.027***	−0.007†	−0.033***
Feedforward rime consistency	−0.017**	−0.025***	−0.023***	−0.026***
Feedback rime consistency	−0.010**	−0.027***	−0.010*	−0.025***
Step 3				
T&B meaningfulness[a]	−0.038***	0.007	−0.012	−0.006
T&B imageability[a]	−0.051***	−0.021**	−0.051***	−0.012
Nelson set size[a]	−0.022**	−0.001	−0.009	0.000
C&F imageability[b]	−0.109***	−0.012**	−0.095***	−0.016***
WordNet[c]	−0.027***	−0.013**	−0.031***	−0.030***
Connectivity[c]	−0.070***	−0.014*	−0.033***	−0.012†

Note. T&B = Toglia & Battig (1978); C&F = Cortese & Fugett (2003).
[a] n = 997. [b] n = 2,342. [c] n = 1,625.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 10
Mean Standardized Accuracy Regression Coefficients From Steps 2 and 3 of the Subject-Level Regression Analyses for Young and Older Adult Lexical Decision Task (LDT) and Naming Performance

	Young		Older	
Predictor variable	LDT	Naming	LDT	Naming
Step 2				
Length	2.87**	1.84*	1.44	2.31*
Objective frequency	7.14***	1.15	4.49**	0.91
Subjective frequency	14.59***	0.91	4.30**	1.42
Neighborhood size	2.04*	1.40	1.73†	0.98
Feedforward onset consistency	1.95*	3.85**	1.27	1.55
Feedback onset consistency	1.26	2.01*	1.05	2.29*
Feedforward rime consistency	2.91*	5.00**	1.31	3.34**
Feedback rime consistency	1.18	2.44*	0.78	2.20*
Step 3				
T&B meaningfulness[a]	1.43	0.51	1.16	0.94
T&B imageability[a]	1.21	0.73	2.57**	1.64
Nelson set size[a]	0.95	0.94	1.24	1.13
C&F imageability[b]	18.47***	1.61	7.74**	2.33*
WordNet[c]	1.27	1.13	1.40	0.61
Connectivity[c]	1.83	0.99	1.78*	1.25

Note. T&B = Toglia & Battig (1978); C&F = Cortese & Fugett (2003).
[a] $n = 997$. [b] $n = 2,342$. [c] $n = 1,625$.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

(based on each subject's overall response latency and standard deviation) as the dependent measure in the regression analyses for this set of items.

The results of this replication are quite clear. First consider the R^2 values from the three steps of the regression models. For the current lexical decision study, the R^2 estimates from Steps 1, 2, and 3 (including all words and the Cortese & Fugett, 2003, imageability norms in Step 3) were .01, .41, and .49, respectively, whereas the R^2 estimates from the words selected from the ELP were .01, .44, and .52. For the naming data, the R^2 estimates from Steps 1, 2, and 3 were .35, .50, and .50, respectively, and for the ELP, the estimates were .38, .57, and .58. More important, as shown in Figures 10 and 11, the lexical decision and naming results from the ELP provide a remarkable replication of not only the reliability of the regression coefficients but also on the size of the coefficients. Hence, even though these results were taken from a much more diverse subject pool from six different universities and the tested words were embedded within mostly multisyllabic words using different screening procedures, there was clear convergence in the regression coefficients. It is also noteworthy that the naming results from the ELP included data collected with a different voice key especially constructed for this project, and, on average, the subjects in the ELP produced response latencies that were on the order of 100 ms slower than those in the current naming study.

Given that these data provide relatively stable estimates of performance at the item level, we are now in a position to discuss the relative contributions of the targeted variables in speeded naming and lexical decision performance in the context of the word recognition literature. We discuss the predictive effects of each of these variables in turn.

Length

The effect of orthographic length has been central in recent models of speeded word naming. An important theoretical observation by Weekes (1997) indicated that length in letters influenced nonword-naming performance but did not influence word-naming performance after other variables were controlled for. This is in contrast to the earlier observation of an effect of length in speeded word naming by Frederiksen and Kroll (1976). However, Weekes pointed out that Frederiksen and Kroll did not control for potentially contaminating variables. Coltheart et al. (2001) recently highlighted this finding as being supportive of a dual-route model, in which the more serial, sublexical pathway is necessary for nonword naming, whereas a more parallel pathway contributes to word naming. Furthermore, Coltheart et al. argued that only the dual-route model can explicitly capture the Lexicality × Length interaction that Weekes observed.

The present study provides unequivocal evidence that longer words take more time to name than shorter words do. In fact, this variable accounted for nearly as much variance as objective frequency did. Hence, our data are inconsistent with the strong conclusion from the Weekes (1997) study that length does not produce a unique effect on word naming. Of course, the important question is why we found a pattern of results different than that reported by Weekes. There are at least three possibilities. First, it is possible that we obtained an effect of length because the range of lengths in the present study was larger (2 to 8 letters) than that in the Weekes study (3 to 6 letters). However, this does not appear to account for the difference in results, because the vast majority of stimuli in the present study were 3 to 6 letters in length, that is, 2,403 words out of the 2,428. There were only 25 words at the extremes, and an items-level regression excluding these items yielded highly reliable effects of length for both young ($p < .0001$) and older adults ($p < .0001$). Thus, restriction of range is not the answer.

A second possible reason for the differing patterns of results is that Weekes (1997) randomly intermixed words and nonwords within the same list. We intentionally did not intermix words and

Table 11
Mean Standardized Regression Coefficients for Nonwords for Both the Item-Level and Subject-Level Response Latency and Accuracy Analyses as a Function of Group

	Young		Older	
Predictor variable	RT	Accuracy	RT	Accuracy
Item level				
Neighborhood size	.45***	−.31***	.40***	−.25***
Length	.53***	−.27***	.41***	−.14***
R^2	.21	.07	.13	.04
Subject level				
Neighborhood size	.16***	.93***	.15***	.95***
Length	.19***	.73***	.16***	.87**

Note. RT = reaction time.
** $p < .01$. *** $p < .001$.

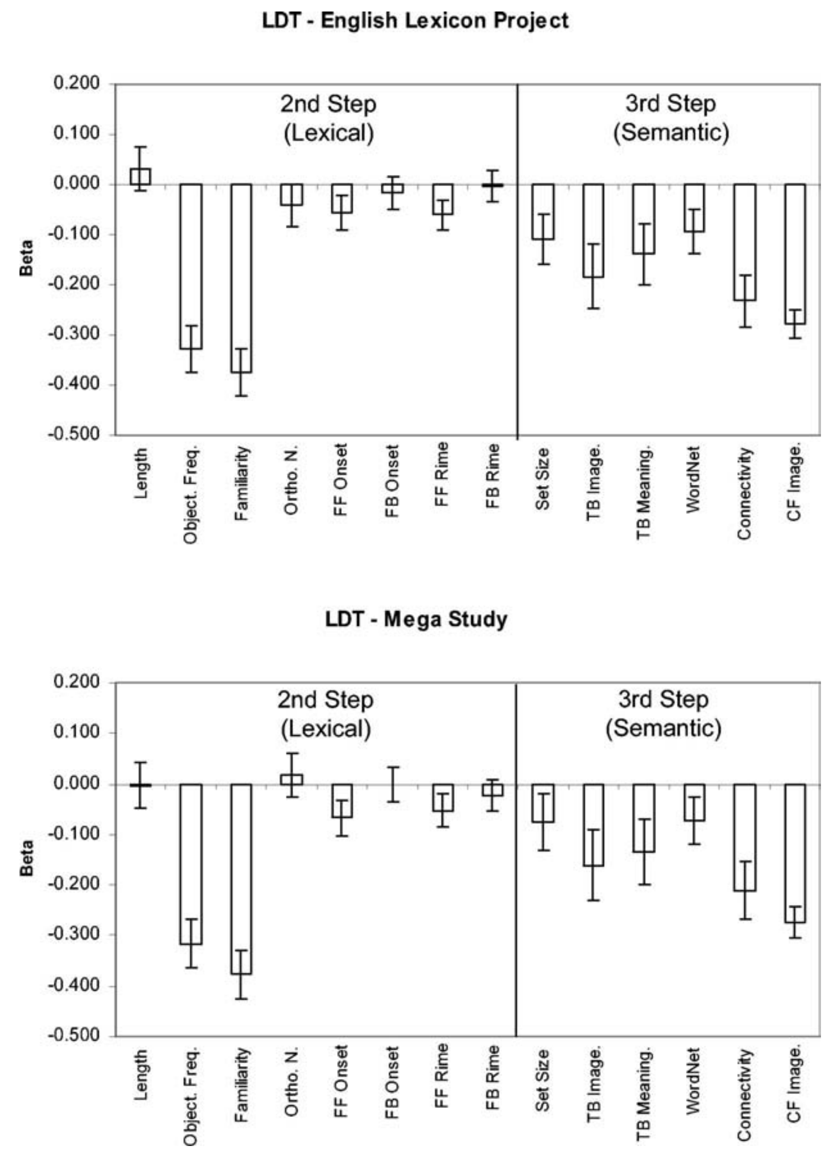


Figure 10. Replication of lexical decision task (LDT) results from items taken from the English Lexicon Project database (Balota et al., 2002). Error bars represent 95% confidence intervals. CF Image. = the Cortese and Fugett (2003) imageability measure; TB Meaning. = the Toglia and Battig (1978) meaningfulness measure; TB Image. = the Toglia and Battig imageability measure; FB Rime = feedback rime consistency; FF Rime = feedforward rime consistency; FB Onset = feedback onset consistency; FF Onset = feedforward onset consistency; Ortho. N. = orthographic neighborhood; Object. Freq. = objective frequency.

nonwords because this might encourage more nonlexical processing (e.g., Monsell et al., 1992; Zevin & Balota, 2000). If this intermixing were the case, then one would actually expect more of a length effect for words in the Weekes study compared with the present study. However, this was not what occurred.

Another possible reason for the difference is that Weekes (1997) covaried out length in phonemes. Because phoneme length and letter length are highly correlated, it is possible that there was no unique variance to be accounted for by letter length after phoneme length was partialed out. To address this possibility, we determined each word's length in phonemes and entered this variable, along with the remaining lexical predictor variables, into the regression equation in Step 2 of the model. The results again yielded highly reliable predictive power of length in letters in speeded naming performance, above and beyond length in phonemes for both young adults ($p < .001$) and older adults ($p < .001$). Moreover, the same pattern held when we entered length in letters after entering length in phonemes precisely as in the Weekes study. Hence, the present letter-length effect cannot be dismissed as a phoneme-length effect.

It is possible that length would have less of an effect for very familiar stimuli (i.e., for common words, the input may be more likely to be processed in parallel), as one might expect from the Coltheart et al. (2001) dual-route model described above. If this

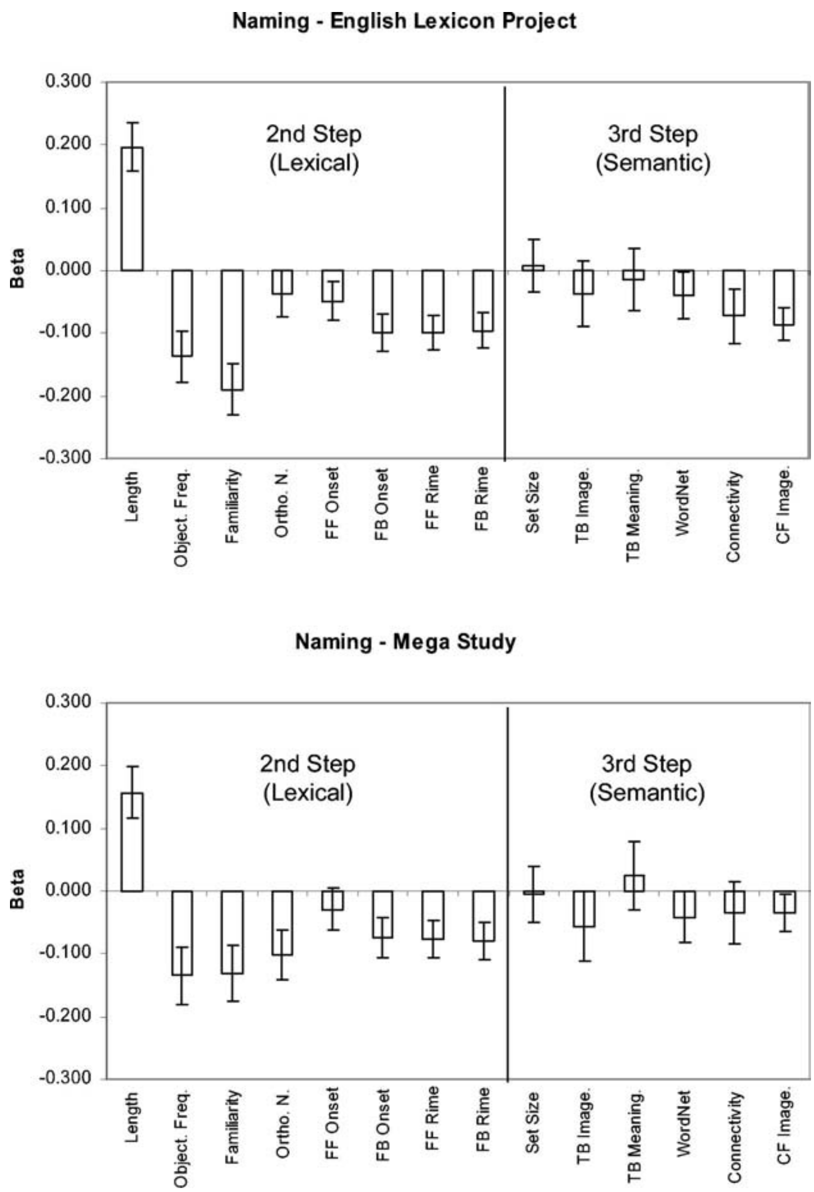


Figure 11. Replication of the naming results from items taken from the English Lexicon Project database (Balota et al., 2002). Error bars represent 95% confidence intervals. CF Image. = the Cortese and Fugett (2003) imageability measure; TB Meaning. = the Toglia and Battig (1978) meaningfulness measure; TB Image. = the Toglia and Battig imageability measure; FB Rime = feedback rime consistency; FF Rime = feedforward rime consistency; FB Onset = feedback onset consistency; FF Onset = feedforward onset consistency; Ortho. N. = orthographic neighborhood; Object. Freq. = objective frequency.

were the case, then one might expect a larger length effect for low-frequency words, where there may be a decreased likelihood of parallel lexical processing. Weekes (1997) did in fact find evidence of larger length effects for low-frequency words than for high-frequency words. However, Weekes did not find a unique effect after remaining variables were controlled for, even for low-frequency words.

To further explore the Length × Word Frequency interaction, we adopted the strategy advocated by Cohen et al. (2003). Although other techniques are available for testing interaction effects in regression, these methods reduce continuous variables to categories and diminish statistical power (Jaccard, Turrisi, & Wan, 1990). The Length × Word Frequency interaction was represented by a predictor variable created from the product of length and log Zeno frequency. The R^2 change between two regression models (one with the interaction term and one without) is measured, and the extent of the change is evaluated. This method uses the full regression model, that is, after all the additional variables have been partialed out in Steps 1 and 2, and hence does not discard any information. One way of conceptualizing this is that it is a test of the interaction while controlling for the potentially confounding influence of the remaining variables.

The results indicated that there was indeed a Length × Word Frequency interaction in naming performance for both young and older adults (both $ps < .01$). Because the full regression model is being used, Cohen et al. (2003) argued that the most appropriate procedure to interpret the interaction is to compare the slope of one variable at different levels of the other variable in the interaction. Hence, we present the slopes of word length at low, medium, and high levels of word frequency. As shown in the two leftmost sections of Figure 12), the effect of length is inhibitory for both high- and low-frequency words (as indicated by positive regression coefficients); however, the effect is larger for low-frequency words than for high-frequency words. This latter pattern is consistent with Weekes's (1997) initial observation that length appears to exert a stronger influence on low-frequency words.

Turning to lexical decision performance, length accounted for less variance than it did in the naming task, as reflected by the reliable effect of task in the subject-level regression analyses. The main effect of length was in fact eliminated for the young adult response latencies, although the effect did appear in the accuracy data. This null effect of length is qualified by a significant Length × Word Frequency interaction, entered in Step 3 of the regression analyses. This interaction term was highly reliable for both young ($p < .001$) and older adults ($p < .01$) and is captured in the two rightmost sections of Figure 12, where we show how the standardized regression coefficient of length varies as a function of age and objective frequency. As shown, length becomes less inhibitory as word frequency increases for both young and older adults, mimicking the effects in naming.

In sum, length is a powerful predictor of naming performance and is a less powerful predictor of lexical decision performance. There is also evidence of a Length × Word Frequency interaction in both naming and lexical decision performance. The length of a word slows response latency primarily for lower frequency words. Clearly, current models of lexical processing must take into consideration the influence of stimulus length and its interactive effects with word frequency.

Orthographic Neighborhood Size

A second variable that has received considerable attention in the lexical processing literature is orthographic neighborhood size. As Andrews (1997) pointed out, the theoretical importance of orthographic neighborhood size is that two of the major models of lexical processing, the serial search models (Forster, 1976; Paap & Johansen, 1994) and the interactive activation models (e.g., McClelland & Rumelhart, 1981), appear to predict that increasing orthographic neighborhood size should increase response latencies in lexical processing tasks. For example, according to serial search models, search sets are defined by orthographic similarity, and hence one would expect that either words with many neighbors would have larger search sets, or, if search set is held constant, subjects would have more difficulty searching through search sets with highly similar orthographically related neighbors (see, however, Forster & Shen, 1996, for an alternative account). Turning to

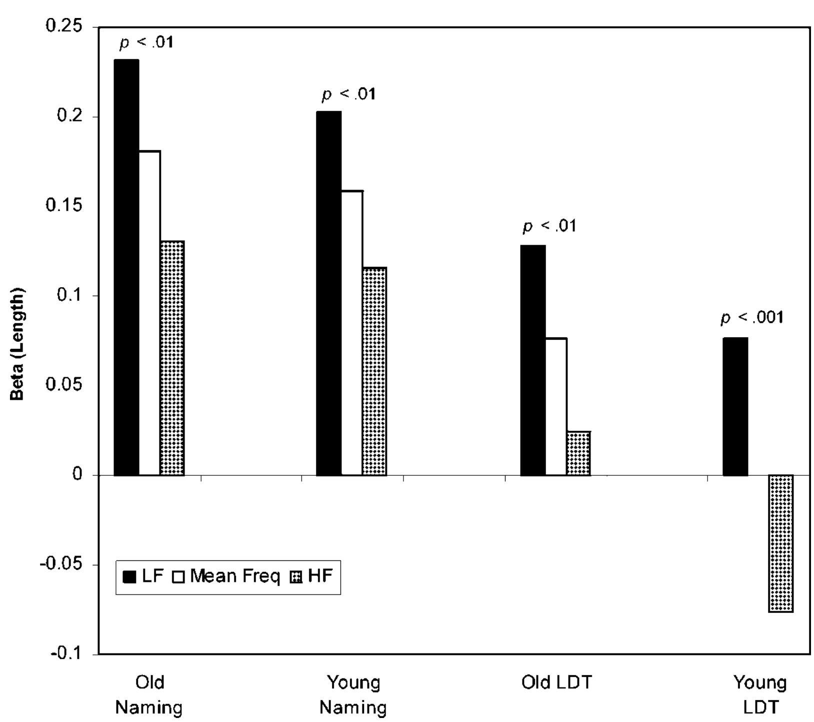


Figure 12. Word Frequency × Word Length interaction as a function of age and task. LF = low frequency; Mean Freq = mean frequency; HF = high frequency; Old Naming = older adults in the naming task; Young Naming = young adults in the naming task; Old LDT = older adults in the lexical decision task; Young LDT = young adults in the lexical decision task.

the interactive activation framework, there are inhibitory connections between lexical candidates, and so there should be more within-level inhibition from words that have more orthographic overlap, that is, orthographic neighbors. The empirical problem that arose in this literature is that there is typically a facilitatory influence of orthographic neighbors instead of the inhibitory influence predicted by the models. However, as Andrews noted, it is critical to consider the specific characteristics of the task to better understand the nature of orthographic neighborhood size effects.

First, consider the literature concerning orthographic neighborhood size and lexical decision performance. The results here have been complex to say the least. For example, in the initial study of neighborhood density, Coltheart et al. (1977) found no effect of neighborhood size for words but a strong inhibitory effect for nonwords. Grainger (1990, 1992), Grainger and Jacobs (1996), and Carreiras, Perea, and Grainger (1997) found evidence for inhibitory influences of neighborhood frequency, that is, words with higher frequency neighbors produced slower lexical decision latencies, consistent with the extant theoretical perspectives. In contrast, Andrews (1989, 1992) and Forster and Shen (1996) found facilitatory effects of neighborhood density. Sears, Hino, and Lupker (1995) also found facilitatory effects of neighborhood density; however, in four of the five studies reported, the effect of neighborhood density did not reach significance by items, suggesting that the effects may not generalize beyond a specific set of items. Johnson and Pugh (1994) found facilitatory effects when illegal nonwords were used and inhibitory effects when legal nonwords were used. Interestingly, Forster and Shen were concerned about possible item-selection problems in this literature and suggested that a multiple regression approach may be a better way of tackling this issue.

The results of the present regression analyses of lexical decision performance are clear. With young adults, there is a replication of the original Coltheart et al. (1977) observation. Specifically, there is no evidence of a unique neighborhood size main effect across a large set of single-syllable words, but there is evidence of a large inhibitory effect for nonwords that are based on these words. However, the null neighborhood size effect in young adults is qualified by a significant interaction between neighborhood size and log frequency ($p < .01$), which we tested by entering the product of the two predictor variables in a third step, as described above. As shown in Figure 13, this interaction reflected the finding that neighborhood size facilitated response latencies for low-frequency words but actually produced some inhibition for high-frequency words. These results are remarkably compatible with Andrews's (1989) and Forster and Shen's (1996) finding that neighbor size facilitated lexical decision performance only for low-frequency words and that neighborhood size effects were either unreliable or inhibitory for high-frequency words (Andrews, 1989, Experiment 2). Within the dual-route model, high-frequency words should be more likely to be influenced by lexical processes than low-frequency words would be. If lexical access involves competition among neighbors (McClelland & Rumelhart, 1981), it is possible that high-frequency words would show greater inhibitory effects of neighborhood size. As described below, there is evidence that neighborhood size consistently helps naming perfor-

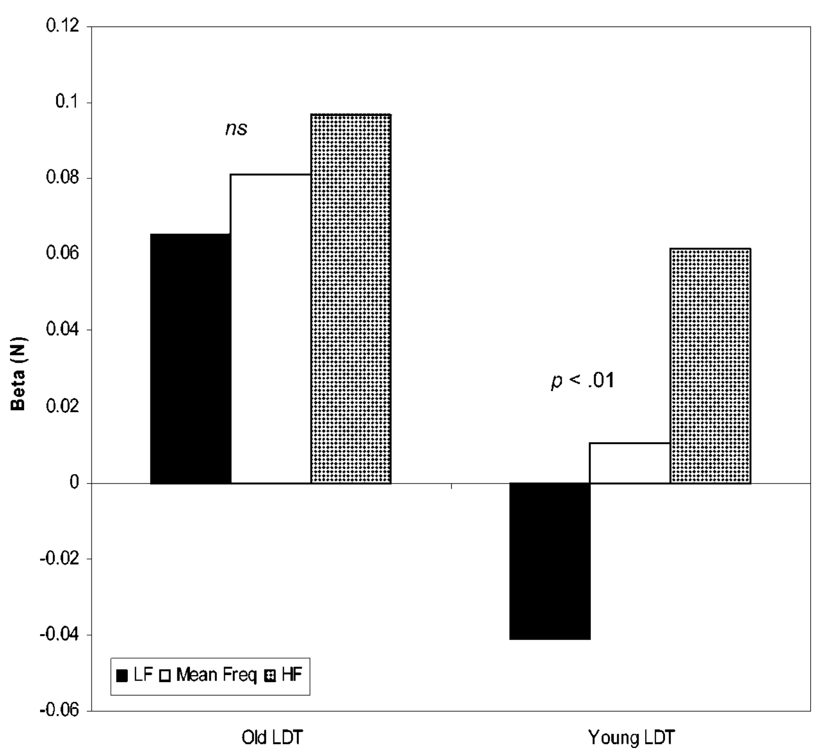


Figure 13. Word Frequency × Orthographic Neighborhood interaction in lexical decision performance as a function of age. LF = low frequency; Mean Freq = mean frequency; HF = high frequency; Old LDT = older adults in the lexical decision task; Young LDT = young adults in the lexical decision task.

mance (Andrews, 1989), supporting the idea that neighborhood size facilitates the nonlexical mappings between spelling and sound. Because low-frequency words are more likely to be influenced by nonlexical processes, one would predict facilitatory neighborhood size effects for such words, which is what we obtained.

In contrast, the older adults produced a consistent inhibitory influence of neighborhood size in lexical decision performance. The larger inhibitory effects of orthographic neighborhood in older adults, compared with those found with young adults, may reflect speed-of-processing differences between older adults and young adults. If neighborhood size effects are modulated by processing speed, one would expect slow young adults to behave more like older adults than fast young. We conducted a median split on the young adults on the basis of their response latencies, and for each speed group, we tested the interaction between neighborhood size and log frequency after partialing out the standard variables in Steps 1 and 2. The interaction was reliable for the fast young adults ($p < .01$) but not for the slow young adults. As shown in Figure 14, as predicted, the interactions indicated that the slow young adults mirrored the pattern obtained for the older adults. Because slow young adults take a longer time to respond, one might argue that lexical activation and competition have more time to build up and exert an effect.

In sum, neighborhood size is inhibitory in older adults' and in slow young adults' lexical decision performance. However, for young adults' low-frequency word performance, the effect of neighborhood size is facilitatory. (This interaction was also replicated, $p < .01$, in the lexical decision data obtained from the ELP.) Orthographic neighborhood size effects in lexical decision appear to vary as a function of word frequency (Andrews, 1989) and processing speed of the subjects. Possibly some of the controversies in the past literature may reflect differences in item and subject samples. Turning to the nonwords, consistent with the available literature, both the young adults and the older adults produced large inhibitory effects of neighborhood density.

Andrews (1997) noted that the influence of neighborhood size, in contrast to its role in lexical decision performance, has been consistently facilitatory in speeded naming performance. Specifically, studies by Andrews (1989, 1992), Grainger (1990), and Laxon, Coltheart, and Keating (1988) have all found facilitatory effects of neighborhood size on speeded naming performance. In the present naming results, neighborhood size produced a highly reliable and unique predictive effect in both young adults and older adults. It is interesting to note that this pattern was found above and beyond the orthographic and phonological consistency of the rime unit. As Andrews pointed out, one interesting question is the extent to which orthographic neighborhood size is simply a surrogate for orthographic rime consistency. The present study clearly indicates that this effect is independent of feedforward and feedback rime consistency.

We also tested whether any additional variance was accounted for by the interactive effects of neighborhood size and word length or neighborhood size and log frequency when these interactions were entered in the third step of the regression analyses. The latter test was based on the results by Andrews (1989, 1992), who found a larger effect of neighborhood size for low-frequency words as compared with high-frequency words. There was reliable additional unique variance picked up by the Neighborhood Size × Log Frequency interaction in both the young adults ($p < .001$) and the older adults ($p < .001$). As shown in Figure 15, both age groups exhibited larger facilitatory effects of neighborhood size for low-

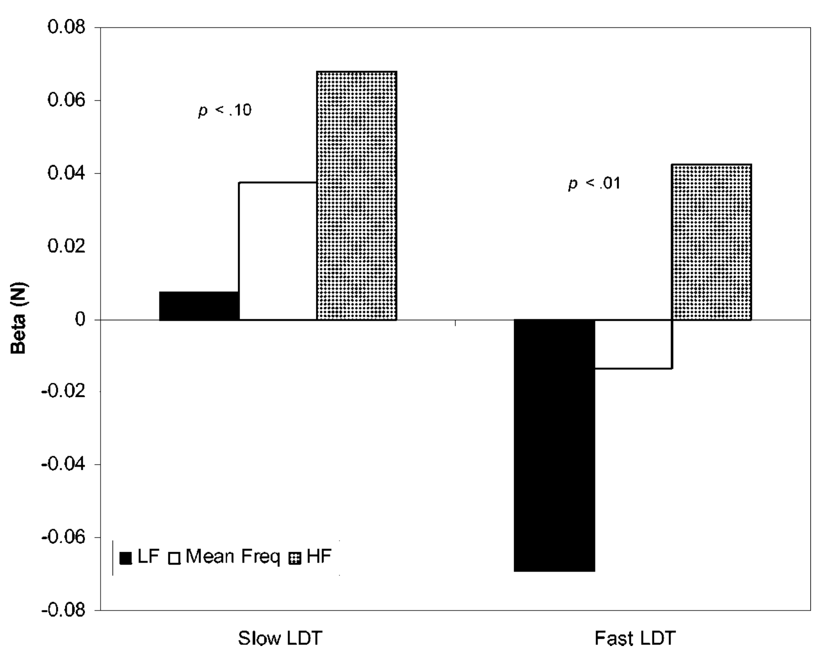


Figure 14. Word Frequency × Orthographic Neighborhood interaction in lexical decision as a function of processing speed in young adults. LF = low frequency; Mean Freq = mean frequency; HF = high frequency; Slow LDT = slow young adults in the lexical decision task; Fast LDT = fast young adults in the lexical decision task.

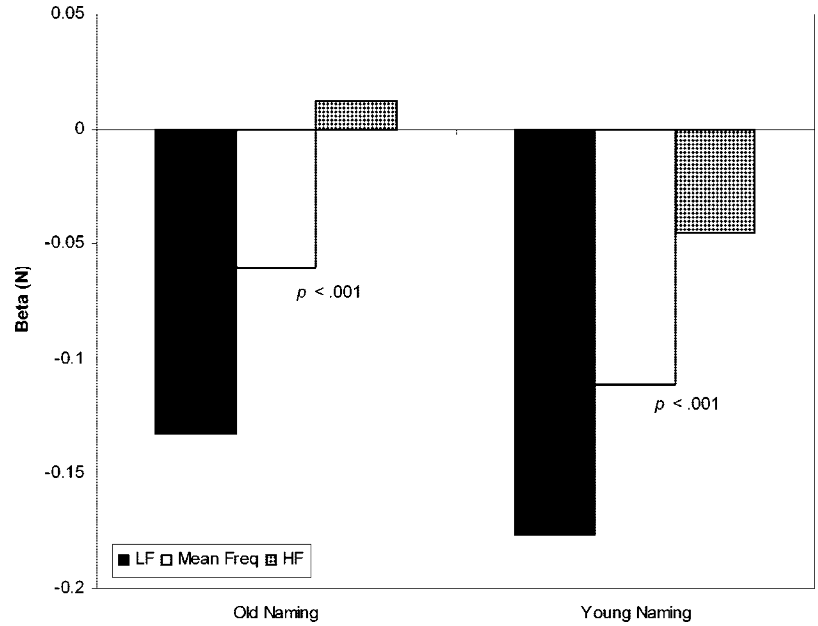


Figure 15. Word Frequency × Orthographic Neighborhood interaction in naming for both young and older adults. LF = low frequency; Mean Freq = mean frequency; HF = high frequency; Old Naming = older adults in the naming task; Young Naming = young adults in the naming task.

frequency words than for high-frequency words. (This interaction was replicated, $p < .01$, in the ELP naming data set.) Thus, neighborhood size plays an especially large and facilitatory role for low-frequency words in the speeded naming task for both young and older adults. Interestingly, as shown in Figure 16, this pattern is complemented by a significant Neighborhood Size × Length interaction ($p < .05$ for older adults and $p < .001$ for young adults) in speeded naming performance, which did not

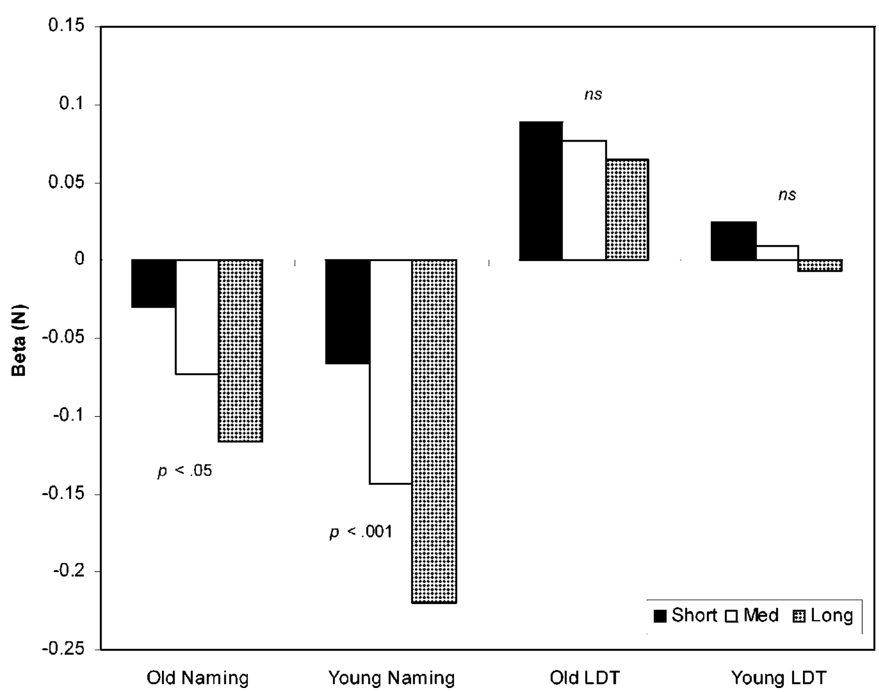


Figure 16. Orthographic neighborhood by length interaction as a function of age and task. Med = medium; Old Naming = older adults in the naming task; Young Naming = young adults in the naming task; Old LDT = older adults in the lexical decision task; Young LTD = young adults in the lexical decision task.

appear in lexical decision performance. Hence, it appears that neighborhood size facilitates naming latencies more for longer low-frequency words than for shorter high-frequency words. However, these appear to be additive effects, because the three-way interaction did not approach significance.

It is possible that the large orthographic neighborhood effects in naming reflect the mapping of letters, graphemes, and even higher level units onto phonological codes and that greater neighborhood density accelerates nonlexical recoding processes. The Word Frequency × Neighborhood Size interaction suggests that low-frequency words, which are more susceptible to nonlexical procedures, especially benefit from having many neighbors. This benefit carries over somewhat to lexical decision performance for fast young adults. High-frequency words, in contrast, are more likely to experience lexical influences in lexical decision. Competition among activated lexical candidates, as suggested in the original McClelland and Rumelhart (1981) framework, would slow down responses to words with many neighbors. Finally, one might argue that the consistent inhibitory effect for nonwords in lexical decision is due to an increase in familiarity that slows the nonword response in this task.

Feedforward and Feedback Consistency

Our measures of consistency were motivated by the empirical work of Jared et al. (1990), who found that the frequency of the stimulus word and the relative frequency of friends to enemies modulate naming performance. We focused on measures of consistency, as opposed to regularity, in the present study because Cortese and Simpson (2000) have demonstrated that consistency is a more powerful predictor variable than regularity is when the two factors are factorially crossed. As described earlier, in computing the consistency measures, we computed a token frequency measure that was based on the log frequency of the friends for a given unit (i.e., rime or onset in the feedforward or feedback direction) divided by the log frequency of the friends and enemies for that unit. Although we used token-based estimates of consistency in the present analyses, it should be noted that one finds the same pattern of reliable consistency effects when one considers a type consistency measure, which does not weight each friend and enemy by the log frequency of that item.

Before we discuss these results, note that the measures of consistency depend on the vocabulary set that one defines consistency against. In the present study, we used 4,444 single-syllable words from the item set used in the ELP that had Zeno et al. (1995) frequency estimates available. To ensure that our consistency measures were representative, we also included analyses that were based on consistency measures independently derived by Kessler et al. (2003) on a set of 3,690 single-syllable words. The correlations between the two sets of token consistency measures were quite high (onset feedforward = .95, onset feedback = .98, rime feedforward = .95, rime feedback = .94). As expected, when we entered the Kessler et al. token consistency estimates in our analyses, we found an identical pattern of reliable effects. Hence, the effects found in the present study do not appear to be biased by the sample of stimuli we used to define consistency.

A second issue that should be noted here is the possibility of correlated variables with our consistency measures. We attempted to address this by entering four additional variables in item-level analyses in Step 2 to determine if such potentially correlated variables modulate the obtained consistency effects. In particular, we obtained estimates from Kessler et al. (2003) to test the influence of the length in letters of the onset unit, length in letters of the rime unit, spelling frequency of the onset unit, and spelling frequency of the rime unit. It is possible that longer and less frequent units are more likely to produce lower consistency values, independent of the feedforward or feedback consistency of these values. In only one instance (feedback onset consistency for the young adult naming latencies) did the addition of these potentially correlated variables influence the pattern of reliable effects of consistency obtained from the present regression analyses.

Onset Consistency

Although the influence of onset consistency did not reach significance in the current naming latencies, the effect was in the predicted direction and there was a reliable effect in the accuracy data. Consistent with predictions made by the dual-route model, previous studies have found effects of onset consistency on naming latencies (e.g., Cortese, 1998; Kawamoto & Kello, 1999; Kessler et al., 2003; Treiman et al., 1995). Moreover, as shown in Figure 11, there was a reliable effect of onset consistency in the ELP naming data set.

Interestingly, feedforward onset consistency was related to lexical decision latencies and accuracy for both young and older adults. This is somewhat inconsistent with the traditional view that lexical decision is relatively impervious to feedforward consistency effects. However, as noted below, further inspection indeed indicates that there is a feedforward consistency effect in the available lexical decision literature.

Turning to feedback onset consistency, there was an effect of this variable in naming but not in lexical decision performance. This pattern was replicated across the two different data sets. However, as noted above, the effect of feedback onset consistency was eliminated in the young adult naming data but not the older adult data when spelling frequency of the onset unit was included as a control variable. Given these results, these data do not provide strong support concerning a unique effect of feedback onset consistency in young adult naming performance.

Rime Consistency

On the basis of the extant literature, at least two predictions can be made. First, feedforward rime consistency should predict naming performance (Jared et al., 1990; Treiman et al., 1995) more strongly than lexical decision performance (Jared et al., 1990). Second, feedback rime consistency should predict lexical decision performance (Stone et al., 1997) more strongly than naming performance (Ziegler, Montant, & Jacobs, 1997).

First, consider the influence of feedforward rime consistency. This variable consistently predicted naming performance more than lexical decision performance. This pattern is consistent with the results by Andrews (1982) and Jared et al. (1990), among others. The simplest interpretation of this pattern is that the consistency of the orthographic rime to phonological rime mapping modulates onset latencies in naming performance, because this is one of the primary codes that subjects use to drive pronunciation. It is interesting that Ziegler, Montant, & Jacobs (1997) and Inhoff and Topolski (1994) have reported that feedforward rime consistency effects persist in delayed naming tasks, suggesting that there may be an influence of this variable in output processes after lexical access. In a recent positron-emission tomography neuroimaging study of speeded naming, Fiez, Balota, Raichle, and Petersen (1999) reported that spelling-to-sound rime consistency modulates motor areas bilaterally. Thus, it appears that feedforward consistency effects may have multiple loci in speeded naming performance (also see Kinoshita & Woollams, 2002).

Although the effect was smaller, feedforward rime consistency also reliably facilitated lexical decision performance in both data sets. This would appear to conflict with the observation that feedforward rime consistency effects are not produced in lexical decision performance. However, a closer inspection of this literature reveals that a relationship indeed exists between feedforward rime consistency and lexical decision performance. First, although Jared et al. (1990) failed to find a consistency effect for lexical decision latencies, there was a significant consistency effect for errors (7.9% for inconsistent words and 4.3% for consistent words). Second, in Stone et al. (1997), when one collapses across feedback-inconsistent and -consistent words, there is a feedforward consistency effect, with faster latencies (755 ms vs. 775 ms) and fewer errors (5.5% vs. 12.4%) for consistent words. Also, Andrews (1982) clearly showed facilitatory effects of feedforward rime consistency in lexical decision. Although it is not yet clear how feedforward rime consistency operates in lexical decision, these findings are compatible with Frost's (1998) argument that phonology is a mandatory process in processing visual words (also see Yates, Locker, & Simpson, in press).

Turning to feedback rime consistency, the present results are intriguing on a number of levels. First, the feedback rime consistency effect was initially found in lexical decision (Stone et al., 1997; also see Pexman, Lupker, & Jared, 2001) and appeared to be stronger in lexical decision than in naming performance (Ziegler, Montant, & Jacobs, 1997). This initial observation was particularly important because it suggested a role for feedback from phonology to orthography and hence was interpreted to support a lexical resonance model. Specifically, the correspondence between phonological codes and orthographic codes in visual word processing may support a feedback process that might facilitate the pattern of activation's settling into a consistent pattern (also see Edwards, Pexman, & Hudson, in press; Pexman et al., 2001). However, when Peereman et al. (1998) attempted to replicate the Ziegler, Montant, and Jacobs finding of a feedback consistency effect in French, Peereman et al. found that they could replicate the effect but that this pattern was likely due to the confounding of familiarity with feedback consistency (also see Kessler et al., 2003).

The present results yielded reliable effects of feedback rime consistency in naming performance after controlling for objective and subjective frequency estimates, along with other related variables. The effects were replicated in both young and older adults, as well as in the ELP database. We also found an effect of feedback rime consistency in lexical decision performance, but this effect was relatively small and only occurred for older adults in the item analyses. (This effect was reliable in the subjects-level analyses for both young and older adults.) The larger effect in the older adults may indicate that slower response latencies may afford more time for feedback rime consistency to play a role. To test this speed-of-processing explanation, we again conducted a median split on just the younger adults, based on their overall mean word response latencies. We then entered the same set of standard variables into the items-level regression equation for both Step 1 and Step 2 for lexical decision performance. The feedback rime consistency was somewhat larger for the slow young adults ($-.024$) than for the fast young adults ($-.007$), suggesting that speed of processing may modulate the presence of a feedback consistency effect. We tested the same variable with subject-level regression analyses. Feedback rime consistency is significant only for the slow young adults ($p = .03$) in lexical decision performance, confirming the idea that the feedback rime consistency effect becomes more salient in slow subjects.

In summary, the present results provide evidence of feedback rime consistency effects in naming and in lexical decision performance, particularly for the slow subjects. In light of these results and given the controversy that this area has generated, we are inclined to believe that the relationship between feedback rime consistency and word recognition deserves further study. Feedback rime consistency effects are theoretically quite intriguing in supporting a highly interactive system in which the consistency of the mapping of the phonological information onto spelling patterns contributes to the naming response as it unfolds across time.

Objective and Subjective Word Frequency

The present results were intriguing on a number of dimensions with respect to the influence of objective and subjective word frequency. First, consider the between-task comparisons. The regression analyses clearly indicated that the predictive power of both frequency estimates was much larger in lexical decision than in naming performance. There are two major accounts of this task difference in the size of frequency effects. One account is based on the dual-route perspective of speeded word naming, in which there is a lexical route that is frequency modulated and a sublexical route that is relatively independent of word frequency (see, e.g., Coltheart et al., 2001). The notion is that subjects can rely on the frequency-independent spelling-to-sound route in naming but not lexical decision performance. This sublexical route may be more influential for lower frequency words, where the lexical route is slower to generate an output, thereby facilitating naming performance for low-frequency words. In support of this position, Monsell, Doyle, and Haggard (1989) provided evidence that the word frequency effect is comparable in naming and lexical decision when one considers orthographic patterns that have irregular mappings onto phonology. The notion is that for irregular words, the lexical route must drive the response in both naming and lexical decision performance, because the sublexical route would produce an error for these items, that is, pronouncing the word *pint* such that it rhymes with *hint*. An alternative account of the smaller word-frequency effect in naming than in lexical decision is simply that the word-frequency effect becomes exaggerated in lexical decision because the constraints of the task emphasize frequency-based information in order for subjects to make the discrimination between familiar words and unfamiliar nonwords. Specifically, high-frequency words are more discriminable from the zero-frequency nonword stimuli than are low-frequency words.

To discriminate between these two accounts, we identified 120 words (approximately 5% of the total data set) that were the most feedforward rime inconsistent. The prediction from the dual-route perspective is that these items would be most likely to drive the lexical route for a correct pronunciation, and so the frequency effects in naming and lexical decision performance should be comparable. To directly compare the two tasks, we conducted hierarchical regression analyses on each subject with the same two steps as in the standard regression analyses and then entered the standardized regression coefficients into an ANOVA to determine if there is a reliable effect of task. There was still a larger effect of subjective word frequency in lexical decision than in naming, $p <$ .01, even for these highly inconsistent words. Hence, it is not the case that naming and lexical decision performance produce comparable word-frequency effects for the words with the most inconsistent spelling-to-sound correspondences in the database.

It is also noteworthy that there was a significant nonlinear relation between log frequency and response latency in lexical decision performance, and this nonlinear component did not occur in naming performance (see Figures 8A–8D). This suggests that a log transformation does not capture the additional increase in response latency for very low-frequency words in lexical decision performance. This finding is at least compatible with the notion that low-frequency words are disproportionately more disrupted by their similarity to the nonword targets than are high-frequency words (see Murray & Forster, in press, for an alternative interpretation). Of course, there has been some speculation in the literature that low-frequency words may be more likely to be engaged in qualitatively different processes, such as retrieving the meaning or spelling of the stimulus word (e.g., Balota & Chumbley, 1985; Besner & Swan, 1982). This framework is also compatible with the fact that the LDT produced a much more skewed RT distribution than the naming task did (see the discussion of the ex-Gaussian analyses in the *RT Distribution Analyses* section).

Finally, there is an intriguing age dissociation involving the objective and subjective frequency estimates. Older adults were influenced more than young adults by objective word frequency, whereas young adults were more influenced than older adults by subjective word frequency. Both of these effects were highly reliable in the subject-level analyses. This may reflect cohort differences. The subjective frequency estimates were based on estimates from young adults, and these estimates may be less appropriate for older adult subjects. Fortunately, we have available subjective frequency estimates using the same procedure from 90 healthy older adults on a randomly selected subset of 480 words. There was still a larger influence of subjective frequency for young adults (−.48) than for older adults (−.40). Hence, this difference does not appear to be due to simple cohort effects.

Of course, one may make the same cohort argument regarding the larger objective frequency effects in older adults compared with young adults. In the present study, we used the objective frequency estimates based on the Zeno et al. (1995) norms. Because this corpus was based on printed material from a variety of sources, it is likely that these frequency estimates may be more tuned to the older adult lexicon than the young adult lexicon. To address this possibility, we used more recent frequency-based information provided by Burgess (see Burgess & Livesay, 1998), which consists of approximately 131 million observations based on all Usenet newsgroups during the month of February in 1995. These should be more tuned to the young adult lexicon than are the Zeno et al. norms. It is interesting that the HAL estimates of objective word frequency (Lund & Burgess, 1996) were more highly correlated with lexical decision performance for young adults (.31) than for older adults (.22), with little difference for naming performance (.09 vs. 12 for young and older adults, respectively). The pattern for lexical decision performance is, of course, opposite of what one finds for the Zeno et al. norms. Hence, although measures from the HAL corpus and the Zeno et al. corpus are highly correlated (e.g., r = .86), there may be subtle cohort effects that make between-age-group comparisons difficult. To our knowledge, this is the first demonstration of such cohort effects of word frequency. It is quite possible that with a smaller scale study, we would not be able to detect such cohort effects.

In sum, the present analyses of subjective and objective word frequency have yielded a number of intriguing findings for monosyllabic words. First, there are powerful effects of subjective word-frequency estimates for both naming and lexical decision performance, which for young adults are actually larger than objective word-frequency estimates. Hence, it appears that subjects (especially young adults) have good metacognitive insights into their frequency of exposure to words. Second, the effects of word frequency are much stronger in lexical decision than in naming, even for highly inconsistent words. Third, in the LDT, the nature of the word-frequency effect is nonlinear even when log transform is used as a predictor, possibly suggesting a qualitatively distinct process for very low-frequency words. Fourth, one can detect subtle cohort effects when using different objective word-frequency estimates across different age groups. This is particularly noteworthy for studies of age-related changes in lexical processing.

Semantic Variables

The present study included three sets of analyses on semantic variables. In the first set of analyses, we used a restricted set of 996 words (the items are available in both the Toglia and Battig, 1978, and the Nelson et al., 1998, norms) and addressed the predictive power of some of the standard semantic variables explored in the literature. We took a relatively conservative approach and added these variables in the third step, after both subjective and objective frequency were partialed out. This was done because previous researchers suggested that the initial evidence for semantic effects was most likely due to uncontrolled influences of other variables, such as familiarity (see Gernsbacher, 1984). The present results provided some evidence of an influence of Toglia and Battig's (1978) imageability estimates on lexical decision performance for both young and older adults and, to a lesser extent, on young adults' naming performance. Toglia and Battig's meaningfulness estimates and Nelson et al.'s (1998) semantic set size estimates consistently predicted only young adults' lexical decision performance. The larger effects of these semantic variables in lexical decision performance were expected because this task places a greater emphasis on the meaningfulness of the stimulus as a useful dimension to make the word–nonword discriminations. Of course, we are not the first to provide evidence of an influence of meaning variables in lexical decision performance (e.g., Hino & Lupker, 1996; James, 1975; Jastrzembski, 1981; Locker, Simpson, & Yates, 2003; see Balota, Ferraro, & Conner, 1991, for a review).

As noted, the researchers making the original observations of meaning influences in lexical decision performance were criticized for not controlling for familiarity of the stimulus (see Gernsbacher, 1984). However, subjects may also rely on meaning information when making untimed global familiarity estimates, and so by controlling for familiarity, one may be throwing out the baby with the bathwater. In fact, Balota et al. (2001) have shown that traditional familiarity estimates are strongly related to meaning variables. This is precisely why we developed the subjective frequency estimates, which are less strongly related to meaning variables than are standard familiarity estimates (see Balota et al., 2001, for a discussion). The important observation is that subjective frequency estimates were indeed a powerful predictor of performance above and beyond objective frequency, and yet meaning-level variables still accounted for a reliable amount of variance in the third step of the hierarchical regression analyses. Hence, lexical decision performance does not appear to provide a window into the magic moment of word recognition, that is, the point in time where the word is recognized before meaning has been accessed.

Interestingly, there was also a consistent effect of imageability on speeded naming performance. This was highlighted by the imageability norms developed by Cortese and Fugett (2003) on the full set of items. This effect occurred in both the item- and subject-level analyses. The reliable effect of imageability in naming performance was replicated in the item-level analyses from the ELP naming database. Hence, even in a task that does not emphasize the discrimination between meaningful word stimuli and nonmeaningful nonword stimuli, one can still obtain an influence of meaning. Again, this effect is above and beyond the influence of a host of variables that influence speeded naming performance. We believe these results are most consistent with a view in which meaning becomes activated very early on, in a cascadic manner, during lexical processing and contributes to the processes involved in reaching a sufficient level of information to drive a lexical decision or a naming response.

In addition to using more standard measures of semantic information, we also conducted a set of semantic analyses motivated by recent work by Steyvers and Tenenbaum (2004). As noted, these researchers analyzed three large databases (*Roget's Thesaurus of English Words and Phrases* [Roget, 1911], WordNet norms [Miller, 1990], and the Nelson et al., 1998, norms) to determine if these networks had what they referred to as small-world structure. Thus, they calculated the degree of connectivity among the words in each of the norms, as reflected by overlap in meaning in the *Roget's Thesaurus* and WordNet source material and number of connections in the Nelson et al. production norms. Small-world structure is reflected by sparse connectivity between nodes, relatively short average distance between any two nodes, a large degree of local clustering, and the finding that connectivity across words follows a power function. As noted, Steyvers and Tenenbaum found evidence of such small-world structure in each of the three large databases that they measured. Moreover, they demonstrated that other recent ways of attempting to ground semantics, such as semantic latent analysis (e.g., Landauer & Dumais, 1997), do not produce such power functions. Steyvers and Tenenbaum suggested that such structure may naturally develop out of emerging semantic networks, wherein a relatively small set of concepts is central in the network (i.e., produces a large degree of clustering), and these concepts serve as the hubs of communication for the rest of the network.

In the present study, we investigated the influence of sheer connectivity based on the databases from WordNet (Miller, 1990) and the Nelson et al. (1998) connectivity measures. Again, we entered the semantic connectivity measures after all the lexical variables were entered into the regression equation. Both measures of connectivity accounted for a reliable amount of variance in both young' and older adults' naming and lexical decision in the subjects analyses, although the latter task, as predicted, produced the larger effects of connectivity. Steyvers and Tenenbaum (2004) also reported evidence in support of this general pattern in the present lexical decision data and also in a smaller set of naming data; however, they did not partial out the variance attributable to the phonological onsets and all of the lexical variables that were entered in the present study. We believe that the small-world structure identified by Steyvers and Tenenbaum in semantic memory may be useful in understanding the organization and development of lexical meaning (see Buchanan, Westbury, & Burgess, 2001, for a similar approach).

Finally, we explored interactions (both two-way and three-way) among the meaning level variables and other variables, such as word frequency and feedforward and feedback consistency. This was motivated by the work of Strain et al. (1995) suggesting that imageability plays more of a role in speeded word naming for items that are rather difficult to name because of their low frequency and inconsistent spelling-to-sound correspondence (also see Hino & Lupker, 1996). Although our regression coefficients were in the same direction as that predicted by Strain et al., they did not reach significance (ps $>$ .20). Hence, the unique interactive effect of meaning-level variables and other variables is relatively modest across the full set of single-syllable words.

Future Directions

Although the present results yielded a number of important observations regarding lexical processing via regression analyses, there are a number of further issues that need to be addressed. Before concluding, we believe it is important to acknowledge these issues and suggest some possible directions for future work.

First, how does one settle on the critical set of predictor variables in the regression equation? On the basis of the available lexical processing literature, one could easily double the number of predictor variables to explore in multiple regression analyses. For example, one might be interested in the number of higher frequency orthographic neighbors or some relative index of orthographic neighborhood that takes into account the frequency of the word and the frequencies of the orthographic neighbors. There are also additional ways to measure consistency. For example, one might argue that a type consistency measure (which is the ratio of friends to total friends plus enemies) as opposed to a token consistency measure (which weights each friend and enemy on the basis of its log-frequency value) may be more appropriate (see Kessler et al., 2003). Although we explored other variables in our initial analyses, we eventually settled on a set of theoretically motivated variables. Our goals were to identify commonly studied variables in the literature, attempt to minimize overlap with the predictor variables, and investigate variables that have produced some controversy in the available literature. This list of predictor variables clearly is not all inclusive and is driven by the goals of the given project. As in the case of factorial designs, one must interpret the unique effect of a given variable in the context of the other variables that are accounted for.

A second issue is the order of entry of the variables into the regression model. We entered the onset variables in the first step, lexical variables in the second step, and semantic variables in the third step. This was motivated by the possibility that the phonological onset variables may capture voice key sensitivity, and we wanted to explore the influence of lexical variables on naming performance after this potentially contaminating effect was removed. We investigated the influence of the semantic variables in the third step because, as noted above, there has been some controversy regarding the unique influence of these variables above and beyond correlated variables, such as familiarity. However, one could argue that the semantic variables should have been entered earlier in the regression model. In fact, one might argue that semantic information is indeed strongly tied to the role of frequency-based information (see, e.g., Plaut et al., 1996, for such an instantiation).

Third, one might argue that regression analyses are relatively difficult to interpret when one is interested in interactive effects of variables. This is a fair criticism if one is more accustomed to the factorial designs that dominate the word recognition literature. However, it is also the case that it is difficult to control for all potentially extraneous variables in each of the cells of a factorial design. We have approached this issue via the examination of interactive effects after we partialed out a series of main effects, that is, controlling for the potentially contaminating variables. Although the results from the analyses of interactions were broadly consistent with the available literature, these results also indicate a need for further exploration.

Fourth, the present analyses relied primarily on linear regression, and it is quite likely that nonlinear regression models may ultimately account for more variance and will be more compatible with predictions from computational models. An example of this in the present results is the influence of word frequency, in which there is a substantial nonlinear component in lexical decision performance after one transforms the word-frequency measure into a log scale. This nonlinear component did not occur in naming performance. We used the simplifying linear analyses because these are the most common approaches to speeded lexical processing and we were interested in directly comparing naming and lexical decision performance. However, it is likely that the next wave of understanding the influence of variables on lexical processing tasks will involve nonlinear influences and, importantly, the manner in which computational models of word recognition can capture such nonlinear influences.

Conclusions

The present study provides an analysis of the monomorphemic single-syllable words from the Kučera and Francis (1967) norms that have been critical in developing the current computational models of word recognition. We have studied these items across the two standard lexical processing tasks (naming and lexical decision) and across two different age groups. We have explored standard predictor variables that have been theoretically motivated by the literature and have been shown to produce stable influences on at least one of the tasks in the literature. The results highlight the differences between naming and lexical decision, suggesting that each task brings online a distinct set of processes. In fact, virtually every variable identified produced a highly reliable effect of task in the subject-level analyses. Because of the importance of task analyses, it will be particularly important to extend these observations to other measures of lexical processing, such as eye fixation durations (see Juhasz & Rayner, 2003; Schilling, Rayner, & Chumbley, 1998, for similar multiple regression approaches with online reading measures).

We also believe that some inconsistencies in the available literature may have arisen from item-selection effects. Allowing the language to define the stimulus set has some advantages over selecting items for specific cells of complex designs. Clearly, multiple regression techniques will not replace well-controlled factorial designs that are theoretically driven. It is likely that the two approaches will provide complementary constraints on theory development. Ultimately, however, we believe that large data sets will be particularly useful for progress in this field so that researchers can access a common set of items, which could provide preliminary tests of experimental hypotheses via either large-scale regression approaches or more specific tests on selected items taking a factorial approach. Our data are available for such tests, and, in the near future, larger, more comprehensive data sets will become available for researchers to access (see the ELP Web site at http://elexicon.wustl.edu/). This should lead to a more cumulative development of knowledge in the word recognition literature.

References

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition, 10,* 565–575.

Andrews, S. (1989). Frequency and neighborhood size effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 802–814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 234–254.

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

Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 545–555.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database.* Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Balota, D. A. (1994). Visual word recognition: The journey from features to meaning. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 303–348). San Diego, CA: Academic Press.

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 and 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 and Language, 24,* 89–106.

Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., & Treiman, R. (2002). *The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords*. Accessed January 30, 2004, on the Washington University Web site: http://elexicon.wustl.edu

Balota, D. A., & Duchek, J. M. (1988). Age-related differences in lexical access, spreading activation, and simple pronunciation. *Psychology and Aging, 3,* 84–93.

Balota, D. A., & Faust, M. (2001). Attention in dementia of the Alzheimer type. In F. Boller & S. Cappa (Eds.), *Handbook of neuropsychology* (2nd ed., Vol. 6, pp. 51–80). New York: Elsevier Science.

Balota, D. A., Ferraro, F. R., & Connor, L. T. (1991). On the early influence of meaning in word recognition: A review of the literature. In P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 187–222). Hillsdale, NJ: Erlbaum.

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

Balota, D. A., & Spieler, D. H. (1998). The utility of item-level analyses in model evaluation: A reply to Seidenberg and Plaut. *Psychological Science, 9,* 238–240.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128,* 32–55.

Bashore, T. R. (1994). Some thoughts on neurocognitive slowing. *Acta Psychologica, 86,* 295–325.

Besner, D., & Bourassa, D. C. (1995, June). *Localist and parallel processing models of visual word recognition: A few more words.* Paper presented at the annual meeting of the Canadian Brain, Behavior and Cognitive Science Society, Halifax, Nova Scotia, Canada.

Besner, D., & Swan, M. (1982). Models of lexical access in visual word recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 34A,* 313–325.

Brinley, J. F. (1965). Cognitive sets, speed and accuracy of performance in the elderly. In A. T. Welford & J. E. Birren (Eds.), *Behavior, aging, and the nervous system* (pp. 114–149). Springfield, IL: Charles C Thomas.

Broadbent, D. E. (1967). Word-frequency effects and response bias. *Psychological Review, 74,* 1–15.

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review, 8,* 531–544.

Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting RT in a basic word recognition task: Moving on from Kučera and Francis. *Behavior Research Methods, Instruments, & Computers, 30,* 272–277.

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of the orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology : Learning, Memory, and Cognition, 23,* 857–871.

Catlin, J. (1973). In defense of sophisticated guessing theory. *Psychological Review, 80,* 412–416.

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

Cerella, J., & Fozard, J. L. (1984). Lexical access and age. *Developmental Psychology, 20,* 235–243.

Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition, 12,* 590–606.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7,* 249–253.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100,* 589–608.

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., Patterson, K., & Marshall, J. C. (1980). *Deep dyslexia*. London: Routledge & Kegan Paul.

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.

Cortese, M. J. (1998). Revisiting serial position effects in reading aloud. *Journal of Memory and Language, 39,* 652–666.

Cortese, M. J., & Fugett, A. (2003). *Imageability ratings for 3,000 monosyllabic words*. Manuscript in preparation.

Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are they? *Memory & Cognition, 28,* 1269–1276.

Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review, 4,* 226–231.

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.

Edwards, J. D., Pexman, P. M., & Hudson, C. E. (in press). Exploring the dynamics of the visual word recognition system: Homophone effects in LDT and naming. *Language and Cognitive Processes*.

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.

Fiez, J. A., Balota, D. A., Raichle, M. E., & Petersen, S. E. (1999). Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. *Neuron, 24,* 205–218.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

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

Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition, 28,* 1109–1115.

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 696–713.

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 361–379.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123,* 71–99.

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.

Glanzer, M., & Ehrenreich, S. L. (1979). Structure and search of the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 18,* 381–398.

Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior, 22,* 24–44.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language, 29,* 228–244.

Grainger, J. (1992). Orthographic neighborhoods and visual word recognition. In R. Frost (Ed.), *Orthography, phonology, morphology, and meaning* (pp. 131–146). Oxford, England: North-Holland.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518–565.

Henderson, L. (1982). *Orthography and word recognition in reading*. London: Academic Press.

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1331–1356.

Humphreys, L. G. (1978). Research on individual differences requires correlational analysis, not ANOVA. *Intelligence, 2,* 1–5.

Inhoff, A. W., & Topolski, R. (1994). Use of phonological codes during eye fixations in reading and in on-line and delayed naming tasks. *Journal of Memory and Language, 33,* 689–713.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). Implications of behavioral decision theory and social marketing for designing social action programs. In J. Edwards, R. S. Tindale, L. Heath, & E. J. Posavac (Eds.), *Social influence processes and prevention* (pp. 103–142). New York: Plenum Press.

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

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29,* 687–715.

Jastrzembski, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology, 13,* 278–305.

Johnson, N. F., & Pugh, K. R. (1994). A cohort model of visual word recognition. *Cognitive Psychology, 26,* 240–346.

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1312–1318.

Kawamoto, A. H., & Kello, C. T. (1999). Effect of onset cluster complexity in speeding naming: A test of rule-based approaches. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 361–375.

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language, 47,* 145–171.

Kessler, B., Treiman, R., & Mullennix, J. (2003). *A new look at the factors that affect oral word reading*. Manuscript submitted for publication.

Kinoshita, S., & Woollams, A. (2002). The masked onset priming effect in naming: Computation of phonology or speech planning? *Memory & Cognition, 30,* 237–245.

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

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

Laxon, V. J., Coltheart, V., & Keating, C. (1988). Children find friendly words friendly too: Words with many orthographic neighbors are easier to read and spell. *British Journal of Educational Psychology, 58,* 103–119.

Locker, L., Simpson, G. B., & Yates, M. (2003). Semantic neighborhood effects on the recognition of ambiguous words. *Memory & Cognition, 31,* 505–515.

Lorch, R. F., Jr., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 149–157.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology, 3,* 215–233.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

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

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 570–590.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19–40.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113,* 181–190.

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.

MetaMetrics, Inc. (2003). *MetaMetrics word frequency counts* [Database]. Durham, NC: Author. (Available from MetaMetrics, Inc., Developers of the Lexile Framework, attention: A. Jackson Stenner, 2327 Englert Drive, Suite 300, Durham, NC 27713)

Miller, G. A. (1990). WordNet: An on-line lexical database. *International Journal of Lexiography, 3, 235-312*.

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.

Monsell, S., Patterson, K., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translations of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 452–467.

Murray, W. S., & Forster, K. I. (in press). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*.

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.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved June 1, 2000, from the University of South Florida Web site: http://www.usf.edu/FreeAssociation/

Paap, K. R., & Johansen, L. S. (1994). The case of the vanishing frequency effect: A retest of the verification model. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1129–1157.

Peereman, R., Content, A., & Bonin, P. (1998). Is perception a two-way street? The case of feedback consistency in visual word recognition. *Journal of Memory and Language, 39,* 151–174.

Perfetti, C. A. (1994). Psycholinguistics and reading ability. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 849–894). San Diego, CA: Academic Press.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988, February 25). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331,* 585–589.

Pexman, P. M., Lupker, S. J., & Jared, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 139–156.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Rastle, K., & Davis, M. H. (2003). *On the complexities of measuring naming*. Manuscript submitted for publication.

Rayner, K. (1998). Eye movements in reading and information processing: Twenty years of research. *Psychological Bulletin, 124,* 372–422.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

Roget, P. M. (1911). *Roget's thesaurus of English words and phrases.* Available from Project Gutenberg, Illinois Benedictine College, Lisle, IL.

Rosenthal, R. (1995). Critiquing Pygmalion: A 25-year perspective. *Current Directions in Psychological Science, 4,* 171–172.

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition, 26,* 1270–1281.

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

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Seidenberg, M. S., & Plaut, D. C. (1998). Evaluating word-reading models at the item level: Matching the grain of theory and data. *Psychological Science, 9,* 234–237.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23,* 383–404.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and post-lexical loci of contextual effects on word recognition. *Memory & Cognition, 12,* 315–328.

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.

Spieler, D. H., & Balota, D. A. (2000). Factors influencing word naming in younger and older adults. *Psychology and Aging, 15,* 225–231.

Steyvers, M., & Tenenbaum, J. B. (2004). *The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth*. Manuscript submitted for publication.

Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36,* 337–359.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1140–1154.

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word recognition. *Journal of Memory and Language, 26,* 608–631.

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

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 Memory and Language, 124,* 107–136.

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review, 7,* 424–465.

Watts, D. J., & Strogatz, S. H. (1998, June 4). Collective dynamics of small world networks. *Nature, 393,* 440–442.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 50A,* 439–456.

Yates, M., Locker, L., & Simpson, G. B. (in press). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review*.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science.

Zevin, J. D., & Balota, D. A. (2000). Priming and attentional control of lexical and sublexical pathways during naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 121–135.

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language, 47,* 1–29.

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review, 5,* 683–689.

Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language, 37,* 533–554.

Ziegler, J., Stone, G. O., & Jacobs, A. M. (1997). What is the pronunciation for *–ough* and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, & Computers, 29,* 600–618.

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 and Performance, 24,* 1131–1161.

Received April 16, 2002
Revision received January 19, 2004
Accepted January 26, 2004 ■