Journal of Experimental Psychology:
Learning, Memory, and Cognition
1995, Vol. 21, No. 5, 1140–1154

0278-7393/95/$3.00

Semantic Effects in Single-Word Naming

Eamon Strain and Karalyn Patterson
Medical Research Council Applied Psychology Unit

Mark S. Seidenberg
University of Southern California

Three experiments demonstrated that, for lower frequency words, reading aloud is affected not only by spelling–sound typicality but also by a semantic variable, imageability. Participants were slower and more error prone when naming exception words with abstract meanings (e.g., *scarce*) than when naming either abstract regular words (e.g., *scribe*) or imageable exception words (e.g., *soot*). It is proposed that semantic representations of words have the largest impact on translating orthography to phonology when this translation process is slow or noisy (i.e., for low-frequency exceptions) and that words with rich semantic representations (i.e., high-imageability words) are most likely to benefit from this interaction.

The basic processes specific to reading—that is, setting aside higher level syntactic and integrative processes presumably common to reading and speech comprehension—involve the transcoding of orthographic information into phonological and semantic representations. The nature of these basic processes is, of course, much debated in the literature on reading. One particular source of contention concerns the role of phonology in comprehension of a written word. Many theorists argue that because reading is a language activity and language is fundamentally phonological, the computation of a phonological representation for the written word is central to reading comprehension rather than an optional or strategic phenomenon (Carello, Turvey, & Lukatela, 1992; Lesch & Pollatsek, 1993; Van Orden, Pennington, & Stone, 1990; and many others). Other models, though perhaps not denying a role for phonology, maintain that the skilled reader also develops some degree of ability to compute meaning from orthographic patterns without reliance on phonology (Coltheart, Curtis, Atkins, & Haller, 1993; Jared & Seidenberg, 1991; Plaut & Shallice, 1993; Seidenberg & McClelland, 1989; and again, many others). Positions taken by individual theorists in this debate often seem to depend largely on intuition and personal preference ("models as toothbrushes," Watkins, 1984). To the extent that the debate has an empirical base, however, this consists largely of one main line of inquiry, at least with regard to normal adult readers. The question asked is whether and how processing of a written word is affected by its phonological characteristics.

In semantic categorization tasks, for example, word homophones or nonword pseudohomophones of real category exemplars provoke a significantly higher error rate than foil words with equivalent orthographic similarity that do not sound like real category exemplars (Coltheart, Patterson, & Leahy, 1994; Jared & Seidenberg, 1991; Van Orden, 1987; Van Orden, Johnston, & Hale, 1988; Wydell, Patterson, & Humphreys, 1993). For theorists who prefer phonological toothbrushes, the homophone effect constitutes evidence that reading comprehension is inescapably linked to phonology. Theorists who prefer a two-pronged brush of parallel orthographic and phonological access to meaning, on the other hand, tend to emphasize certain qualifications to the strength and ubiquity of the homophone effect (Coltheart et al., 1994; Jared & Seidenberg, 1991).

In our study, instead of asking whether performance in a comprehension task is affected by phonological characteristics of the written word, we turned the tables and asked whether performance in a phonological task—word naming—is affected by semantic characteristics of the word. The semantic dimension manipulated was imageability, that is, the extent to which the representation of a word's meaning has sensorimotor properties. Highly imageable words (like *corkscrew, red,* or *sparkling*) have meanings with many sensorimotor properties; low-imageability words like *naive* or *presumption* have more abstract meanings. We chose word imageability because it is one of the best predictors of oral-reading performance in certain acquired disorders of reading. Neurological patients with severe phonological deficits, whose residual reading abilities are assumed to rely mainly or solely on direct access from orthography to meaning, are often markedly more successful in reading imageable than abstract words (see, for example, Coltheart, Patterson, & Marshall, 1980; Funnell, 1987; Plaut & Shallice, 1993).

We are not the first to ask whether a semantic variable like imageability affects word naming by normal readers; however, previous experiments have yielded either very small (de Groot,

Eamon Strain and Karalyn Patterson, Medical Research Council (MRC) Applied Psychology Unit, Cambridge, United Kingdom; Mark S. Seidenberg, Neuroscience Program, University of Southern California.

The research was supported by National Institutes of Mental Health Grant MH 47566 and is part of a research program in collaboration with J. L. McClelland and D. C. Plaut, whose contributions we gratefully acknowledge. We are also grateful to Stephen Goldinger for numerous helpful comments from which the article benefited.

Correspondence concerning this article should be addressed to Eamon Strain, MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF United Kingdom. Electronic mail may be sent via Internet to eamon.strain@mrc-apu.cam.ac.uk.

1989) or nonsignificant (Brown & Watson, 1987) advantages for imageable words. There are principled reasons, however, for predicting that semantic effects should be modulated by other characteristics of the words not manipulated in these earlier studies. Because the task of interest is written-word pronunciation, performance should be governed primarily by the computation of phonology directly from orthography (hereinafter, orth-to-phon translation). If a word's semantic characteristics can have any influence on the process of naming it, this influence might be observed chiefly on words in which orth-to-phon translation is somewhat inefficient, slow, or error prone. What sorts of words are these? For both skilled human readers and computational models, extensive orth-to-phon training eventually results in correct translation of the majority of all trained words (Plaut & McClelland, 1993; Seidenberg & McClelland, 1989); however, both humans and models are always somewhat slower, more error prone, or both, on low-frequency words with inconsistent and atypical spelling–sound correspondences. We therefore predicted that low-frequency exception words, which have had the smallest impact on setting the weights for orth-to-phon translation, should be the items most likely to reveal a role for word meaning in word naming. Details of this idea are explored at greater length in the General Discussion section, after the results from three experiments designed to test this prediction are presented.

Experiment 1

In the first experiment we used a straightforward word-naming paradigm, with an orthogonal manipulation of word frequency, regularity, and imageability. Our prediction was that normal adult readers' accuracy and speed of word naming should reveal a three-way interaction between these variables, with a significant advantage for high- (over low-) imageability words—particularly for low-frequency exception words.

Method

Participants. The data were collected from 20 members of the Applied Psychology Unit participant panel who were paid for their participation. The participants were aged between 19 and 45 years, and about two thirds were women.

Materials and design. The experimental stimuli were selected from a pool of monosyllabic and disyllabic words for which imageability ratings were available. Most of the ratings came from the Medical Research Council (MRC) psycholinguistic database (Coltheart, 1981). The imageability ratings contained in this database are derived from a merging of the Colorado norms (Toglia & Battig, 1978), the Paivio norms (unpublished; these are an expansion of the norms of Paivio, Yuille, & Madigan, 1968), and the Gilhooley norms (Gilhooley & Logie, 1980). Imageability ratings for a set of 192 matched regular and exception words, obtained in a previous study using a sample of 20 Applied Psychology Unit panel members, were also used. Finally, ratings for a small number of words were collected from 40 staff members at the MRC Applied Psychology Unit before this experiment (discussed below). All ratings were on a scale ranging from 1 (*low imageability*) to 7 (*high imageability*). For words present in more than one source, the rating used was the average over sources. Words with ratings between 4.9 and 7, inclusive, were classified as high imageability, and words with ratings between 1 and 4.3, inclusive, were classified as low imageability. We chose these unequal ranges because the distribution of imageability values was slightly skewed toward the higher end of the imageability scale. Although these imageability ratings played an important role in the original selection of stimuli, they were mainly intended as initial guidelines. Because some of the norms used were both rather out of date and obtained from American samples (our participants were British), we collected ratings from our participants after the experiment and used these to confirm the classification of the stimuli.

The words were further separated into four groups: low-frequency regular (LFR), low-frequency exception (LFE), high-frequency regular (HFR), and high-frequency exception (HFE). A word was categorized as high frequency if it had a Kučera and Francis (1967) value greater than 70 per million and as low frequency if its value in the Kuçera and Francis norms was below 30 per million. A word was classified as an exception if its pronunciation was inconsistent with grapheme to phoneme rules (Venezky, 1970). A further criterion concerning consistency was also applied: We excluded from the exception set any word belonging to an orthographic body neighborhood in which the pronunciation of most of the members conflicts with grapheme to phoneme rules (e.g., *bold*). Orthographically strange words (e.g., *yacht*) were also excluded. A word was classified as regular if both (a) its pronunciation was consistent with grapheme to phoneme rules and (b) it belonged to a consistent orthographic neighborhood (e.g., *bank* was acceptable because all _*ank* words rhyme, but *barn* was not because of the existence of *warn*). Finding words with the appropriate characteristics of frequency and regularity that also had imageability ratings proved to be a severe limitation. We therefore collected ratings for a further 58 words from underrepresented classes (e.g., HFEs) from members of the Applied Psychology Unit staff. Eleven of these words were eventually used in the analysis reported below.

Within each of the four word type groups (i.e., LFE, HFE, LFR, and HFR), each high-imageability word was paired with a low-imageability word closely matched on initial phoneme, number of letters, and log frequency. With respect to the matching in terms of initial phoneme, in cases in which it was not possible to obtain an exact match, items were matched with words beginning with phonetically similar sounds (e.g., a word beginning with /p/ could be matched with one beginning with /t/ or /k/). Also, as many of these pairs of words as possible were matched along the same dimensions with a similar pair from each of the other three word type groups. It was possible to find 16 sets of four pairs of items matched in this way. These 64 pairs constituted a thoroughly matched subset within a larger set of 100 word pairs, in which every low-imageability item was matched with a specific high-imageability item of the same word type. In addition, a further eight backup items were included. These were items that could be used to replace certain matched words if the imageability ratings provided by our participants were very discrepant from the preexisting norms. In the event, none of these backup items were used in the analysis.

Participants in the experiment named this complete set of 208 words in a series of four blocks (separated by brief rest periods). The eight different word types were evenly spread throughout the four blocks. Each of the four experimental blocks began with three starter items, which were medium-frequency regular words. Each block, therefore, contained 55 items in all.

Before the data were analyzed, this original set of words was further modified as a result of information collected from our participants during the experiment. First, two words (*boatswain* and *carrel*) were removed because participants showed insufficient knowledge of them (all but 3 participants pronounced *boatswain* incorrectly, and no participants [when asked after the experiment] knew the meaning of the word *carrel*). Second, the imageability ratings collected from our participants after the experiment indicated that 10 words had been wrongly classified in terms of the imageability factor (6 words classified as low-imageability gave rise to unsatisfactorily high ratings, and the converse was true for 4 words classified as high imageability). Deletion of these words necessitated the removal of the words paired with them, and, if they also belonged to the fully matched subset, that subset had to be reduced accordingly. When this was done, there remained 12 eight-way matched sets of words embedded within a larger set of 88 pairs of words. The numbers of items remaining in each category in the whole group were as follows: 18 high- and low-imageability pairs in the LFE group, 25 LFR pairs, 18 HFE pairs, and 27 HFR pairs. There were 56 one-syllable and 40 two-syllable items in the eight-way matched set and 98 one-syllable and 78 two-syllable items in the whole set. Table 1 gives the mean imageability ratings (the average of the preexisting values from norms and those obtained from our participants), Kučera and Francis (1967) frequency, and word length values for the items in the eight-way matched group, with the appropriate values for the whole group also indicated. The words from the eight-way matched set can be found in Appendix A.

In addition to these experimental stimuli, 20 medium-frequency (Kučera & Francis, 1967, range: 30–70 per million) regular consistent words were selected as practice items.

Apparatus. Using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993), we presented the stimuli in the center of a Macintosh PowerBook 170 computer screen. The words were presented in black lowercase print, Geneva 24 point, within a white rectangle (10 cm × 5.5 cm) on a dark gray background. The screen was placed approximately 60 cm from the participant. Naming responses were recorded by using an AKG acoustics C410 microphone and a Sony TC-525 tape recorder. The microphone was also connected to a voice-activated relay interfaced to the computer, which timed response latencies in milliseconds from the appearance of the stimulus to the onset of the participant's response.

Procedure. Participants were tested one at a time in a quiet room. They were given written instructions (on the computer screen) explaining that their task was to name the words aloud as quickly and as accurately as they could. The instructions were followed by a block of 20 practice trials and then the four experimental blocks. The intertrial interval in both practice and experimental blocks was 1,000 ms. Within each trial, participants first saw a fixation point, which remained on the screen for 750 ms. Immediately at the offset of the fixation, the word to be named appeared. As soon as the participant named the word, it disappeared, and the cycle was repeated after the intertrial interval. The words within each block were presented in a different random order for each participant, and the order of presentation of the four experimental blocks was randomly counterbalanced between participants. At the end of each block, the participants could rest for as long as they wished before starting the next block. The experimenter recorded mispronunciations and voice key errors by hand during the experiment, and these were checked with the tape recordings after the experiment.

Imageability rating task. When the participants had completed all four experimental blocks, they were asked to rate all of the experimental word stimuli for imageability on a 7-point scale. They were given a set of instructions to read (that was based on the instructions used by Toglia & Battig, 1978) along with verbal instructions that, in the case of ambiguous words (e.g., *well*), they should rate the most familiar meaning rather than attempting to find some compromise value. We also stressed to participants that any sensory experience counted in determining imageability, not just visual sensations. Using PsychLab (Gum & Bub, 1988), we presented the words one at a time in black, lowercase script, Geneva 24 point, and centered on the PowerBook 170 white screen. The rating scale was simultaneously presented on screen below the target word, with the extremes appropriately labeled as a reminder. Each word remained on screen until the participant indicated his or her response by pressing one of seven keys labeled from 1 to 7 on the keyboard. There was no time limit for response, although participants were asked to work as quickly and as carefully as they could. As soon as each response was made, the next word to be rated was presented. The words were presented in a different random order for each participant, and responses were automatically stored in a text file for analysis.

Results

Separate analyses were carried out on the data for the complete set of 88 pairs of items and for the more precisely controlled eight-way matched subset. Because these produced similar results, for conciseness, only the analyses for the better controlled eight-way matched group are presented here. The mean latencies in response to each word type are shown in Table 2. Analyses of variance (ANOVAs) were performed on latency and error data, using both subject (F_1) and item (F_2) means.

Latencies. Responses to 4% of the words (83 out of a total of 1,920 responses) were excluded from the analysis. Four observations (0.2%) were removed because of voice key errors. A further 18 observations (0.9%) were removed after being classified as outliers based on the following procedure: The response times (RTs) collected from the 20 participants for each word were ranked, and the extreme observations for each word were subjected to the test for outliers described in

Table 1
Characteristics of Stimuli in Experiment 1

	High-frequency exception		Low-frequency exception		High-frequency regular		Low-frequency regular	
Characteristic	HI	LI	HI	LI	HI	LI	HI	LI
Mean frequency[a]								
Matched set	249.70	228.70	7.75	9.25	175.80	179.80	9.33	7.25
Whole set	224.20	272.00	7.40	9.40	151.10	161.60	8.40	7.70
Mean no. of letters								
Matched set	5.00	5.40	5.80	6.00	5.25	5.40	5.25	4.90
Whole set	5.20	5.40	5.90	6.00	5.00	5.20	5.40	5.20
Mean imageability								
Matched set	611.90	310.00	615.30	325.10	595.40	321.00	597.10	342.60
Whole set	599.00	311.00	616.00	326.10	594.00	315.20	591.00	354.20

Note. HI = high imageability; LI = low imageability.
[a]From Kučera and Francis (1967).

Table 2
Response Times (in Milliseconds) for Each Word Type in Experiment 1

Word type	Matched set M	Matched set SD	Whole set M	Whole set SD
High-frequency exception				
HI	521.00	48.00	526.00	25.00
LI	530.00	56.30	530.00	31.50
Low-frequency exception				
HI	552.00	55.70	560.00	49.40
LI	580.00	89.50	593.00	51.20
High-frequency regular				
HI	517.00	40.00	523.00	25.40
LI	536.00	67.80	531.00	28.58
Low-frequency regular				
HI	525.00	53.10	532.00	32.00
LI	539.00	70.80	548.00	35.65

Note. HI = high imageability; LI = low imageability.

Johnson and Leone (1968). Any value with a probability of .01 or less of being a genuine member of the distribution of RTs for that particular word was discarded and replaced by the cutoff value. Both voice key errors and outliers were fairly evenly distributed across the different types of word. Finally, 61 observations (3%) were removed from the RT analysis because a participant made an error in naming the word. Missing values were replaced by the individual participant's mean RT for the appropriate word group.

The three variables included in the ANOVA conducted on the latency data were regularity (regular vs. exception), imageability (high vs. low), and frequency (high vs. low). These variables were treated as within-subject variables in the analysis by subjects and between-items variables in the analysis by items.

Participants were significantly faster in naming high-frequency words (526 ms) than low-frequency words (549 ms): $F_1(1, 19) = 16.83$, $MSE = 1{,}230.98$, $p < .001$; $F_2(1, 88) = 9.49$, $MSE = 1{,}395.58$, $p < .005$. They were also significantly faster to name regular words (529 ms) than exception words (545 ms): $F_1(1, 19) = 21.45$, $MSE = 485.32$, $p < .0002$; $F_2(1, 88) = 4.90$, $MSE = 1{,}395.58$, $p = .05$. High-imageability words yielded significantly shorter RTs (529 ms) than low-imageability words (546 ms): $F_1(1, 19) = 10.30$, $MSE = 1{,}177.43$, $p < .005$; $F_2(1, 88) = 5.43$, $MSE = 1{,}395.58$, $p = .02$. Furthermore, there was a significant interaction between the regularity and frequency variables: $F_1(1, 19) = 29.32$, $MSE = 417.75$, $p < .0001$; $F_2(1, 88) = 5.72$, $MSE = 1{,}395.58$, $p < .02$. The means relevant to this interaction are plotted in Figure 1 and show (a) that the frequency effect was strongest for exception words and (b) that the regularity effect was present only for low-frequency words. This interpretation was confirmed by analysis of simple effects. LFRs were named with significantly shorter average RTs (532 ms) than LFEs (566 ms): $F_1(1, 19) = 54.15$, $MSE = 417.75$, $p < .0001$; $F_2(1, 88) = 10.60$, $MSE = 1{,}395.58$, $p < .002$; whereas responses to HFRs (527 ms) and to HFEs (525 ms) were equally fast. The 40-ms difference between HFEs and LFEs was highly reliable, $F_1(1, 19) = 77.60$, $MSE = 417.75$, $p < .0001$; $F_2(1, 88) = 15$, $MSE = 1{,}395.58$, $p < .0002$; whereas the 5-ms difference between HFRs and LFRs was nonsignificant by both subjects and items.

Although no other effects reached significance, there was a trend toward a significant three-way interaction (Frequency × Regularity × Imageability), but only in the subjects analysis, $F_1(1, 19) = 3.20$, $MSE = 447.01$, $p < .09$. The relationship between these three variables is plotted in Figure 2 and may be summarized as follows: Only the exception words showed a pronounced frequency effect, and there was an interaction between imageability and frequency such that LFEs showed a much larger imageability effect than HFEs. HFRs and LFRs produced roughly similar RTs, with no interaction between imageability and frequency.

Errors. Naming errors were categorized into four types. Visual–phonological word errors were those in which the participant produced a different word that was visually or phonologically similar to the target (e.g., *wrest* → "west"). Visual–phonological nonword errors were those in which the participant produced a nonword that was visually or phonologically similar to the target word (e.g., *suave* → "sove"). Regularization errors were cases in which participants pronounced an exception word as if it were regular (e.g., *dread* → "dreed"). Finally, all other errors (e.g., stuttering) were categorized as "other." Regularization was the most frequent type of error, with 42 errors (69% of the total errors and 78% of all errors to exception words) being of this type. As can be seen from the error rate percentages in Figure 2, very few errors were made in response to any regular word type. The distribution of errors over the four exception word conditions is shown in Table 3, in which we compare regularization errors with all other error types. Many more errors, in particular regularization errors,

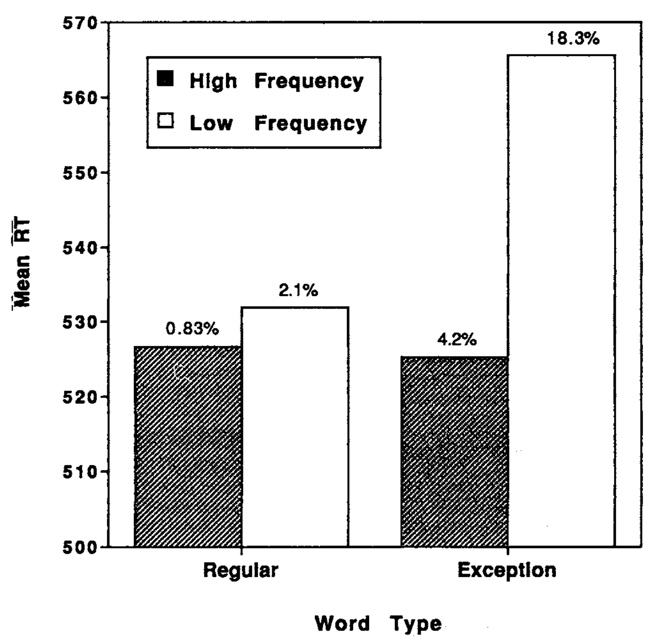

Figure 1. The interaction between regularity and frequency in response times (RTs; in milliseconds) from Experiment 1, with percentage of errors.

occurred in response to low-imageability LFEs than to any other type of word.

The error data for the exception words were further assessed with a two-variable ANOVA, namely frequency (high vs. low) and imageability (high vs. low), in which the dependent variable was the number of regularization errors. Before statistical analysis, error data were square root transformed (Myers, 1979). This analysis showed that participants made fewer regularization errors on high-frequency than low-frequency words, $F_1(1, 19) = 107.91$, $MSE = 0.07$, $p < .0001$; $F_2(1, 44) = 4.76$, $MSE = 0.66$, $p < .04$, and fewer regularization errors to high-imageability than to low-imageability words, F_1 $(1, 19) = 31.16$, $MSE = 0.16$, $p < .0001$, and marginal by items, $F_1(1, 44) = 2.16$, $MSE = 0.66$, $p = 0.15$. Most notably, there was an interaction between these variables, $F_1(1, 19) = 73.90$, $MSE = 0.10$, $p < .0001$, and marginal by items, $F_2(1, 44) = 2.99$, $MSE = 0.66$, $p < .09$. The source of the interaction is the fact that there was a pronounced imageability effect for LFEs (with many more regularization errors to low- than to high-imageability words) but no such effect for HFEs. This interpretation is borne out in both subjects and items analyses of simple effects. There was a significant effect for imageability at low frequency, $F_1(1, 19) = 124.23$, $MSE = 0.10$, $p < .0001$; $F_2(1, 44) = 5.12$, $MSE = 0.66$, $p < .03$, but not at high frequency (nonsignificant by subjects and by items). The simple effects analyses also revealed a significant frequency effect for low-imageability words, $F_1(1, 19) = 147.79$, $MSE = 0.10$, $p < .0001$; $F_2(1, 44) = 7.65$, $MSE = 0.66$, $p < .01$, but not for high-imageability words (nonsignificant by subjects and by items).

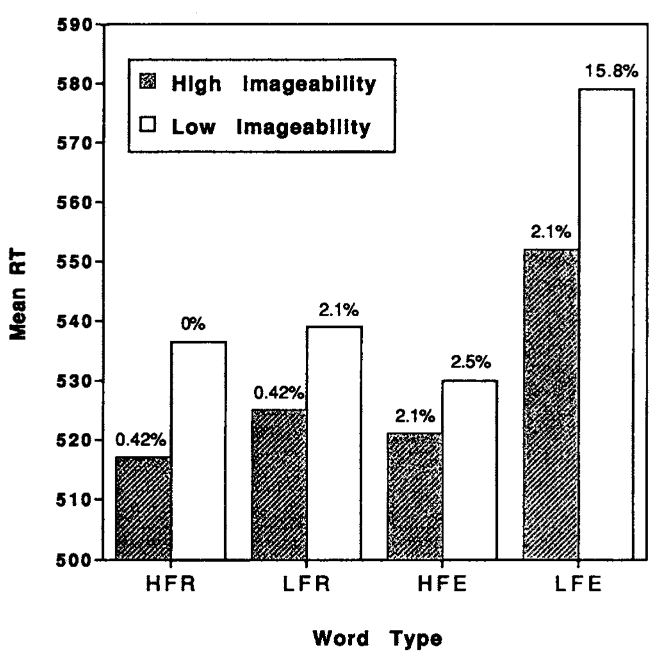

Figure 2. The relationship between regularity, frequency, and imageability in response times (RTs; in milliseconds) from Experiment 1 with percentage of errors. HFR = high-frequency regular; LFR = low-frequency regular; HFE = high-frequency exception; LFE = low-frequency exception.

Table 3
Error Rates (%) in Response to the Various Exception Word Types in Experiment 1

Frequency/imageability	Regularizations	Others
High		
High	1.25	0.80
Low	0.40	2.10
Low		
High	1.25	0.80
Low	14.60	1.25

Discussion

The results of Experiment 1 indicate that normal adult readers are both slower and less accurate in naming abstract low-frequency exception words than imageable low-frequency exception words; this imageability effect was not as strongly evident for any other word type. These results support the hypothesis offered in the introduction that low-frequency exception words should be the items most likely to reveal a role for word meaning in naming. Although this hypothesis was clearly confirmed by the error data, however, it was not convincingly supported by the RT data, in which the critical three-way interaction was only marginal by subjects and nonsignificant by items. One reason for this pattern of results may be that our items analysis lacked power, because it was based on only 12 words per condition. Another possibility is that a small number of specific words in the low-imageability, LFE condition may consistently have produced long latencies. For example, some items in this group, though well matched on other factors, might be orthographically more unusual than those in the other groups. To investigate this possibility, we calculated positional bigram frequencies (which were based on the Solso & Juel, 1980, norms) for each of the items; the means for the word types do indeed reveal lower average PBF values for the low-imageability, low-frequency exceptions (3,927.70) than for items in the other seven groups (ranging from 4,867.80 to 7,571.60). Experiment 2 was designed, in part, to address this concern.

Experiment 2

The chief concern in the second experiment was to increase the sensitivity of the design. This was primarily achieved by improving and by enlarging the set of stimuli. Only low-frequency words were used because it was only with this group of words that we expected to see significant effects, an expectation confirmed by Experiment 1. This meant that matching on the relevant variables only had to be achieved for quartets, rather than octets, of words, allowing both a larger number of matched sets and better matching. In particular, in addition to the factors in Experiment 1, items in Experiment 2 were matched in terms of Solso and Juel's (1980) positional bigram frequencies. Also, the number of participants tested in this experiment was doubled to 40.

Method

Participants. The data were collected from 40 members of the Applied Psychology Unit participant panel who were paid for their

Table 4
Characteristics of Stimuli in Experiments 2 and 3

	Low-frequency exception		Low-frequency regular	
Characteristic	HI	LI	HI	LI
Mean frequency[a]	6.50	5.60	6.10	5.60
Mean no. of letters	5.30	5.80	5.60	5.70
Mean imageability	573.70	352.00	592.50	328.10
Mean positional bigram frequency[b]	4,679.90	4,486.50	4,602.50	4,436.40

Note. HI = high imageability; LI = low imageability.
[a]From Kučera and Francis (1967). [b]From Solso and Juel (1980).

participation. The participants were aged between 22 and 70 years, and about two thirds were women.

Materials and design. Only low-frequency items were used in this experiment. These were selected from the low-frequency words within the pool of items with imageability ratings collected before Experiment 1 (refer to the *Method* section of Experiment 1 for details on how the words were generated and divided into word type groups). As before, high-imageability words were matched with low-imageability items within each word group (now consisting of two groups: LFR and LFE), and then each pair was matched as closely as possible with a pair from the other group. We achieved 16 quartets matched on the same factors as in Experiment 1 (frequency, imageability, number of letters, and initial phoneme, or at least class of phoneme) plus positional bigram frequency. Forty of these items were one-syllable, and the remaining 24 were two-syllable. In addition to the 64 items selected in this way, a further 14 backup items were included, once again in case some of the selected items received inappropriate imageability ratings from our participants.

Participants named these 78 words in two blocks separated by a brief rest period. The four different word types were evenly spread throughout the two blocks. Each of the experimental blocks began with three starter items (medium-frequency regular words) and therefore contained a total of 42 items. As in Experiment 1, after completing the naming experiment, participants were asked to rate each of the items for imageability. Two words received imageability ratings inconsistent with their classification before the experiment: Both *ghoul* and *clam* had been classified as highly imageable but received low ratings from our participants and were therefore replaced by their backup items, *ghost* and *cliff,* respectively. Table 4 gives the mean imageability ratings (i.e., the average of the preexisting values from norms and those obtained from our participants), frequency, positional bigram frequency, and word length values for the items used in the analysis of this experiment. The words themselves can be found in Appendix B.

In addition to these experimental stimuli, 26 medium-frequency (Kučera & Francis, 1967, range: 30–70 per million) regular consistent words were used as practice items.

Apparatus. Using PsyScope running on an Apple Macintosh IIci computer, we presented the stimuli in the center of a Macintosh Portrait black and white display. The words were presented in black, lowercase print, Geneva 24 point, on a white screen placed approximately 60 cm from the participant. Naming responses were recorded by using a Yamaha MH100 headset microphone connected to the voice key port of a CMU button box (see Cohen et al., 1993, for details), which was interfaced to the computer allowing it to time response latencies in milliseconds. The microphone was also connected to a Sony TC-525 tape recorder that recorded the participants' responses.

Procedure. The procedure in this experiment was identical to that in the first, except that participants were only required to name one practice and two experimental blocks (rather than four experimental blocks). As before, after the naming experiment, participants were asked to rate the items for imageability.

Results

Latencies. Responses to 8% of the original word-naming data (191 out of a total of 2,560 responses) were excluded from the analysis: 26 observations (1%) were removed because of voice key errors, and 8 observations (0.3%) were classified as outliers and replaced according to the same technique described for Experiment 1. Both voice key errors and outliers were fairly evenly distributed across the different types of word. Finally, 157 observations (6%) were removed from the RT analysis because a participant made an error in naming the word. Missing values (i.e., voice key errors and naming errors) were replaced by the individual's mean RT for the appropriate word group.

Because we had such a wide age range of participants (22–70 years), a preliminary ANOVA was carried out to see whether age interacted with either or both of the two important variables: imageability and regularity. Three variables were included in this ANOVA: regularity (regular vs. exception), imageability (high vs. low), and age (young vs. old). To produce the two age groups, we divided the participants into two groups of 20, the age range of the young group being 22–46 years (M 36.85), and that of the old group being 57–70 ($M = 62.30$). Regularity and imageability were treated as within-subject variables in the analysis by subjects and between-items variables in the analysis by items. Age was treated as a between-subjects variable in the analysis by subjects and a within-item variable in the analysis by items. The older participants were slower (568 ms) in naming words than the younger participants (528 ms): $F_1(1, 38) = 3.29$, $MSE = 19{,}263.31$, $p < .08$; $F_2(1, 60) = 213.46$, $MSE = 175.61$, $p < .0001$. Of importance, however, age did not interact with either of the other two variables; therefore, for the main analyses, data for the older and younger participants were combined. Two variables were therefore included in the main ANOVA conducted on the latency data: imageability and regularity. These variables were treated as within-subject variables in the analysis by subjects and between-items variables in the analysis by items.

Participants were significantly quicker in naming regular words (538 ms) than exception words (558 ms): $F_1(1, 39) = 35.04$, $MSE = 426.97$, $p < .0001$; $F_2(1, 60) = 4.26$, $MSE = 1{,}404.83$, $p < .05$. They were also significantly faster in naming high-imageability words (538 ms) than low-imageability words (560 ms): $F_1(1, 39) = 53.45$, $MSE = 447.70$, $p < .0001$; $F_2(1, 60) = 6.80$, $MSE = 1{,}404.83$, $p = .01$. Furthermore, there was a reliable interaction between regularity and imageability, $F_1(1, 39) = 20.62$, $MSE = 610.23$, $p < .0001$, and marginal by items, $F_2(1, 60) = 3.58$, $MSE = 1{,}404.80$, $p = .06$. The means relevant to this interaction are plotted in Figure 3 and show (a) that the regularity effect is present only for low-imageability words and (b) that the imageability effect is present only for exception words. This interpretation was confirmed by analysis of simple effects. For low-imageability words, RTs to regular words (542 ms) were significantly faster than those to exception words

(579 ms): $F_1(1, 39) = 45.05$, $MSE = 610.23$, $p < .0001$; $F_2(1, 60) = 7.82$, $MSE = 1{,}404.83$, $p < .01$; by contrast, for high-imageability items, regular words and exception words produced virtually identical naming times (535 and 537 ms, respectively). Similarly, there was a highly reliable difference between high- and low-imageability exception words, $F_1(1, 39) = 58.35$, $MSE = 610.23$, $p < .0001$; $F_2(1, 60) = 10.12$, $MSE = 1{,}404.83$, $p < .002$, whereas the corresponding comparison for regular words was nonsignificant by both subjects and items.

Errors. As in Experiment 1, naming errors were categorized into regularizations, visual–phonological word errors, visual–phonological nonword errors, and others. Regularizations were the most frequent type of error, with 103 incorrect pronunciations (66% of the total errors) being of this type. There were 19 visual–phonological word errors (12%), 16 visual–phonological nonword errors (10%), and 19 errors classified as others (12%). The distribution of errors over the four word conditions is shown in Table 5, in which we compare regularization errors with all other error types.

An ANOVA was carried out on the error data, with imageability and regularity both treated as within-subject variables by subjects and between-items variables by items. In the initial error analysis, the dependent variable was the total number of errors (i.e., all error types were combined). Before statistical analysis, error data were square root transformed. The ANOVA indicated (a) that there were significantly more errors with exception words (142) than regular words (15), $F_1(1, 39) = 116.29$, $MSE = 0.26$, $p < .0001$; $F_2(1, 60) = 30.71$, $MSE = 0.79$, $p < .0001$, (b) that there were significantly more errors with abstract words (126) than imageable words (31), $F_1(1, 39) = 77.63$, $MSE = 0.15$, $p < .0001$; $F_2(1, 60) = 12.25$, $MSE = 0.79$, $p < .001$, and (c) that there was a significant interaction between regularity and imageability: $F_1(1, 39) = 73.91$, $MSE = 0.20$, $p < .0001$; $F_2(1, 60) = 10.62$, $MSE = 0.79$, $p < .002$.

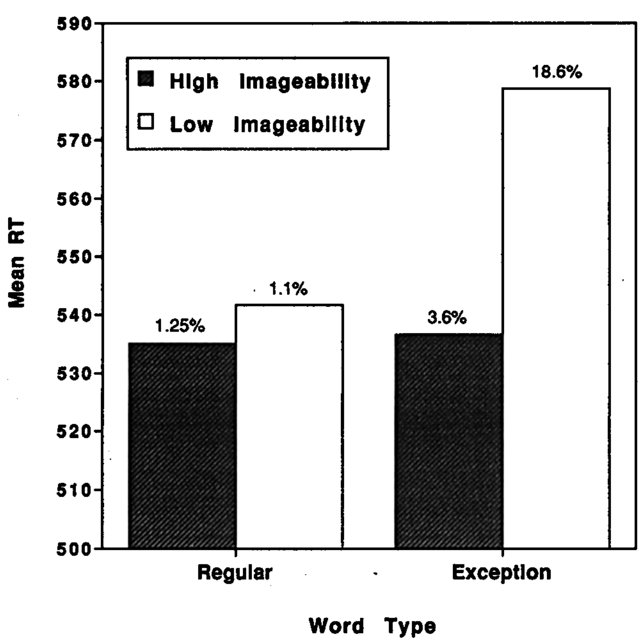

Figure 3. The interaction between regularity and imageability in response times (RTs; in milliseconds) from Experiment 2 with percentage of errors.

Table 5
Error Rates (%) in Response to the Various Word Types in Experiment 2

Word type/imageability	Regularizations	Others
Regular		
High	0	1.25
Low	0	1.10
Exception		
High	2.00	1.60
Low	14.10	4.50

Analysis of simple effects indicated that more errors were made to exception words than to regular words for both high-imageability items, significant by subjects, $F_1(1, 39) = 7.25$, $MSE = 0.20$, $p < .01$, marginal by items, $F_2(1, 60) = 2.61$, $MSE = 0.79$, $p = .1$, and low-imageability items, $F_1(1, 39) = 220.54$, $MSE = 0.20$, $p < .0001$; $F_2(1, 60) = 38.73$, $MSE = 0.79$, $p < .0001$. In addition, although there was no imageability effect for regular words (eight errors to imageable and seven errors to abstract regular words), there was a highly reliable difference for exception words between imageable (23 errors) and abstract (119 errors) items: $F_1(1, 39) = 133.43$, $MSE = 0.20$, $p < .0001$; $F_2(1, 60) = 22.84$, $MSE = 0.79$, $p < .0001$. This result closely corresponds to the latency data.

Referring back to Table 5, it appears that the significant difference in overall error rates between high- and low-imageability exceptions is attributable primarily to a difference in the number of regularization errors. This interpretation was tested by performing a second ANOVA on the exception word error data, including the variables imageability and error type variables. Error type was a within-subject variable both by subjects and by items and contained two levels: regularizations versus all other errors. As well as main effects for both imageability and error type, there was also a significant interaction between the error type and imageability variables: $F_1(1, 39) = 22.12$, $MSE = 0.29$, $p < .0001$; $F_2(1, 30) = 4.37$, $MSE = 1.04$, $p < .05$. First, analysis of simple effects showed that there were significantly more regularization errors (90 errors) than all other error types (29) in response to abstract words, $F_1(1, 39) = 45.54$, $MSE = 0.29$, $p < .0001$; $F_2(1, 30) = 8.51$, $MSE = 1.04$, $p < .01$, whereas there were similar numbers of regularizations (13) and all other errors (10) in response to imageable words (nonsignificant by subjects and by items). Second, there was a highly reliable effect of imageability for regularization errors: $F_1(1, 39) = 86.65$, $MSE = 0.29$, $p < .0001$; $F_2(1, 60) = 16.11$, $MSE = 1.000$, $< .0001$; this effect was much weaker for other error types, significant only by subjects: $F_1(1, 39) = 7.06$, $MSE = 0.29$, $p < .0001$; $F_2 < 1$. This analysis, therefore, supported the claim that the greater error rate for low- over high-imageability exceptions is primarily due to regularization errors.

Familiarity. The preceding analyses demonstrated a significant interaction between regularity and imageability, observable in both the latency and error data. This evidence lends strong support to the view that aspects of word meaning are activated in the process of word naming and that meaning has the largest impact on the naming of low-frequency exception words. One potential obstacle to this interpretation is that, although stimuli in the high- and low-imageability groups were well matched on a number of relevant variables including frequency, they were not explicitly matched for familiarity. If the efficiency with which phonological representations are activated by orthography or are translated into speech is modulated by word familiarity, and if our high-imageability exception words were more familiar than the low-imageability exceptions, then the significant differences in RT and error rate might be attributable to familiarity, potentially a phonological variable, rather than imageability, a semantic one.

Because of the lack of appropriate norms for the sort of stimuli we required, it was impossible to control for familiarity beforehand. To address this issue after the fact, however, we asked 30 participants (drawn from the same population as those in the naming experiment) to provide paper-and-pencil familiarity ratings for our stimuli. Participants were asked to give each word a rating on a 7-point inclusive scale ranging from 1 (*low familiarity*) to 7 (*high familiarity*). The items were presented to each participant in one of several different random orders, except that the first six words for every participant were starter items of widely differing familiarity levels, to allow the participant to obtain a sense of the familiarity range before they began rating the actual stimuli. These ratings showed that there was no major confound between imageability and familiarity in this stimulus set, with all word types receiving similar average values (low-imageability regulars = 3.40, high-imageability regulars = 3.80, low-imageability exceptions = 3.60, and high-imageability exceptions = 3.90). The fact that participants were slower and less accurate in naming abstract than imageable exceptions seems very unlikely to find an explanation in terms of familiarity, given the small difference in mean familiarity ratings for these two word sets.

Familiarity ratings give an estimate of how often participants come in contact with whole-word patterns and consequently might help predict naming times. In addition, the familiarity of *subword* patterns might also be an important determinant of naming latency, which is why we matched our items in terms of positional bigram frequency. An alternative measure of the familiarity of subword patterns is the N metric, which was proposed by Coltheart, Davelaar, Jonasson, and Besner (1977). The N value for a word is the number of different words that can be produced by changing just one letter of the original word, preserving other letter positions, and reflects the size of the word's orthographic neighborhood. Because of N's potential effect on naming latencies, we thought it prudent to check how well our items were matched on this variable, albeit in a post hoc manner. We used the MRC psycholinguistic database (Coltheart, 1981) to calculate the Ns for our items, which were based on all the items in the database. As with the familiarity ratings, there was no major confound between imageability and N in this stimulus set, with all word types receiving roughly similar average values (low-imageability regulars = 3.25, high-imageability regulars = 4.80, low-imageability exceptions = 4.44, and high-imageability exceptions = 6.60). A two-variable ANOVA (regularity and imageability were variables), with N as the dependent variable, showed that these values were not significantly different from each other.

For a more complete assessment of the potential impact of familiarity and N on our results, however, we performed a by-items analysis of covariance (ANCOVA) on the RT data; we used familiarity and N (log transformed) as covariates and regularity and imageability as the other two variables. With the influence of familiarity and N partialled out in this way, the critical interaction between regularity and imageability remained significant, indeed at a slightly higher level of reliability, $F_1(1, 58) = 5.39$, $MSE = 1{,}160.16$, $p < .03$. This seems to rule out the hypothesis that our effects might be due to a confounding of imageability with familiarity of our stimuli, either at the word or subword level.

The above analyses showed that the high- and low-imageability exception groups were well matched in terms of familiarity and that this variable did not influence the results. However, it might still be argued that the observed imageability effect was due to one or two atypical words in the low-imageability, low-frequency exception group giving rise to particularly long latencies and large numbers of errors. To demonstrate that this was not the case, the individual RTs and the number of regularization errors for each of the low-frequency exception words from this experiment are given in Appendix C. The spread of both RTs and regularization errors in the low-imageability condition would argue against an explanation in terms of one or two aberrant items affecting the results for the whole group. Indeed, only 4 out of the 16 low-imageability items failed to produce some regularization errors, and all but 2 of those that produced errors gave rise to 3 or more. Contrast this with the fact that only 6 out of the 16 high-imageability items produced this error type, and only 1 of these produced more than two errors.

Discussion

The results of Experiment 2, in which we used only lower frequency words, provide additional support for the hypothesis outlined in the introduction: In both latency and accuracy of word naming, normal adult readers showed a reliable interaction between regularity and imageability. This interaction reflects approximately equivalent performance (both accuracy and RT) on imageable regular words, abstract regular words, and imageable exception words, with significantly slower and less accurate performance only on abstract exception words. This pattern of results suggests (a) that representations of word meaning are activated in the course of orth-to-phon translation and (b) that semantic representations for imageable words make a useful contribution to this computation in the case of low-frequency exception words.

Experiment 3

In Experiment 3, we hypothesized that a manipulation that reduced the impact of word meaning on the naming process should significantly increase the error rate in response to high-imageability, low-frequency exceptions but have a much smaller effect on the errors produced in response to low-imageability exceptions. Assuming that any semantic contribution to phonological activation builds up gradually, one way to reduce this contribution might be to force participants to name words more rapidly than they are naturally inclined to do. A speeded naming paradigm was therefore used in Experiment 3, with the hypothesis that, in comparison with normal naming, we would observe a reduced effect of word imageability.

Method

Participants. The data were collected from 40 members of the Applied Psychology Unit participant panel who were paid for their participation. The age range of the participants in this experiment was similar to that in Experiment 2 (21–70 years), with an equal number of men and women.

Materials and design. The identical experimental stimuli to those in Experiment 2 were used. Because the task was more demanding and required a higher level of attention, however, the words were presented in a series of four short blocks, rather than two long blocks, to provide more rest periods. Each of these four blocks contained five starter items, followed by the 16 targets and 2 fillers. Before being presented with the four experimental blocks, participants were given a practice block consisting of 25 medium-frequency regular words.

Apparatus. The apparatus was identical to that used in Experiment 2.

Procedure. Participants were tested one at a time in a quiet room. They were given on-screen instructions telling them that their task was to name the items as quickly as they could. The importance of speed was stressed by telling the participants that, at the speed required, it was possible that they might make some mistakes, but they were not to worry about that. Furthermore, 250 ms after the onset of the target word, a tone was presented, and 100 ms later the word disappeared. Participants were told that they should try to say each word before it disappeared, and that if they named the word very quickly, they might occasionally prevent the tone from sounding. When it was clear that the participants understood the instructions, they were given a block of 26 practice trials and then the four experimental blocks. The intertrial interval in both practice and experimental blocks was 750 ms. Within each trial, participants first saw a fixation point, which remained on the screen for 1,000 ms. Immediately at the offset of the fixation, the word to be named appeared. This word remained on the screen for 350 ms, with a tone presented 250 ms after the onset of the word. If the participant named the word with a latency of 350 ms or less, the word disappeared as soon as the naming response began; on those rare occasions when a participant produced a naming latency of 250 ms or less, the tone was not sounded. This cycle was repeated after the intertrial interval. Participants' responses and response latencies were recorded. The words within each block were presented in a different random order for each participant, and the order of presentation of the four experimental blocks was counterbalanced between subjects. At the end of each block, the participants could rest for as long as they wished before starting the next block. The experimenter recorded mispronunciations and voice key errors by hand during the experiment, and these were checked by using the tape recordings after the experiment.

Results

All correct responses (except for outliers in the RT distribution; listed below) were included in the analysis. That is, although there was in a sense a response deadline (because participants were asked to try to start naming each word before it disappeared), the purpose of the deadline was merely to encourage fast responses, not to exclude slower ones (see Vitkovitch & Humphreys, 1991, for a similar treatment of speeded responses in picture-naming experiments). ANOVAs were performed on the latency and error data by using subject and item means as units of analysis. The speeded manipulation was clearly effective, in the sense that naming latencies were on average approximately 100 ms shorter in this experiment than in the previous two. The individual RTs and numbers of regularization errors for each of the low-frequency exception words from this experiment are given in Appendix C.

Latencies. Of the original word-naming latencies, 357 out of a total of 2,560 (14%) were excluded from the analysis. Twelve observations (0.5%) were removed because of voice key errors; 9 observations (0.3%) were classified as outliers (by using the same technique as in Experiments 1 and 2) and were removed and replaced with the cutoff values. Both voice key errors and outliers were fairly evenly distributed across the different types of word. Finally, 336 observations (13%) were removed because the participants made an error in naming the word; this is approximately twice as many errors as in Experiment 2, when the identical words were named without the instructions and cues designed to speed naming. Missing values (i.e., voice key errors and naming errors) were replaced by the individual's mean RT for the appropriate word group.

As a striking difference in results from Experiments 1–2, under speeded conditions there was no main effect of regularity on latency of correct responses, with essentially identical mean RTs to regular (428 ms) and exception (429 ms) words (though see the error analysis reported below). Participants were significantly slower to name low-imageability words (434 ms) than high-imageability words (424 ms): $F_1(1, 39) = 15.95$, $MSE = 259.43, p < .0005$; $F_2(1, 60) = 3.44, MSE = 483.96, p = .07$. Furthermore, the regularity and imageability variables interacted with each other, significant by subjects, $F_1(1, 39) = 14.02$, $MSE = 201.55$, $p < .001$, and marginal by items, $F_2(1, 60) = 2.35$, $p = .13$, and this interaction is shown in Figure 4. As in Experiment 2, the imageability effect was strongly present only for exception words. This interpretation was confirmed by analysis of simple effects of both the subject and item means. The naming times for high-imageability exception words (420 ms) were reliably faster than those for low-imageability exception words (439 ms): $F_1(1, 39) = 34.25$, $MSE = 201.55, p < .0001$; $F_2(1, 60) = 5.73$, $MSE = 483.96$, $p < .02$, whereas there was no difference between high- and low-imageability regular words (427 and 429 ms, respectively).

One notable feature of these results is the lack of the regularity effect. If we assume, first, that speeding the naming response reduces the impact of semantics and, second, that semantic information assists naming for exception words more than for regular words, then we might expect the regularity effect to become larger under speeded naming. However, bearing in mind the high error rate induced by the instructions, the RT results of Experiment 3 should be trusted with caution. The error rate for Experiment 3 was twice as high as that for Experiments 1–2, with a very high proportion of the errors occurring in response to exceptions. Exception

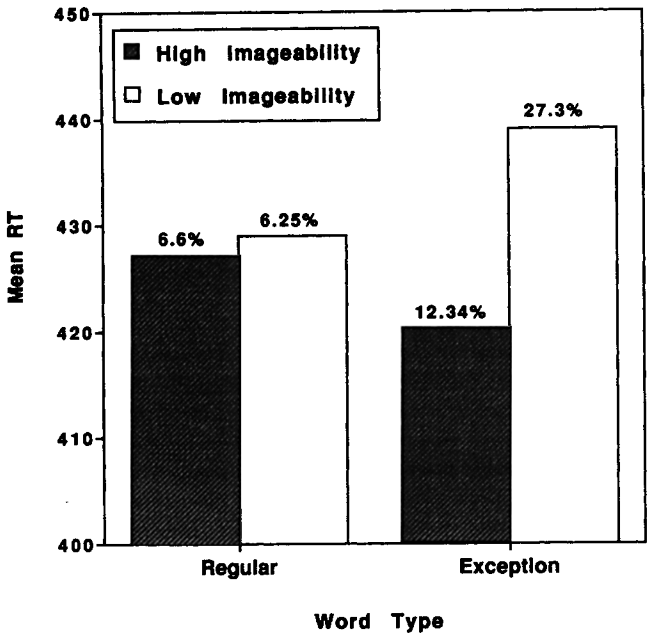

Figure 4. The interaction between regularity and imageability in the in response times (RTs; in milliseconds) from Experiment 3 with percentage of errors.

words normally take longer to name than regular words: If participants responded to the speeded manipulation by trading speed for accuracy on the exception words, this might well produce exception RTs close to those for regular words, but at the cost of many more errors.

Errors. The error pattern produced in this experiment was of particular interest because the instructions stressed speed over accuracy. Errors were categorized into the same four groups as in the previous experiments. Regularizations were the most frequent type of error, with 115 errors (34% of the total errors) being of this variety. There were 85 visual–phonological word errors (25%), 63 visual–phonological nonword errors (19%), and 73 errors classified as others (22%).

For the following analyses, error data from the present speeded-naming experiment and the previous normal-naming experiment (with identical stimulus words) were combined, giving a sample of 80 participants in all. All error data were square root transformed. Two separate analyses were carried out: one on all errors combined and one just on regularization errors.

ANOVA on all errors. Three variables were included in the ANOVA conducted on total errors: regularity (regular vs. exception), imageability (high vs. low), and naming condition (normal vs. speeded). In the by-subjects analysis, regularity and imageability were within-subject variables, and naming condition was a between-subjects variable. In the by-items analysis, naming condition was a within-item variable, and regularity and imageability were between-items variables. Participants mispronounced exception words more often than regular words, $F_1(1, 78) = 200.22$, $MSE = 0.32$, $p < .0001$; $F_2(1, 60) = 34.35$, $MSE = 1.28$, $p = .0001$, made significantly more errors with abstract words than imageable words, $F_1(1, 78) = 61.83$, $MSE = 0.23$, $p < .0001$; $F_2(1, 60) = 12.51$, $MSE = 1.28$, $p < .001$, and, of course, made significantly more mistakes under conditions of speeded naming than normal naming, $F_1(1, 78) = 31.87$, $MSE = 0.72$, $p < .0001$; $F_2(1, 60) = 66.66$, $MSE = 0.44$, $p < .0001$. In this combined analysis, as in the analysis specifically for Experiment 2, there was also a significant interaction between regularity and imageability: $F_1(1, 78) = 76.773$, $MSE = 0.21$, $p < .0001$; $F_2(1, 60) = 10.86$, $MSE = 1.28$, $p = .005$. There were no other significant effects.

The interaction between imageability and regularity was similar to that obtained in Experiment 2 (refer to Table 4). Analysis of simple effects indicated that more errors were made to exception words than regular words at both levels of imageability: high, $F_1(1, 78) = 38.88$, $MSE = 0.21$, $p < .0001$; $F_2(1, 60) = 4.34$, $MSE = 1.28$, $p < .05$, and low, $F_1(1, 78) = 346.97$, $MSE = 0.21$, $p < .0001$; $F_2(1, 60) = 38.50$, $MSE = 1.28$, $p < .0001$. The significant interaction is explained by the fact that high- and low-imageability regular words produced about the same number of errors (nonsignificant by subjects and by items), whereas there was a highly reliable difference between numbers of errors produced in response to high- and low-imageability exception words: $F_1(1, 78) = 146.15$, $MSE = 0.21$, $p < .0001$; $F_2(1, 60) = 20.81$, $MSE = 1.28$, $p < .0001$.

This first error analysis provided no support for our hypothesis that speeded naming would have the most harmful effect on high-imageability exception words. Had this been the case, a three-way interaction between regularity, imageability, and naming condition should have emerged. There is, however, good reason to question the sensitivity of this analysis including all error types. Under speeded-naming conditions, many of the errors were trivial ones (e.g., false starts and stutters, clearly provoked by the pressure to start speaking quickly), and this extra noise in the data may have masked the effect of interest. A more informative analysis would focus on the relative numbers of regularization errors produced under normal and speeded naming. In Experiments 1–2, the accuracy difference between high- and low-imageability exception words was due mainly to a greater number of regularization errors on the low-imageability words. If, as we have proposed, it is semantic information for the high-imageability exceptions that helps to prevent regularization errors under normal naming; and further, if speeded naming reduces the impact of such a semantic contribution; then we would predict that, under speeded naming, imageable exceptions should engender a greater increase in the number of regularizations than abstract exceptions.

ANOVA on regularization errors. Only responses to exception words were included in this two-way ANOVA involving imageability (high vs. low) and naming condition (normal vs. speeded). In the by-subjects analysis, imageability was a within-subject variable, and naming condition was a between-subjects variable; in the by-items analysis, vice versa.

Participants produced significantly more regularization errors with low-imageability words than with high-imageability words: $F_1(1, 78) = 85.27$, $MSE = 0.28$, $p < .0001$; $F_2(1, 60) = 7.82$, $MSE = 2.69$, $p < .01$. As predicted, there was also a significant interaction between naming condition and imageability, $F_1(1, 78) = 9.02$, $MSE = 0.28$, $p < .005$, marginal by items, $F_2(1, 60) = 3.81$, $MSE = 0.31$, $p = .06$. The interaction between imageability and naming condition is illustrated in Figure 5. In moving from normal to speeded naming, the number of regularization errors in response to *high-imageability* exception words more than doubled (from 13 to 36 errors), and analysis of simple effects showed that this increase was statistically significant, $F_1(1, 78) = 19.41, MSE = 0.28, p < .0001$, marginal by items, $F_2(1, 60) = 3.92, MSE = 0.31, p < .06$. By contrast, the number of regularization errors in response to low-imageability words actually decreased slightly from normal (90 errors) to speeded naming (79 errors), though the simple-effects analysis showed this decrease to be nonsignificant by subjects and by items.

In conclusion, forcing participants to speed their word naming increased the number of regularization errors to high-imageability exceptions, whereas speeding had no significant effect on the number of regularization errors to low-imageability exceptions. On the assumption that speeding the naming process reduces the impact of meaning on orth-to-phon translation, this finding supports the claim that semantic information assists the correct naming of high-imageability (low-frequency) exception words.

General Discussion

The results of Experiments 1–3 can briefly be summarized as follows. In Experiment 1, participants produced many more errors in response to low- than to high-imageability words only for low-frequency exceptions. The augmented error rate in this word group consisted almost entirely of regularization errors. The latency data, though not as strong, were consistent with this pattern. In Experiment 2, we used only low-frequency words but with a larger N than Experiment 1 for both items and subjects and also better balancing of the stimulus items, which produced a significant interaction between regularity and imageability for RT as well as for accuracy; that is, for low-frequency exception but not for regular words, words with imageable referents were named more rapidly and more accurately than words with abstract meanings. Once again, the difference in error rate was almost entirely attributable to more regularization errors on abstract than on imageable exception words. Experiment 3, in which participants were encouraged to speed their word-naming responses by an average of almost 100 ms as compared with Experiment 2, produced a significant increase in regularization errors only to low-frequency exception words with high-imageability referents.

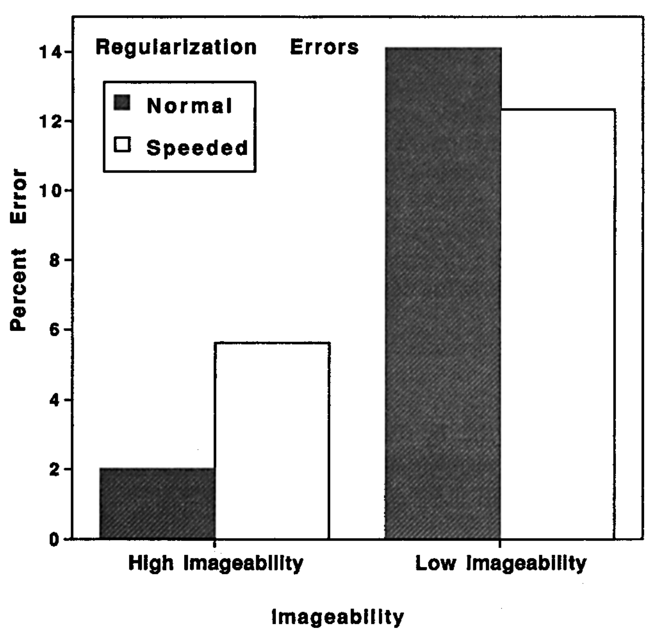

Figure 5. The interaction between imageability and naming condition in the regularization errors from Experiments 2 and 3.

Imageability is admittedly a complex variable with only subjective rather than objective values and is moreover correlated with other variables known or suspected to influence the efficiency of word processing; we therefore cannot promise that our pattern of results is explicable only in terms of the semantic difference between high- and low-imageability words. We have, however, made every effort to balance the words on other dimensions, either by stimulus matching or by an ANCOVA. In particular, in Experiment 2, we were able to rule out word familiarity, a variable that plausibly has its impact at the stage of phonological rather than semantic representations, as a likely alternative explanation of the advantage for high-imageability words.

Our account of this pattern of results is facilitated by reference to Figure 6, the framework for word pronunciation and comprehension proposed by Seidenberg and McClelland (1989). The orth-to-phon computational networks implemented by Seidenberg and McClelland and Plaut and McClelland (1993) demonstrate that this single procedure is capable of learning to produce accurate pronunciations for essentially all monosyllabic words, whether these embody typical or atypical spelling-to-sound correspondences. This demonstration is not, however, equivalent to a claim that no other representations or processes are involved or required when human readers translate orth-to-phon. The framework in Figure 6 permits semantic representations of words to affect the computation of their pronunciations in one or both of two ways. The most obvious is that because readers may gradually learn to access meaning directly from orthographic representations and have already—early in life—learned to activate phonology from meaning, then pronunciation of written words may be a cooperative venture between the two procedures: orth-to-phon and orth-to-meaning-to-phon. The second is that because this sort of model assumes graded activations of distributed representations rather than all-or-none retrieval of whole-word representations, then interaction between phonological and semantic representations may occur during the period between initial phonological activation (on the basis of orth-to-phon) and the point at which the phonological representation becomes sufficiently strong and stable to support a response. In other words, even if the principal activation of phonology is directly from orthography, semantic representations might still have a detectable impact on word naming because of cascaded processing from phonology to meaning and back again.

One more concept is required for our discussion and that is some notion of how semantic representations of imageable and abstract words might differ. The field of cognitive psychology

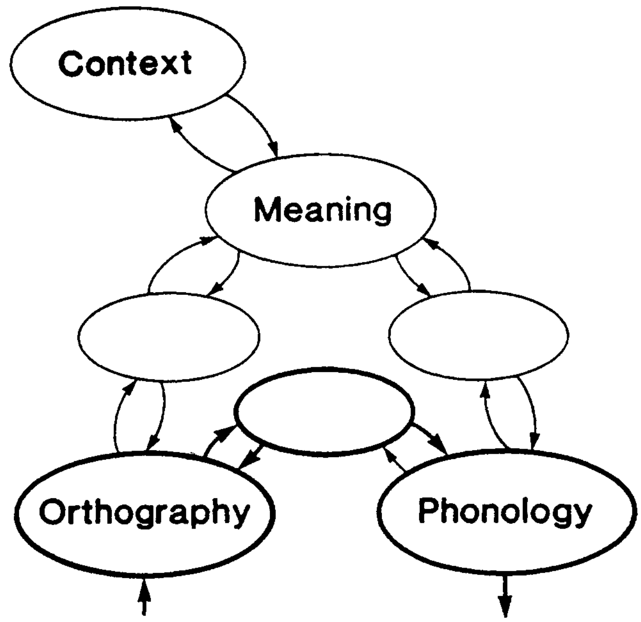

Figure 6. Framework for word pronunciation and comprehension. Adapted from "A Distributed, Developmental Model of Word Recognition and Naming" by M. S. Seidenberg and J. L. McClelland, 1989, *Psychological Review, 96,* p. 22. Copyright 1989 by the American Psychological Association.

has rather little to say about the nature of semantic representation, but this particular distinction is often cast in some form of "richer" semantic representations for high- than for low-imageability words (Allport, 1985; see also Breedin, Saffran, & Coslett, 1994, for an excellent review of this topic). One form of operationalizing the notion of richness is simply by the number of semantic features, with imageable words having more defining features than abstract words (Jones, 1985; Plaut & Shallice, 1993). Another explains the difference in terms of relationship to context, with imageable words bringing their own rich base of contextual information (Schwanenflugel, Akin, & Luh, 1992), whereas abstract words depend more for their meanings on the linguistic context in which they occur (Saffran, Bogyo, Schwartz, & Marin, 1980). It is even possible that, because high-imageability words link to sensory and motor properties of nonlinguistic semantic memory (Paivio, 1991), their meanings are represented in a qualitatively and neuroanatomically distinct way from those for low-imageability words. This latter conjecture is supported by several detailed case studies of patients with progressive but selective impairments of semantic memory, demonstrating a significant advantage for comprehension of abstract as opposed to imageable words (see Breedin et al., 1994, and Warrington, 1975, for case studies and Patterson & Hodges, 1995, for discussion). Whatever the nature of the difference, it seems that, for a person with intact semantic memory, high-imageability words have richer, more stable, more readily computable, and contextually independent meanings than low-imageability words.

Now, armed with both the framework in Figure 6 and some notion of how semantic representations might differ as a function of imageability, we return to an interpretation of our word-naming results. On presentation of a word to be pronounced, the orth-to-phon computation begins to activate elements of phonology. For words with regular, consistent spelling–sound correspondences, an unambiguous phonological representation should quickly be achieved; this will be true even for a low-frequency consistent word because the computation of its phonology will be supported by weights on connections established by other words (friendly neighbors) sharing both spelling and sound patterns. For a word with inconsistent correspondences, a single, clear phonological representation may also emerge rapidly, so long as it is a high-frequency word: Although the computation of a word's pronunciation is clearly affected by the characteristics of its neighbors, the most important factor is the frequency with which that individual word has been processed (Plaut & McClelland, 1993; Seidenberg & McClelland, 1989). Only low-frequency exception words should fail to produce rapid, noise-free phonological representations because these are boosted neither by their own nor by their neighbors' impact on weight settings. If activation of correct phonology rises to criterion more slowly for these items, then there will be time for semantic information to enter and influence the course of phonological processing. Furthermore, given some ambiguity in the computed phonological representation—for example, if there is simultaneous activation of phonological elements corresponding to both the correct and regularized vowel pronunciation of a word like *pint* —then the semantic information might help to resolve this ambiguity. Whether the semantic influence is on speed, accuracy, or both, the fact that words with imageable referents have better semantic representations should mean that low-frequency exception words benefit from semantic support primarily if they are high- rather than low-imageability words.

It is perhaps worth noting that the idea of computing several competing elements of a pronunciation more or less simultaneously is consistent with observations of the behavior of both humans and computational models. In naming experiments with time pressure, normal readers sometimes start to give a regularized pronunciation of a low-frequency exception word and then correct it. The orth-to-phon computational model of Plaut and McClelland (1993) sometimes activates two competing phonological elements (typically for the vowel), especially for low-frequency inconsistent words. Although neither of these observations demonstrates that the ambiguity is (for the human reader) or would be (for the computer model) resolved by information about the word's meaning, some support for precisely this hypothesis comes from the reading performance of patients with semantic deficits. Virtually all of these patients make frequent regularization errors in reading (Patterson & Hodges, 1992), and several have been observed either (a) to offer both pronunciations of an exception word without a preference (e.g., *deaf* → "it's deef or deaf") or (b) to change their minds, starting equally often with a regularization followed by the correct pronunciation (*deaf* → "deef, no deaf") or with a correct response followed by a regularization (*deaf* → "deaf, no deef," Behrmann & Patterson, 1993; Funnell, 1993). Normal readers never offer two pronunciations without a preference and never change correct responses to incorrect ones; we suggest that this is because communication with meaning reinforces the correct phonological representation for the intact readers but not for the patients.

The ideas just discussed provide an account of the results of Experiments 1 and 2. Participants were slower and made significantly more regularization errors in pronouncing low-frequency, irregular abstract words than low-frequency, irregular imageable words because although the orth-to-phon computation is relatively inefficient for both of these two word groups, the high-imageability words benefit more from activation of semantic representations. Regular words and high-frequency exception words did not show an imageability effect because their orth-to-phon translation is too efficient and self-sufficient. An account of the results from Experiment 3 with speeded naming requires a further notion: When participants are not allowed to wait until phonological representations settle into a clear, stable pattern before initiating their naming responses, then any beneficial effects from activation of semantic information will be diminished. This should, and did, have the effect of increasing regularization errors on low-frequency exception words with high-imageability meanings.

We are not, of course, suggesting that only low-frequency, high-imageability exception words automatically access semantic representations—all words surely do. The claim is merely (a) that phonological processing of regular words and high-frequency words is too efficient to allow much impact of the word's meaning and (b) that, in a single-word naming task, the phonological representation for a low-imageability word will not markedly be assisted by interaction with its semantic representation, perhaps because low-imageability words have rather context-dependent meanings.

At a general level, the results presented in this article are congruent with current notions of interactive activation between the representations of orthography, phonology, and meaning that are basic to reading (Seidenberg & McClelland, 1989; Van Orden & Goldinger, 1994). At a more specific level, although many studies concerned with such interactions have established the importance of phonology in computing the meaning of a printed word, this is one of very few demonstrations of how meaning may affect computation of phonology.

References

Allport, D. A. (1985). Distributed memory, modular systems and dysphasia. In S. K. Newman & R. Epstein (Eds.), *Current perspectives in dysphasia* (pp. 30–60). Edinburgh, Scotland: Churchill Livingstone.

Behrmann, M., & Patterson, K. (1993, July). *A patient for the decade: Frequency and neighbourhood effects in a surface alexic patient.* Paper presented to a joint meeting of the Canadian Society for Brain, Behaviour and Cognitive Science and the Experimental Psychology Society, Toronto, Ontario, Canada.

Breedin, S. D., Saffran, E. M., & Coslett, H. (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology, 11,* 617–660.

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

Carello, C., Turvey, M. T., & Lukatela, G. (1992). Can theories of word recognition remain stubbornly nonphonological? In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning: Advances in psychology* (Vol. 94, pp. 211–226). Amsterdam: North-Holland.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers, 25,* 257–271.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A,* 497–505.

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. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Coltheart, V., Patterson, K., & Leahy, J. (1994). When a ROWS is a ROSE: Phonological effects in written word comprehension. *Quarterly Journal of Experimental Psychology, 47A,* 917–955.

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

de Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 824–845.

Funnell, E. (1987). Morphological errors in acquired dyslexia: A case of mistaken identity. *Quarterly Journal of Experimental Psychology, 39A,* 497–539.

Funnell, E. (1993, March). *Capturing the task of reading.* Paper presented to the British Neuropsychological Society, London, England.

Gilhooley, K. J., & Logie, R. H. (1980). Age of acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1944 words. *Behavior Research, Methods, & Instrumentation, 12,* 395–427.

Gum, T., & Bub, D. (1988). *PsychLab software.* Montreal, Quebec, Canada: PsychLab.

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120,* 358–394.

Johnson, L. J., & Leone, F. C. (1968). *Statistics and experimental design in engineering and the physical sciences, Vol. 1.* New York: Wiley.

Jones, G. V. (1985). Deep dyslexia, imageability, and ease of predication. *Brain and Language, 24,* 1–19.

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

Lesch, M. F., & Pollatsek, A. (1993). Automatic access of semantic information by phonological codes in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 285–294.

Myers, J. L. (1979). *Fundamentals of experimental design.* Boston: Allyn & Bacon.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45,* 255–287.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery and meaningfulness values for 925 words. *Journal of Experimental Psychology Monograph Supplement, 76* (3, Pt 2).

Patterson, K., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia, 30,* 1025–1040.

Patterson, K., & Hodges, J. R. (1995). Disorders of semantic memory. In A. D. Baddeley, B. A. Wilson, & F. Watts (Eds.), *Handbook of memory disorders* (pp. 167–186). New York: Wiley.

Plaut, D. C., & McClelland, J. L. (1993). Generalization with componential attractors: Word and nonword reading in an attractor network. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 824–829). Hillsdale, NJ: Erlbaum.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10,* 377–500.

Saffran, E. M., Bogyo, L. C., Schwartz, M. F., & Marin, O. S. M. (1980). Does deep dyslexia reflect right-hemisphere reading? In M. Coltheart, K. Patterson & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 381–406). London: Routledge & Kegan Paul.

Schwanenflugel, P. J., Akin, C., & Luh, W. (1992). Context availability and the recall of abstract and concrete words. *Memory & Cognition, 20,* 96–104.

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

Solso, R. L., & Juel, C. L. (1980). Positional frequency and versatility of bigrams for two through nine letter English words. *Behavior Research, Methods, & Instrumentation, 12,* 297–343.

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

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition, 15,* 181–198.

Van Orden, G. C., & Goldinger, S. D. (1994). The interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1269–1291.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 371–386.

Van Orden, G., Pennington, B., & Stone, G. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97,* 488–522.

Venezky, R. L. (1970). *The structure of English orthography.* The Hague, The Netherlands: Mouton.

Vitkovitch, M., & Humphreys, G. W. (1991). Perseverant responding in speeded naming of pictures: It's in the links. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 664–680.

Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology, 27,* 635–657.

Watkins, M. J. (1984). Models as toothbrushes. *The Behavioural and Brain Sciences, 7,* 86.

Wydell, T. N., Patterson, K. E., & Humphreys, G. W. (1993). Phonologically mediated access to meaning for Kanji: Is a rows still a rose in Japanese Kanji? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 491–514.

Appendix A
Eight-Way Matched Stimuli Used in Experiment 1

Low-frequency regular		Low-frequency exception		High-frequency regular		High-frequency exception	
HI	LI	HI	LI	HI	LI	HI	LI
blade	blunt	breast	blown	bill	best	blood	break
blister	blunder	boulder	broader	doctor	district	building	greatest
ditch	deed	dove	debt	dark	deal	dead	done
dump	dodge	dough	dread	black	bring	death	does
mirror	mercy	monarch	mischief	market	manner	money	measure
mustard	mister	monkey	nowhere	morning	method	mother	nothing
pepper	parry	treasure	twofold	picture	training	people	toward
pickle	pious	croquet	toughness	teacher	trying	police	trouble
sack	sane	sword	suave	saw	stay	foot	flow
scout	scorn	swamp	scarce	space	sense	steak	sure
weed	wisp	wealth	wont	wife	west	war	want
wick	yore	worm	wrath	write	wrong	watch	worth

Note. HI = high imageability; LI = low imageability.

(Appendixes continue)

Appendix B
Stimuli Used in Experiments 2 and 3

Low-frequency exception		Low-frequency regular	
LI	HI	LI	HI
broader	boulder	blessing	banner
cache	climb	cleft	cliff
caste	comb	clause	corpse
chasm	croquet	custom	coffin
dose	dove	deed	duck
guise	ghost	gait	groin
mischief	meadow	madness	mattress
scarce	soot	scribe	straw
sleight	sword	scorn	spike
soften	shovel	stanza	sandal
stingy	fatigue	figment	scarlet
suave	swamp	fraud	snail
toughness	treasure	traitor	trumpet
trough	pear	truce	trout
warn	worm	whence	wreck
wrath	wand	wrest	witch

Note. LI = low imageability; HI = high-imageability.

Appendix C
Response Times (RTs; in Milliseconds) and Number of Regularization Errors for High- and Low-Imageability Exception Words in Experiments 2 and 3

Low imageability			High imageability			Low imageability			High imageability		
Word	RT	No. of errors	Word	RT	No. of errors	Word	RT	No. of errors	Word	RT	No. of errors
Experiment 2						Experiment 3					
broader	569	0	boulder	532	0	broader	450	1	boulder	427	0
cache	602	3	climb	509	0	cache	445	5	climb	408	0
caste	589	4	comb	523	0	caste	423	5	comb	414	0
chasm	634	9	croquet	582	1	chasm	458	11	croquet	449	2
dose	539	8	dove	511	2	dose	401	4	dove	403	1
guise	544	0	ghost	502	0	guise	432	0	ghost	407	0
mischief	518	11	meadow	494	0	mischief	419	9	meadow	397	0
scarce	597	2	soot	572	2	scarce	470	1	soot	435	11
sleight	627	16	sword	573	1	sleight	448	13	sword	430	11
soften	564	3	shovel	556	0	soften	448	0	shovel	436	0
stingy	605	17	fatigue	589	0	stingy	472	15	fatigue	445	0
suave	635	5	swamp	578	0	suave	464	5	swamp	435	0
toughness	550	0	treasure	534	0	toughness	431	0	treasure	428	0
trough	607	2	pear	518	0	trough	433	1	pear	411	6
warn	505	0	worm	498	6	warn	399	0	worm	406	4
wrath	576	10	wand	514	1	wrath	433	9	wand	398	1

Received May 18, 1994
Revision received October 24, 1994
Accepted November 14, 1994 ■