When Words Compete: Levels of Processing in Perception of Spoken Words

Michael S. Vitevitch and Paul A. Luce

Psychological Science 1998 9: 325
DOI: 10.1111/1467-9280.00064

Research Report

WHEN WORDS COMPETE: Levels of Processing in Perception of Spoken Words

Michael S. Vitevitch and Paul A. Luce

Language Perception Laboratory and Center for Cognitive Science, Department of Psychology, University at Buffalo

Abstract—*Current theories of spoken-word recognition posit two levels of representation and process: lexical and sublexical. By manipulating probabilistic phonotactics and similarity-neighborhood density, we attempted to determine if these two levels of representation have dissociable effects on processing. Whereas probabilistic phonotactics have been associated with facilitatory effects on recognition, increases in similarity-neighborhood density typically result in inhibitory effects on recognition arising from lexical competition. Our results demonstrated that when the lexical level is invoked using real words, competitive effects of neighborhood density are observed. However, when strong lexical effects are removed by the use of nonsense word stimuli, facilitatory effects of phonotactics emerge. These results are consistent with a two-level framework of process and representation embodied in certain current models of spoken-word recognition.*

Without doubt, understanding spoken words is one of the primary perceptual and cognitive activities in which people engage each day. For many years, this routine but complex mental process has been the focus of much research and theory. In particular, researchers have attempted to determine how the listener maps a continuously varying acoustic signal onto one of thousands of representations of words in memory with such remarkable speed and accuracy.

Current theories of how people perceive spoken words propose that the recognition process consists of at least two levels of representation and process: lexical and sublexical (see McClelland & Elman, 1986). The *lexical* level includes representations that correspond to words and is often characterized as consisting of competitive processes among these lexical representations (Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris, 1994; see also Colombo, 1986). Although there is less agreement regarding the precise nature of representations at the *sublexical* level (see Pisoni & Luce, 1987), many theories of spoken-word recognition also assume that independent representational entities corresponding to components of spoken words are activated during perception (McClelland & Elman, 1986; Vitevitch, 1997).

Recent research (see Vitevitch, 1997; Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997) has suggested that the manner in which the recognition system deals with probabilistic phonotactic information may provide insights into the processes associated with lexical and sublexical units of representation. Probabilistic phonotactics refers to the frequencies of segments and sequences of segments in syllables and words (Trask, 1996), and a number of studies have demonstrated that listeners are sensitive to probabilistic differences among phonotactic patterns (Treiman, Kessler, Knewasser, Tincoff, & Bowman, 1996; Vitevitch et al., 1997).

Michael Vitevitch is now at the Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington.

Address correspondence to Paul A. Luce, Department of Psychology, University at Buffalo, Buffalo, NY 14260; e-mail: paul@deuro.fss.buffalo.edu.

For example, we (Vitevitch et al., 1997) found that participants' preference ratings of spoken nonsense words vary as a function of phonotactic probability. Nonwords composed of common segments and sequences of segments were rated as being better possible words than nonwords composed of less common segments and sequences. In addition, we used the auditory naming task to demonstrate that processing—and not simply subjective judgment—is affected by the probabilistic phonotactic characteristics of spoken stimuli. In this task, participants hear a spoken stimulus and must repeat it as quickly and accurately as possible. We found that reaction times mirrored the pattern of results from the rating experiment: Nonwords composed of high-probability phonotactic patterns were repeated faster than nonwords composed of less common patterns, suggesting that phonotactic probabilities not only are represented in memory, but also have consequences for processing.

Of particular interest to modeling the process of spoken-word recognition is the locus of probabilistic phonotactic effects. These effects may arise at the lexical or sublexical level. On the one hand, segmental and sequential probabilities may be represented by the activation levels of representations of phonetic segments. On the other hand, effects of probabilistic phonotactics (even for nonsense words) may arise from interactions among lexical items themselves, implicating no direct role of sublexical units (McClelland & Elman, 1986; Taraban & McClelland, 1987). Thus, identifying the locus of previously demonstrated effects of phonotactics may provide crucial insights into the nature of the architecture of the word recognition system (see Norris, 1994).

An apparently dissimilar phenomenon in the literature on spoken-word recognition may have direct relevance to the issues of phonotactic processing and, ultimately, lexical and sublexical processes and representations. The number and frequency of similar words activated in memory during perception have predictable effects on recognition speed and accuracy. Using a number of different experimental paradigms (including the naming task), Luce and Pisoni (1998) demonstrated that words that sound similar to many other words (i.e., words in dense similarity neighborhoods) are recognized less accurately and more slowly than words that sound similar to few other words (i.e., words in sparse neighborhoods).

A problem arises when one attempts to reconcile the previously demonstrated effects of phonotactics and similarity neighborhoods: Words in high-density neighborhoods tend to have high-probability phonotactic patterns, whereas words in low-density neighborhoods are typically composed of less frequent segments and sequences of segments (see Landauer & Streeter, 1973). High-probability phonotactic patterns are high in probability precisely because there are many words sharing the component segments. Simply put, there is a positive correlation between phonotactic probability and neighborhood density. However, in one study (Vitevitch et al., 1997), stimuli with high-probability phonotactic patterns were responded to more quickly than stimuli with low-probability phonotactic patterns, whereas in another study (Luce & Pisoni, 1998), stimuli in high-density neighborhoods were responded to more slowly than stimuli in low-density neighborhoods.

One possible explanation for this discrepancy presents itself: Perhaps effects of probabilistic phonotactics and neighborhood density arise at two different levels of processing. Facilitatory effects of probabilistic phonotactics may reflect differences among activation levels of sublexical units, whereas effects of similarity neighborhoods may arise from competition among lexical representations. To test this possibility, we directly manipulated similarity neighborhoods and probabilistic phonotactics of two sets of stimuli—words and nonsense words—in a naming task. We expected that if the effects of probabilistic phonotactics and similarity neighborhoods arise from different levels of processing, we would see differences between participants' responses to word stimuli, which have lexical representations in memory, and nonword stimuli, which do not have representations that can be directly activated in the mental lexicon. In particular, we predicted facilitatory effects of phonotactics for nonwords, but competitive effects of lexical neighborhoods for words.

METHODS

Participants

Thirty native English-speaking adult subjects were recruited from the University at Buffalo community. All subjects reported no history of a speech or hearing disorder and were paid $5 for their participation.

Materials

Two sets of consonant-vowel-consonant (CVC) stimuli were selected: 240 nonsense words and 150 real words. (The numbers of nonwords and words differ because they were chosen in part to provide comparisons with bisyllabic stimuli used in other experiments; see Vitevitch, 1997.) Half of each set consisted of stimuli with high-probability phonotactic patterns and high neighborhood density. The other half of both sets of stimuli consisted of low-density, low-probability phonotactic patterns. Thus, there were four sets of experimental stimuli: high-density, high-probability words and nonwords, and low-density, low-probability words and nonwords. No syllable was used more than once.

Phonotactic probability

Two measures were used to determine phonotactic probability: positional segment frequency (i.e., how often a particular segment occurs in a given position in a word) and biphone frequency (i.e., the segment-to-segment co-occurrence probability, which itself is almost perfectly correlated with segmental transitional probability; see Gaygen, 1997). These metrics were based on frequency-weighted counts of words in an electronic version of Webster's (1967) pocket dictionary. This electronic dictionary, which was created at the Massachusetts Institute of Technology in 1982, contains approximately 20,000 computer-readable phonemic transcriptions.

Nonwords and words that were classified as high-probability patterns consisted of segments with high positional segment probabilities. For example, the consonant /k/ is relatively frequent in initial position, the vowel /i/ is relatively frequent in the medial position, and the consonant /k/ is relatively frequent in the final position. Therefore, /kik/ ("keek") was classified as a high-probability nonword. In addition, a high-probability phonotactic pattern had high biphone probabilities; that is, the co-occurrence of the initial consonant and vowel had a high probability, as did the co-occurrence of the vowel and final consonant (e.g., /b/ followed by /æ/ and /æ/ followed by /p/ in the nonsense word /bæp/).

Nonwords and words that were classified as low-probability patterns consisted of segments with low positional segment probabilities and low biphone probabilities. Despite being relatively rare, none of the patterns were phonotactically illegal in English. Indeed, the transitions in the nonword stimuli occur in real English words.

Similarity neighborhoods

Frequency-weighted similarity neighborhoods were computed by comparing a given phonemic transcription (constituting the stimulus) to all other transcriptions in the Webster's lexicon (see Luce & Pisoni, 1998). A neighbor was defined as any transcription that could be converted to the transcription of the stimulus word or nonword by one phoneme substitution, deletion, or addition in any position.

The log frequencies of the neighbors were then summed for each word and nonword, rendering frequency-weighted neighborhood-density measures. For the low-density nonwords, the mean log-frequency-weighted neighborhood-density value was 13.46, whereas for the high-density nonwords, the mean log-frequency-weighted neighborhood-density value was 44.61, $F(1, 238) = 588.75$, $p < .0001$. For the low-density words, the mean log-frequency-weighted neighborhood-density value was 40.00, whereas for the high-density words, the mean log-frequency-weighted neighborhood-density value was 56.43, $F(1, 148) = 42.45$, $p < .0001$.

Isolation points

Previous research (Marslen-Wilson & Tyler, 1980; see also Luce, 1986) has shown that the point at which a word diverges from all other words in the lexicon affects recognition. Therefore, we determined isolation points using the computerized lexicon. The isolation points for the high-probability, high-density words and the low-probability, low-density words were equivalent $F(1, 148) = 1.59$ $p = .20$. The mean isolation point for the high-probability, high-density words was 2.98 phonemes, and the mean isolation point for the low-probability, low-density words was 2.93 phonemes. That is, the CVC words all diverged at approximately the third phoneme. Because of the manner in which they were constructed, all nonwords had the same isolation point at the final segment.

Frequency

Log frequency (Kučera & Francis, 1967) was also equivalent for the two word conditions, $F(1, 148) < 1$ (for high-probability, high-density words, mean frequency = 2.68; for low-probability, low-density words, mean frequency = 2.59).

Initial segments

For a naming task, one needs to ensure that the potential differential sensitivity of the voice key to various initial segments of the stimuli does not confound the reaction times. Our previous work (Vitevitch et al., 1997), using nonword stimuli with the same initial segments as those of the nonword stimuli in the present experiment, demonstrated that the voice key employed in this experiment does not, in fact, play a confounding role in estimating phonotactic effects on processing time. Nonetheless, we balanced the word and nonword stimuli according to two classes of initial segments that could potentially induce minor differences in the response of the voice key. These classes were abrupt and nonabrupt onsets. (These two classes roughly correspond to ± continuant, Chomsky & Halle, 1968, or to continuous-discontinuous, Jakobson, Fant, & Halle, 1951.) The number of segments falling into these categories did not differ significantly across probability-density conditions for words ($\chi^2[1, N = 150] = 1.28, p = .26$) or nonwords ($\chi^2[1, N = 240] = 0.11, p = .74$).

Stimulus preparation

The word and nonword stimuli were recorded by a trained phonetician. All the stimuli were spoken one at a time in a list. Stimulus durations were equivalent for the two nonword conditions, $F(1, 238) = 2.54, p > .10$, as well as for the two word conditions, $F(1, 148) < 1$. All stimuli were low-pass filtered at 4.8 kHz and digitized at a sampling rate of 10 kHz using a 12-bit analog-to-digital converter. All words were edited into individual files and stored on computer disk.

Procedure

Participants were tested individually. Each participant was seated in a booth equipped with a computer terminal and a pair of headphones. A boom microphone was positioned immediately in front of the participant's lips. The microphone was connected to a voice key interfaced to a PDP 11/34 computer. The voice key registered a response as soon as the participant began speaking. Presentation of stimuli and response collection were controlled by the computer.

A typical trial proceeded as follows: A prompt appeared on the computer screen, and one of the spoken stimulus items was presented at a comfortable listening level. The participant then repeated the item as quickly and as accurately as possible into the microphone.

Each participant received a list containing either all of the words or all of the nonwords. Lexicality (word vs. nonword) was blocked by participant to maximize the probability that participants would primarily process the stimulus items at the lexical (for the word stimuli) or sublexical (for the nonword stimuli) level, if differential processing at these levels is indeed possible.

Reaction time was measured by the computer from the onset of the stimulus to the onset of the participant's verbal response. All responses were recorded on audiotape for accuracy analysis. Accuracy was assessed by listening to the participants' responses and comparing them with a written transcription of the stimuli. A response was scored as correct if there was an identical match on all segments of the stimulus.

Each participant received 10 practice trials. These trials were used to familiarize the participants with the task and were not included in the final data analysis. Because of the ease of the task, 10 trials were sufficient to familiarize the participants with the requirements of the experiment.

RESULTS

The mean reaction time for each probability-density condition is shown separately for words and nonwords in Figure 1. Separate 2 (lexicality) × 2 (probability-density) analyses of variance were performed on reaction times and accuracy scores.

For the reaction times, no main effect of probability-density was obtained, $F < 1$. However, the main effect of lexicality was significant, $F_1(1, 28) = 17.76$, $p < .001$, and $F_2(1, 386) = 447.04$, $p < .0001$, indicating that real words were repeated faster than nonwords. A significant interaction of lexicality and probability-density was also obtained, $F_1(1, 28) = 12.31$, $p < .01$, and $F_2(1, 386) = 6.90$, $p < .01$.

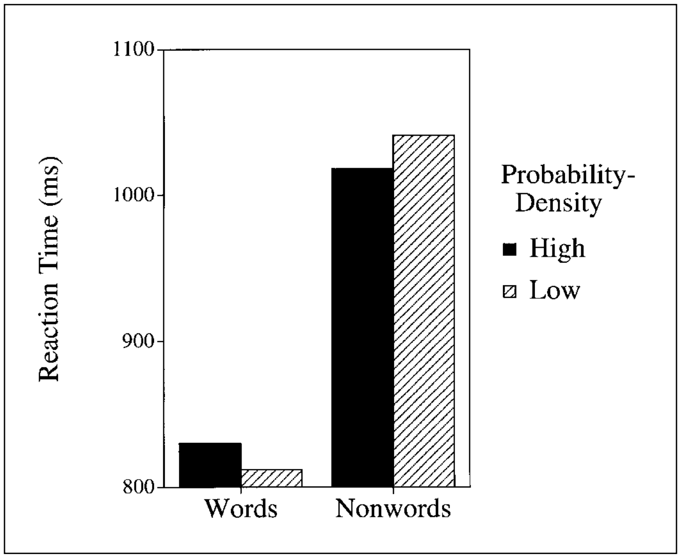

Fig. 1. Reaction times in milliseconds for the words and nonwords for each probability-density condition.

Additional analyses based on the significant interaction were performed to assess the effects of probability-density on the words and the nonwords separately. In the case of real words, low-probability, low-density words were repeated faster than high-probability, high-density words, $F_1(1, 14) = 16.40$, $p < .001$, and $F_2(1, 148) = 3.89$, $p < .05$. For the nonwords, however, high-probability, high-density stimuli were repeated faster than low-probability, low-density stimuli, $F_1(1, 14) = 4.56$, $p < .05$, and $F_2(1, 238) = 3.80$, $p < .05$. Although the magnitude of the effects may appear somewhat small, it is well within the range of effect sizes typically observed in research on spoken-word recognition. Moreover, these effects are consistent across both subjects and items.

No significant effects were obtained for accuracy, $F < 1$. The average percentage correct was 88% for each condition, with the exception of the nonwords in the low-probability, low-density condition, which had an average of 87%.

GENERAL DISCUSSION

The results of this experiment demonstrate that the simultaneous manipulation of probabilistic phonotactics and similarity-neighborhood structure has opposite effects depending on the lexical status of the stimulus: Monosyllabic words show effects of lexical neighborhood competition. Words occurring in dense similarity neighborhoods were responded to more slowly than those in sparse neighborhoods, despite the fact that the stimuli occurring in dense neighborhoods were composed of high-probability segments and sequences of segments. Monosyllabic nonwords, in contrast, were responded to more quickly when they consisted of high-probability phonotactic patterns, despite the fact that they occurred in high-density neighborhoods.

Why does this pattern of results emerge? The crucial variable that mediates these opposing effects is lexicality. Words presumably have lexical nodes or units in memory; nonwords do not. Models of spoken-word recognition such as Trace (McClelland & Elman, 1986), Shortlist (Norris, 1994), and the neighborhood activation model (NAM; Luce & Pisoni, 1998; see also Auer & Luce, 1998) propose that lexical representations compete with or inhibit one another. Thus, words occurring in dense similarity neighborhoods succumb to competition among similar-sounding words activated in memory, resulting in slower processing. Apparently, any benefit these high-density words accrue from having high-probability phonotactic patterns is overshadowed by effects of lexical competition.

Because nonwords do not make direct contact with single lexical units, and thus do not initiate—at least early in processing—large-scale lexical competition, effects of segmental and sequential probabilities emerge for these stimuli. That is, in the absence of strong lexical-competition effects associated with word stimuli, higher activation levels of sublexical units (associated with higher phonotactic probabilities) afford an advantage to high-probability nonwords. Note that this account does not presume that lexical competition is entirely absent for nonwords, nor that facilitatory effects of phonotactics are inoperative for words. Instead, we propose that lexical competition dominates for words, whereas effects of phonotactics are the primary determinant of processing times for nonwords.

It is uncertain whether the effects of phonotactics observed in the present experiment are due to differences among initial activation levels of the segments themselves or are due to changes in segmental activation resulting from lexical feedback (McClelland & Elman, 1986). Nevertheless, the present results support a two-level framework, with effects of phonotactics varying as a function of level of processing. At present, it appears unlikely that a single level of representation and process could produce opposite effects of phonotactics and neighborhood structure. Thus, these results are broadly consistent with models of spoken-word recognition that posit both lexical and sublexical levels of processing.

In addition, our results provide further support for the now widespread assumption in many models (e.g., Trace, Shortlist, NAM) that lexical representations compete—in one way or another—in the recognition process. Clearly, models that fail to incorporate mechanisms of lexical competition, such as the cohort model (Marslen-Wilson & Tyler, 1980), and models that fail to specify a sublexical level of representation at which effects of phonotactics may operate, such as NAM, are inadequate (although a new version of NAM, dubbed PARSYN, that has recently been proposed incorporates a segmental level of representation; see Auer & Luce, 1998). At present, models such as Trace and Shortlist appear to embody the requisite architecture for accounting for opposite effects of probability and density as a function of lexicality, although whether these models as currently instantiated can actually simulate the differential effects observed in our experiment is unclear. Nonetheless, our results provide strong constraints on the nature of the basic architecture of an adequate model of spoken-word recognition.

The finding that effects of phonotactic probability are manifested in opposite ways at the lexical and sublexical levels may provide a coherent framework for interpreting seemingly disparate results in the literature. For example, Newman, Sawusch, and Luce (1996) demonstrated facilitatory effects of increases in neighborhood density on phoneme identification in nonword stimuli. The use of nonword stimuli in the Newman et al. study—as well as the task requirement that participants base their responses on a segmental level of representation—would encourage the type of facilitatory effect found for high-density, high-probability nonword stimuli in the present experiment. In another set of studies, Luce and Pisoni (1998; also replicated in Vitevitch, 1997), using a lexical decision task in which participants were required to decide whether a stimulus was a word or nonword, found that spoken nonwords residing in high-density neighborhoods were responded to more slowly than nonwords in sparse neighborhoods. Although this result apparently contradicts the present finding for nonwords in the naming task, it is easily accounted for under the assumption that phonotactic and neighborhood effects have different loci. In particular, nonword decisions in the lexical decision task (in contrast to auditory naming) require the discrimination of nonwords from words, thus invoking the lexical level of representation and producing similarity-neighborhood effects associated with lexical competition. (It should also be noted that our results from auditory naming are primarily due to perceptual and not production processes. Levelt and Wheeldon, 1994, have demonstrated that pattern probability affects production in a manner opposite to the effects observed in the present experiment.)

In short, the hypothesis that there are two levels of representation with dissociable and diverse effects on processing underscores the complexity of the recognition process: In predicting processing speed for spoken words, one must simultaneously consider the level of representation that dominates the response (Cutler & Norris, 1979; Foss & Blank, 1980), the nature of the task used to interrogate the recognition process, and the phonotactic probability and similarity-neighborhood structure of the stimuli under scrutiny.

In summary, our results have demonstrated that both probabilistic phonotactic information and similarity-neighborhood structure have demonstrable, albeit opposing, effects on recognition of spoken stimuli. The finding that two distinct levels of processing are revealed by the differential effects of probability and density supports a general class of models that includes such current connectionist approaches as Trace and Shortlist. Our findings also provide further support for the now widespread hypothesis that lexical competition plays a major role in the recognition process. Finally, the present results add to a growing body of literature on the effects of phonotactics in the perception of spoken words and demonstrate that effects of phonotactics may manifest themselves in disparate ways depending on the level of processing that dominates recognition.

Acknowledgments—This research was supported in part by Research Grants R01 DC 0265801 and T32 DC00036 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. We would like to thank Jan Charles-Luce, James R. Sawusch, Gayle Beck, and Elaine Hull for their advice and assistance on various aspects of this project.

REFERENCES

Auer, E.T., & Luce, P.A. (1998). *PARSYN: A processing model of neighborhood activation and phonotactics in spoken word recognition.* Unpublished manuscript, University at Buffalo, Buffalo, NY.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper and Row.

Colombo, L. (1986). Activation and inhibition with orthographically similar words. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 226–234.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W.E. Cooper & E.C.T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113–134). Hillsdale, NJ: Erlbaum.

Foss, D.J., & Blank, M.A. (1980). Identifying the speech codes. *Cognitive Psychology, 12,* 1–31.

Gaygen, D.E. (1997). *The effects of probabilistic phonotactics on the segmentation of continuous speech.* Unpublished doctoral dissertation, University at Buffalo, Buffalo, NY.

Jakobson, R., Fant, G., & Halle, M. (1951). *Preliminaries to speech analysis.* Cambridge, MA: MIT Press.

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

Landauer, T.K., & Streeter, L.A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12,* 119–131.

Levelt, W.J.M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition, 50,* 239–269.

Luce, P.A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics, 39,* 155–158.

Luce, P.A., & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19,* 1–36.

Marslen-Wilson, W.D., & Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition, 8,* 1–71.

McClelland, J.L., & Elman, J.L. (1986). The Trace model of speech perception. *Cognitive Psychology, 18,* 1–86.

Newman, R.S., Sawusch, J.R., & Luce, P.A. (1996). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 873–889.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189–234.

Pisoni, D.B., & Luce, P.A. (1987). Acoustic-phonetic representations in word recognition. *Cognition, 25,* 21–52.

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

Trask, R.L. (1996). *A dictionary of phonetics and phonology.* London: Routledge.

Treiman, R., Kessler, B., Knewasser, S., Tincoff, R., & Bowman, M. (1996, May). *English speakers' sensitivity to phonotactic patterns.* Paper presented at the Fifth Conference on Laboratory Phonology, Chicago.

Vitevitch, M.S. (1997). *Phonotactics and spoken word recognition.* Unpublished doctoral dissertation, University at Buffalo, Buffalo, NY.

Vitevitch, M.S., Luce, P.A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40,* 47–62.

Webster's Seventh Collegiate Dictionary. (1967). Los Angeles: Library Reproduction Service.

(RECEIVED 9/10/97; ACCEPTED 12/12/97)