

The effect of polysemy on lexical decision time: Now you see it, now you don't

MICHELLE L. MILLIS

Santa Clara University, Santa Clara, California

and

SCOTT B. BUTTON

New College of the University of South Florida, Sarasota, Florida

Gernsbacher (1984) found that number of word meanings (polysemy) did not influence lexical decision time when it was operationalized as number of dictionary definitions. This finding supports her contention that subjects do not store all possible dictionary meanings for words in memory. The present experiments extended Gernsbacher's research by determining whether more psychologically valid measures of polysemy affect lexical decision time. Three metrics were used to represent the meanings that subjects actually access from memory (*accessible polysemy*): (1) the first meanings subjects think of when asked to define stimulus words, (2) all the meanings subjects generate for words, and (3) the average number of meanings subjects generate. The results showed that the second and third metrics of polysemy influenced lexical decision time, whereas the first metric (representing mostly the access to dominant meanings for words) only approached significance.

Rubenstein, Garfield, and Millikan (1970) examined the effects of word frequency and number of word meanings (*polysemy*) on lexical decision time. They measured polysemy by having subjects write down the first meaning they could think of for each of a set of words. The results showed that the words that were given two to four different meanings (*homographs*) were recognized more quickly than the words that were given only one meaning. The results also showed that lexical decision latencies decreased as the frequency of the words increased.

To account for these findings, Rubenstein et al. (1970) proposed a word recognition model that involves four processes: (1) the stimulus is divided into letters, (2) a subset of lexical entries is marked for comparison with the stimulus, (3) the entries are compared, and (4) an entry is selected as a response. The word frequency effect is obtained because during the marking process, entries are selected for comparison according to frequency. Thus, high-frequency words are responded to more quickly than low-frequency words because their entries are marked first. The polysemy effect is obtained because during the comparison process lexical entries are accessed at random. Thus, the probability that one of these entries will be accessed is greater the more lexical entries a word has.

Clark (1973) argued that Rubenstein et al. (1970) failed to demonstrate statistically that their findings would generalize beyond the particular stimulus set that was used. Clark reanalyzed their data treating both subjects and

items as random effects and found that the effect of polysemy disappeared. Only the effect of frequency on lexical decision time was shown to be statistically reliable.

Forster and Bednall (1976) also compared lexical decision latencies for words that varied in frequency and polysemy. They replicated Rubenstein et al.'s (1970) finding that reaction time was faster for high-frequency words than for low-frequency words. However, polysemy was not found to affect lexical decision time. These results are not surprising in light of Clark's (1973) contention that the polysemy effect obtained in Rubenstein et al.'s research was peculiar to the specific stimulus set that was used.

Although Forster and Bednall (1976) failed to replicate Rubenstein et al.'s (1970) findings, they retained many aspects of the original word recognition model. For example, they agreed that the access of word entries is directed by frequency and that different word meanings are stored as distinct, separate entries. However, they argued that the different meanings of words are not searched for randomly, but rather are accessed sequentially according to frequency of use.

Jastrzemski (1981) and Jastrzemski and Stanners (1975) argued that previous research may have failed to show an effect of polysemy on lexical decision time because the words that were used did not vary sufficiently in number of meanings. In addition, they stated that very few words have only one definition listed in an unabridged dictionary. Thus, instead of comparing homographs with words having single meanings, they compared lexical decision times for words having several meanings with words having few meanings. They also questioned the use

Address correspondence to Michelle L. Millis, Department of Psychology, Santa Clara University, Santa Clara, CA 95053.

of subjects' reports to determine number of word meanings. Instead, they operationalized polysemy as the number of definitions listed for a word in an unabridged dictionary. Both Jastrzembki (1981) and Jastrzembki and Stanners (1975) found that lexical decision time decreased as number of dictionary definitions increased. In addition, Jastrzembki replicated the finding that lexical decision time decreased as word frequency increased.

Jastrzembki (1981) tested the prediction from Rubenstein, Lewis, and Rubenstein's (1971) word recognition model that, regardless of number of meanings, reaction time should be quicker for high-frequency words than for low-frequency words. His results failed to support the model by showing that lexical decision time for high-frequency words with few meanings was slower than that for low-frequency words with several meanings.

Gernsbacher (1984) questioned the psychological validity of Jastrzembki's (1981) operationalization of polysemy as number of dictionary definitions. She argued that it seems unlikely that subjects store all possible dictionary meanings for a word in memory. For example, she showed that professors were able to think of only 3 of the 15 definitions listed in an unabridged dictionary for the word "fudge." Gernsbacher suggested that Jastrzembki may have obtained an effect of number of dictionary definitions on lexical decision time because the frequency of the stimulus words was not adequately controlled. In Jastrzembki's experiments, word frequency was operationalized as the number of times words appear in printed text. The word frequencies were taken from Kučera and Francis's (1967) *Computational Analysis of Present-Day American English*. Gernsbacher suggested that a better measure of frequency of occurrence is the extent to which subjects encounter words in their everyday experience (*experiential familiarity*). She proposed that Jastrzembki's manipulation of polysemy may have been confounded with experiential familiarity; that is, the number of dictionary definitions may have increased with the experiential familiarity of the words. Thus, reaction time may have been quicker for words with several definitions because they were more familiar to the subjects.

To test this possibility, Gernsbacher (1984) presented subjects with words that varied orthogonally in experiential familiarity and number of dictionary definitions. Familiarity was measured by having subjects rate the stimulus words on a 7-point scale ranging from *very unfamiliar* to *very familiar*. The results showed that lexical decision time decreased as experiential familiarity increased. However, no difference in reaction time was obtained for words with more than 10 dictionary meanings and for words with only 1 dictionary entry.

The present experiments were designed to extend Gernsbacher's (1984) research by determining whether more "psychologically real" measures of polysemy influence lexical decision time. Rather than by counting the number of dictionary definitions, polysemy was operationalized as the number of different meanings that subjects were able to think of for a word (*accessible poly-*

semy). It was hypothesized that accessible polysemy would influence lexical decision time because it represents the number of meanings that subjects actually access from memory.

Accessible polysemy was measured in three different ways. In Experiment 1, subjects wrote down the first meaning they could think of for each of a set of words (Rubenstein et al., 1970). The total number of different meanings supplied across subjects for each word was then used as the *first-meaning metric* of polysemy. In Experiments 2 and 3, two new measures of polysemy were introduced. Instead of recording only one meaning, subjects in these two experiments were given unlimited time to write down all of the meanings they could think of for each word. In Experiment 2, polysemy was measured as the total number of different meanings that were generated across subjects for a word (*total-meanings metric*). In Experiment 3, polysemy was measured as the average number of meanings that subjects supplied (*average-meanings metric*).

EXPERIMENT 1

Experiment 1 was designed to determine whether Rubenstein et al.'s (1970) metric of polysemy (i.e., first-meaning metric) would affect lexical decision time if the stimulus words differed by several meanings. Following Rubenstein et al.'s technique, subjects were asked to write down the first meaning for each of a set of words. Accessible polysemy was operationalized as the total number of different meanings that were generated across subjects for each word. To avoid Jastrzembki's (1981) criticism that Rubenstein et al.'s stimuli did not differ adequately in number of meanings (most of their homographs had only two meanings), we selected, for the present research, homographs that have four and five different meanings.¹

Following Gernsbacher's (1984) technique, we measured familiarity using a 7-point scale. Initially, familiarity was to be varied orthogonally with polysemy. However, because few words were given both low familiarity ratings and several different meanings, familiarity was held constant and number of meanings was varied. The purpose of Experiment 1 was to determine whether polysemy, as assessed by the first-meaning metric, influences lexical decision latency when the stimulus words vary widely in number of meanings.

Method

Materials. A list was composed of 104 five-letter words that occur within the two frequency ranges 11–13 and 200–397 (Kučera & Francis, 1967). Because frequency and familiarity are highly correlated (Gernsbacher, 1984), this selection was made to ensure that the words would vary widely in familiarity. All the words were ordered randomly and typed onto five pages. Seventeen subjects not used in the lexical decision task were asked to rate each of these words for familiarity on a 7-point scale, ranging from *unfamiliar* to *very familiar*. Then they were asked to write one meaning for each word. Two judges determined the total number of meanings

that were generated for each word. Only the words on which the judges agreed were used in the lexical decision task.

Forty words were selected from the initial set of words to serve as stimuli in Experiment 1. Half of these words were given one meaning, and the other half were given an average of 4.45 meanings ($SD=0.60$). The two groups of words were equated for experiential familiarity ($M=5.00$, $SD=.97$; $M=4.79$, $SD=.90$, respectively) and for printed word frequency ($M=149.45$, $SD=161.51$; $M=147.80$, $SD=149.69$, respectively). All 40 of the stimulus words are listed in Appendix A. The 40 nonwords that were used in this experiment were pronounceable, but unfamiliar and meaningless.

Apparatus and Procedure. Stimulus generation and data collection were controlled by an Apple IIe microcomputer. The words were presented in uppercase white letters on a 7½-in. black-and-white Sanyo monitor. At a viewing distance of 38.19 in., each five-letter word subtended a visual angle of 1.50° horizontally and 0.38° vertically.

At the beginning of each trial, a fixation point was presented in the center of the screen simultaneously with a brief warning tone. The stimulus item was presented 500 msec later and remained on the screen until the subject made a response. The subject responded by pressing the "key of the keyboard to indicate that the item was a word, and the A key of the keyboard to indicate that the item was not a word. (Responses made on any keys on the right side of the keyboard were recorded as words and those made on the left side were recorded as nonwords.) The offset of the stimulus item was followed by a 1,000-msec blank interval before the next trial began. The subjects were initially given 10 trials with a practice set of five words and five nonwords.

Subjects. Eighteen New College undergraduates with normal or corrected-to-normal vision were paid to serve as subjects in the lexical decision tasks. These students were taken from the same population as the subjects who had generated the word meanings and familiarity ratings. None of the students who had rated the words served as subjects in these experiments.

Results and Discussion

Two reaction times for correct responses were omitted from the data analyses because they exceeded 2,000 msec (Jastrzemski, 1981). The mean reaction times were submitted to an analysis of variance, with number of meanings as a within-subjects factor. Following Clark (1973), separate analyses were conducted, with subjects and items treated as random effects. These analyses revealed that polysemy, as assessed by the first-meaning metric, did not significantly affect reaction time when subjects was treated as the random factor [$F_1(1,17) = 3.15$, $p > .05$, $MS_e = 16.08$], or when both subjects and items were random [$\text{min}F'(1,22) = 2.75$, $p > .05$]. Number of meanings had a significant effect only when items was treated as random [$F_2(1,38) = 21.93$, $p < .01$, $MSe = 2.31$]. The words with several meanings were recognized in 589 msec, and the words with one meaning were recognized in 676 msec.

The proportions of errors out of total trials for the several-meanings and one-meaning conditions were 3.3 and 5.0, respectively. In this and in the remaining experiments, faster reaction times were always accompanied by fewer errors. Thus, the reaction time data in the three experiments reported here cannot be accounted for in terms of speed-accuracy tradeoffs.

These results show that even when the stimulus words varied widely in number of meanings, the first-meaning

metric of polysemy showed only a nonsignificant trend to influence lexical decision time. Perhaps this metric is an insensitive measure of polysemy because it reflects mostly the dominant meanings of words rather than all the meanings that subjects can access. For example, consider the situation in which every subject thinks of the same meaning for a word when asked to report the first meaning that comes to mind. This consistency in responding does not necessarily indicate that the reported meaning is the only meaning represented in the subjects' memories for that word. Several meanings may be stored, but a well-agreed-upon meaning may dominate over all the others and be accessed first. When given the word PEN, for example, all subjects may first think of "a writing instrument." However, other meanings such as "the act of writing," "a place to keep pigs," "the act of trapping," and "a slang word for jail" may also be represented in memory. Using the first-meaning metric of polysemy, PEN would be determined to be a nonhomograph even though it has at least five different meanings. On the other hand, a word with two equiprobable meanings, such as BARK (e.g., sound dogs make and tree covering), would be determined to be a homograph even though it actually has fewer meanings than PEN. Thus, the first-meaning metric is an inappropriate measure of the meanings that subjects can access from memory.

EXPERIMENT 2

The purpose of Experiment 2 was to establish a better measure of accessible polysemy than the first-meaning metric. Instead of reporting the first meaning that came to mind, subjects were asked to record all the meanings they could think of for each of a set of words. The total number of meanings supplied across subjects for each word was then used as the total-meanings metric. It was hypothesized that this measure of accessible polysemy would affect lexical decision time because it represents dominant and more subordinate meanings of words.

Method

Stimuli. A list was composed of 55 five-letter words. These words were ordered randomly and typed onto four pages. Twenty subjects not used in the lexical decision task were asked to rate the words for familiarity on a 7-point scale. A separate group of 14 subjects was given unlimited time to write all the meanings they could think of for each of these words. This change in procedure from Experiment 1 ensured that the familiarity ratings did not affect the meanings that subjects recorded for the words. Two independent judges determined the total number of meanings supplied for each word. Again, only the words on which the judges agreed were used in the lexical decision task.

Twenty-four words were chosen from this initial group to be used in the present experiment. Half of these words were selected to have an average of 3.08 meanings ($SD=0.90$), and the other half an average of 7.25 meanings ($SD=1.14$). The two groups of words were equated for printed word frequency ($M=143.67$, $SD=104.37$; $M=142.17$, $SD=76.71$, respectively); however, it was not possible to match the words for familiarity. Thus, the words with few meanings were chosen so that they had slightly higher experiential familiarity ratings than the words with many meanings ($M=4.65$,

$SD=0.90$; $M=4.28$, $SD=1.12$, respectively). Because reaction time increases with word familiarity, the finding that lexical decision time is quicker for words with many meanings than for words with few meanings cannot be attributed to differing word familiarities. All of the stimulus words are listed in Appendix B. The 24 non-words that were used in this experiment were pronounceable, but unfamiliar and meaningless.

Apparatus and Procedure. The apparatus and procedure were identical to those of Experiment 1.

Subjects. Eighteen Santa Clara University undergraduates with normal or corrected-to-normal vision served as subjects in the lexical decision task to receive course credit. These students were taken from the same population as the subjects who had generated the word meanings and familiarity ratings. None of the students who had rated the words for familiarity or supplied meanings served as subjects in the lexical decision experiment.

Results

Four reaction times were omitted from the data analyses. The analyses of reaction time revealed that lexical decision times were significantly faster for words with several meanings (609 msec) than for words with few meanings (748 msec) when subjects was treated as the random factor [$F_1(1,17) = 280.23$, $p < .01$, $MS_e = 0.52$], when items was treated as random [$F_2(1,22) = 36.34$, $p < .01$, $MS_e = 4.01$], and when both subjects and items were random [$\text{min}F'(1,28) = 32.17$, $p < .01$]. The proportions of errors for these two conditions were 3.2 and 9.7, respectively. These results show that polysemy does affect lexical decision time when it is operationalized as the total number of meanings that subjects generate for words.

Taken together, the results of Experiments 1 and 2 suggest that the total-meanings metric is a better measure of the meanings that subjects access from memory than is the first-meaning metric. Presumably, the first-meaning metric does not significantly influence reaction time because it does not represent both the subordinate and dominant meanings of words. Thus far, no direct evidence has been provided for this contention. Therefore, Experiment 3 was designed to compare the meanings that are recorded when subjects are asked to generate all the meanings they know for a word versus when subjects are asked to write down only the first meaning that comes to mind. The prediction was that subjects should record subordinate meanings as well as the more well-accepted meanings for words when asked to record all of the meanings they know, but that they should show a tendency to report mostly dominant meanings when using the first-meaning metric.

A second purpose of Experiment 3 was to test an alternative metric of accessible polysemy that seems to be a more appropriate measure of the meanings that subjects can access from memory than the total-meanings metric used in Experiment 2. Consider the situation in which one subject reports two meanings for a word and a second subject reports two entirely different meanings for that same word. According to the total-meanings metric, that word has four accessible meanings. However, neither of the subjects alone is able to access more than two meanings

for that word. Thus, a more appropriate measure of accessible polysemy would be to estimate the word as having two meanings, the average number of meanings that each subject can access for a word. Consequently, subjects were again asked to record all of the meanings they could think of for the stimulus words. However, instead of using the total number of meanings, the average number of meanings generated across subjects was used as the metric of accessible polysemy (average-meanings metric).

EXPERIMENT 3

Method

Stimuli. A list was composed of 166 four-letter words. These words were ordered randomly and typed onto five pages. Thirty-five subjects not used in the lexical decision task were asked to rate these words for familiarity. Sixty of these words that had similar experiential familiarity ratings were selected to be used in the next phase of the experiment ($M=5.71$, $SD=0.18$). These 60 words were ordered randomly and typed onto eight pages. Accessible polysemy was then measured in two ways. One group of 21 subjects recorded all the meanings they could think of for each of these words, and 2 judges determined the average number of meanings that were generated for each word (average-meanings metric). Another group of 25 subjects wrote the first meaning that came to mind for each of the same words, and the judges determined the total number of meanings that were supplied (first-meaning metric). Again, only the words on which the judges agreed were used in the lexical decision task.

From the set of 60 words, 24 were chosen to be used in the actual experiment. These words were chosen on the basis of the number of meanings that were generated by the average-meanings metric; half had an average of 1.38 meanings ($SD=0.34$), and half had an average of 2.36 meanings ($SD=0.34$). The two groups of words were equated for familiarity ($M=5.77$, $SD=0.16$; $M=5.72$, $SD=0.16$, respectively) and for printed word frequency ($M=63.92$, $SD=70.42$; $M=63.83$, $SD=74.58$, respectively). All of the stimulus words are listed in Appendix C. The 24 nonwords that were used in this experiment were pronounceable, but unfamiliar and meaningless.

Apparatus and Procedure. The apparatus and procedure were identical to those of the first two experiments, except that the stimulus words contained four letters instead of five. Thus, each word subtended a visual angle of 1.13° instead of 1.50° horizontally.

Subjects. Twenty Santa Clara undergraduates with normal or corrected-to-normal vision served as subjects in the lexical decision task to receive course credit. Like the students participating in Experiment 2, these students were taken from the same population as the subjects who had generated the word meanings and familiarity ratings. None of the students who had rated the words served as subjects in the lexical decision experiment.

Results

Five reaction times were omitted from the data analyses. The analyses of reaction time revealed that lexical decision times were significantly quicker for words with 2.4 meanings (661 msec) than for words with 1.4 meanings (732 msec) when subjects was treated as the random factor [$F_1(1,19) = 37.49$, $p < .01$, $MS_e = 1.03$], when items was treated as random [$F_2(1,22) = 19.70$, $p < .01$, $MS_e = 1.96$], and when both subjects and items were random [$\text{min}F'(1,39) = 12.91$, $p < .01$]. The proportions

of errors for these two conditions were 3.8 and 6.3, respectively. These results show that polysemy influences reaction time when it is operationalized as the average number of meanings that subjects access for words.

To determine the suitability of the first-meaning metric as a measure of accessible polysemy, the meanings that were generated by this metric were compared with the meanings generated by the average-meanings metric for the 24 stimulus words. As predicted, subjects in the average-meanings metric group recorded subordinate as well as dominant meanings for the words. Subjects in the first-meaning metric group, however, showed a tendency to report mostly dominant meanings. For example, when given the word TELL, every subject in the first-meaning metric group recorded "to inform." This meaning was also recorded by all of the subjects in the average-meanings metric group. However, these subjects generated three other meanings as well, including: "to explain," "to relate in detail," "to understand." As another example, when given the word LOAD, subjects in the first-meanings group recorded either "pack" or "quantity of material assembled." These two meanings were also the most well-agreed-upon meanings generated by the average-meanings group. However, these subjects recorded six additional meanings, such as "to put a load in a carrier," "courses taken," "a burden."

A review of these meanings showed that when subjects were asked to record the first meaning that came to mind, several of the meanings that may have been accessed from memory were not. Thus, the first-meaning metric seems to be an inadequate measure of accessible polysemy. Two findings from Experiment 3 support this contention. First, the Pearson product-moment correlation for the first-meaning metric and reaction time was not significant ($r = -.31, p > .05$). This result is consistent with the finding in Experiment 1 that the first-meaning metric of polysemy did not significantly affect lexical decision time. Second, indirect evidence comes from the finding that the average-meanings metric of polysemy did significantly influence reaction time even though the two conditions differed by only one meaning (2.4 vs. 1.4). This result suggests that Rubenstein et al. (1970) may have failed to obtain a polysemy effect not because their stimulus words differed by too few meanings, but because their measure of polysemy (first-meaning metric) was inappropriate. However, because the results of Experiment 1 approached significance, it is possible that the first-meaning metric may be an adequate measure of polysemy, but that neither the present research nor Rubenstein et al.'s experiments were sensitive enough to obtain an effect.

GENERAL DISCUSSION

Whereas several investigators claimed that number of word meanings influences lexical decision time (e.g., Jastrzemski, 1981; Rubenstein et al., 1970), others have failed to obtain a polysemy effect (e.g., Gernsbacher, 1984). One possible reason for the inconsistency of these

results is that polysemy was operationalized in different ways. The findings from these experiments cannot be directly compared because the studies also differed on a number of other factors, including (1) the stringency of the statistical analyses performed on the data, (2) control of word familiarity, and (3) the variation in number of word meanings. The purpose of the present research was to examine the effects of a variety of measures of polysemy on reaction time while taking these three factors into account. In all three experiments separate analyses were conducted with subjects and items treated as random effects. In all three experiments word familiarity was controlled. In Experiments 1 and 2 the stimulus words varied widely in number of meanings, whereas in Experiment 3 the two conditions differed by only one meaning. The finding that polysemy significantly affected reaction time in Experiments 2 and 3 but not in Experiment 1 suggests that the polysemy effect may depend not on the extent to which number of meanings differ, but rather on the measure of polysemy that is used.

Three measures of accessible polysemy were examined: the first meanings that came to mind for a set of stimulus words (first-meaning metric), the total number of meanings that subjects could think of for words (total-meanings metric), and the average number of meanings that subjects could generate (average-meanings metric). It was hypothesized that polysemy would influence reaction time using each of these measures because they represent the meanings that subjects actually access from memory.

The results of Experiment 1 showed only a non-significant trend for polysemy to influence reaction time when it was operationalized as the first-meaning metric, even when the stimulus words varied widely in number of meanings. The results of Experiments 2 and 3, however, showed that the total-meanings metric and the average-meanings metric did influence lexical decision time. We concluded that these latter two metrics are good measures of polysemy because they represent the subordinate as well as the dominant meanings that subjects access for words. Perhaps the effect of polysemy on reaction time is weakest when operationalized as the first-meaning metric because this measure does not represent enough of the subordinate meanings of words.

At first glance, the results of these experiments seem to be inconsistent with Gernsbacher's (1984) finding that number of word meanings (operationalized as dictionary counts) did not influence lexical decision latency. However, Gernsbacher herself claimed that number of dictionary definitions may not adequately reflect the number of meanings that are represented in memory. Her failure to obtain an effect of polysemy on lexical decision time supports her contention that this manipulation is not psychologically valid. Even Jastrzemski (1981) stated that "there certainly is some limit to the number of meanings of a word with which subjects are acquainted. Once this limit has been reached, it would not matter how many additional meanings a word has" (p. 290). The present research extends Gernsbacher's research by show-

ing that polysemy does influence lexical decision time when the metric that is used reflects the meanings that subjects access for a word.

Gernsbacher (1984) argued that experiential familiarity is a better measure of word frequency than is the number of times that words appear in printed text. Similarly, the present research argues that accessible polysemy is a better measure of number of word meanings than is the number of definitions listed in an unabridged dictionary. The experiments reported here support Gernsbacher's general contention that word properties should be measured by determining how words are represented "in the head" rather than by using objective characteristics of the printed word, such as frequency counts and dictionary definitions.

The finding in this research that polysemy influences lexical decision time indicates that the effect of number of word meanings must be accounted for in any model of the lexicon. A recent version of Morton's (1979) logogen model can be used to account for the polysemy effects obtained in Experiments 2 and 3. According to this model, each meaning of a word is represented by a separate logogen. Thus, words with several meanings have more corresponding logogens than words with few meanings. Morton did not adequately explain how the extra representations in lexical memory facilitate recognition for words with multiple meanings. However, Jastrzembski (1981) extended Morton's model by proposing that a word is recognized whenever the threshold of any of its corresponding logogens has been exceeded. All of the logogens representing a particular word are presumed to have the same threshold level. According to this proposal, the greater the number of logogens representing a particular word, the greater the probability that one of these logogens will accumulate enough information for that word to be recognized. This proposal suggests that words with several meanings should be more likely to be recognized than words with few meanings because they have more logogens in lexical memory.

The contention that logogens representing the various meanings of a word have the same threshold value seems oversimplified. In one's everyday experience, the different meanings of a word are used with different frequencies. Both the present research and Gernsbacher's (1984) experiments show that the subject's experience with a word is a very important determinant of lexical decision time. Thus, it seems more reasonable to assume that the logogens for the different meanings of a word have varying thresholds, depending on how often those particular meanings are used in the subjects' everyday experience. One possibility is that all the logogens for a word accumulate information at the same time, but that the logogen representing the most frequently used meaning exceeds its threshold first (*lowest-threshold hypothesis*). Another possibility is that the threshold value is averaged across logogens such that the amount of information accumulated by all the logogens determines when the word be-

comes available as a response (*average-threshold hypothesis*).

A variation of the metrics used in Experiments 2 and 3 could be used to test between these two hypotheses by requiring subjects to rate the familiarity of each meaning that is generated for a word. For example, consider the situation in which subjects generate two meanings for the word BARK (e.g., sound dogs make; covering on trees), and two meanings for the word BANK (e.g., place money is kept; land next to rivers). Suppose further that, using a 7-point scale, subjects rate the familiarity of the "dog" and "tree" meanings of BARK as 6 and 4, respectively, and the "money" and "river" meanings of BANK as 6 and 2. The lowest-threshold hypothesis would predict the BARK and BANK should have equal lexical decision times because the dominant meaning for each word has the same familiarity rating (i.e., 6) and, thus, the same threshold value. The average-threshold hypothesis, on the other hand, would predict that reaction time should be quicker to BARK than to BANK because the average familiarity rating for BARK $[(6 + 4)/2 = 5]$ is higher than the average rating for BANK $[(6 + 2)/2 = 4]$. Thus, the overall threshold value of BARK should be lower than the average threshold value of BANK. This possibility is currently being tested.

Another direction for future research would be to determine whether the total-meanings metric (used in Experiment 2) or the average-meanings metric (used in Experiment 3) is the better measure of accessible polysemy. Consider the situation in which one subject records two meanings for the word BOMB (e.g., an explosive device; to destroy), and a second subject records two different meanings for that word (e.g., a failure; a long pass in football). The total-meanings metric would estimate the word as having four accessible meanings, whereas the average-meanings metric would estimate the word as having two meanings. Now, consider the situation in which one subject records two meanings for the word SUIT (e.g., a formal outfit; a suit charged against someone), and a second subject reports the same two meanings. Both the total-meanings and the average-meanings metrics would estimate SUIT as having two accessible meanings. To determine which metric is the better measure of accessible polysemy, the lexical decision times to the words BOMB and SUIT could be directly compared. If the total-meanings metric is the appropriate measure to use, then reaction time should be quicker to the word BOMB than to the word SUIT, because it has more meanings. On the other hand, if the average-meanings metric is appropriate, then the reaction time for the two words should not differ because they have the same number of meanings. This possibility is also currently being tested.

The important finding of the present research is that polysemy affects lexical decision time when the metric that is used reflects the meanings that subjects can access from memory for a word. These experiments, therefore, show that it is important to control for number of mean-

ings when investigating the effects of other variables on word processing. Thus far, most experiments on lexical decision have not considered the role of this possible confounding variable on reaction time.

REFERENCES

CLARK, H. H. (1973). The language-as-a-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior*, **12**, 335-359.

FORSTER, K. I., & BEDNALL, E. S. (1976). Terminating and exhaustive search in lexical access. *Memory & Cognition*, **4**, 53-61.

GERNSBACHER, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, **113**, 256-281.

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

JASTRZEMBSKI, J., & STANNERS, R. (1975). Multiple word meanings and lexical search speed. *Journal of Verbal Learning & Verbal Behavior*, **14**, 534-537.

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

MORTON, J. (1979). Word recognition. In J. Morton & J. C. Marshall (Eds.), *Psycholinguistics 2: Structure and processes*. Cambridge, MA: MIT Press.

RUBENSTEIN, H., GARFIELD, L., & MILLIKAN, J. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning & Verbal Behavior*, **9**, 487-494.

RUBENSTEIN, H., LEWIS, S. S., & RUBENSTEIN, M. A. (1971). Homographic entries in the internal lexicon: Effects of systematicity and relative frequency of meaning. *Journal of Verbal Learning & Verbal Behavior*, **10**, 57-62.

NOTE

1. Rubenstein et al. (1970) did not report the number of words that were given two, three, and four meanings. However, it is possible to determine from their data analysis that 22 words had two meanings and 18 words had more than two meanings. Even if all 18 words were given the maximum of four meanings, the average number of meanings for Rubenstein's homographs could not exceed 2.9: $[(2 \times 22) + (4 \times 18)]/40 = 2.9$.

APPENDIX A
Stimuli Used in Experiment 1

| One Meaning | Many Meanings |
|-------------|---------------|
| exert | shaft |
| rabbi | mound |
| overt | crest |
| cliff | flush |
| imply | point |
| lyric | order |
| couch | force |

APPENDIX A (Continued)

| One Meaning | Many Meanings |
|-------------|---------------|
| shelf | front |
| aloud | glaze |
| among | globe |
| buddy | array |
| waist | harsh |
| north | stray |
| south | level |
| whole | handy |
| death | known |
| often | young |
| group | close |
| white | class |
| thing | light |

APPENDIX B
Stimuli Used in Experiment 2

| Few Meanings | Many Meanings |
|--------------|---------------|
| story | level |
| earth | stock |
| novel | front |
| spoke | range |
| build | force |
| aside | bound |
| chest | check |
| share | stand |
| group | floor |
| sense | block |
| paper | court |
| staff | sport |

APPENDIX C
Stimuli Used in Experiment 3

| Few Meanings | Many Meanings |
|--------------|---------------|
| huge | lane |
| lake | fool |
| pond | lock |
| past | dull |
| mile | horn |
| pair | dive |
| ship | rear |
| loan | tell |
| chin | tail |
| sand | fine |
| suit | load |
| plug | bond |

(Manuscript received March 9, 1987;
revision accepted for publication July 5, 1988.)