The Neighborhood Characteristics of Malapropisms

Michael S. Vitevitch
Language Perception Laboratory
Department of Psychology
and Center for Cognitive Science
University at Buffalo
Buffalo, NY 14260
mikev@deuro.fss.buffalo.edu

Fay and Cutler (1977) examined several linguistic factors of malapropisms. They defined a malapropism as a real word that erroneously intrudes on an intended, or target, word. The target and error words are not semantically related but share a close relationship in their pronunciations.

Although Fay and Cutler (1977) found many interesting results regarding the stress pattern, grammatical category, number of syllables, and several other characteristics that malapropisms shared with the intended word, their work provided little information that could predict which words might be likely to be malapropisms. The current analysis was undertaken to examine other characteristics of the intended words that might be useful for predicting which words might be likely to be malapropisms.

Evidence suggests that word frequency affects both word *recognition* and word *production* (Dell, 1990; Stemberger and MacWhinney, 1986) in demonstrable ways. Do other factors that influence speech perception also influence speech production? More specifically, do the characteristics of phonologically similar neighborhoods (Luce, Pisoni, and Goldinger, 1990) --word frequency, neighborhood frequency (the frequency of the words in a neighborhood), and neighborhood density (the number of words in memory that are similar to a target word)-- influence word production? To examine this possibility, the current analysis of a malapropism error corpus was conducted.

138 words from Fay and Cutler (1977) were found in an on-line version of the 20,000 word Webster's Pocket Dictionary. The neighborhood characteristics for these 138 words were then calculated. The computations were performed on the phonetic transcriptions contained within the computerized lexicon.

Word frequency was assessed by using the log-frequency of each word. A word was considered to be a low frequency word if the value of its frequency was below the value of log 10. A word with a value above log 10 was considered to be a high frequency word. Neighborhood frequency was assessed in a similar fashion. Neighborhood density was assessed by finding the median value of the data. Values above the median were classified as having dense neighborhoods. Values below the median were classified as having sparse neighborhoods.

The results show that high frequency words from dense neighborhoods tended to be very prevalent in errorful productions. Also, low frequency words from sparse neighborhoods tended to be very prevalent in errorful productions. Additionally, more target words tended to "slip" to words that were of a higher frequency than of a lower frequency compared to the frequency of the target word.

The current results may best be accounted for with a slight modification to an activation based model of speech production (see Dell, 1986, 1988, 1990). A high frequency word that is normally highly activated may be relatively less activated than another word due to the densely populated neighborhood it is found in. A result of a high frequency word being "hidden" by its many neighbors may be the selection, or mis-selection in this case, of an alternate word that is relatively more frequent. The selected word will most likely resemble the intended word due to the organization of the lexicon. These factors may conspire to make a different, more frequent word, even more likely to be selected by the speech production system, resulting in the system producing an error.

Conversely, low-frequency words may not have enough activation on their own to be selected by the speech production system's criteria of "pick the most active candidate." Low-frequency words may need the supportive activation of their neighbors in order to be selected by the speech production system. Thus, low-frequency words in sparse neighborhoods do not receive enough supportive activation from their few neighbors. Consequently, these words may be overlooked by the speech production system, and be involved in more speech errors. These "conspiracy effects" (see Taraban and McClelland, 1987) among words and their neighbors may account for the differential effects of frequency and neighborhood density on the speech production system.