DOCUMENT RESUME

ED 116 126 CS 002 278

AUTHOR Matlin, Margaret W.
TITLE Relationship between Affect and Estimates of English Word Frequency.
PUB DATE 31 Aug 75
NOTE 12p.; Paper presented at the Annual Meeting of the American Psychological Association (83rd, Chicago, Illinois, August 31, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS *American English; Language Research; Language Tests; *Psychological Tests; Vocabulary; *Word Frequency
IDENTIFIERS Francis (W N); Kucera (H)

ABSTRACT

The experiment reported in this document extended the study by Milburn and Bell (1969) of English word frequencies. Specifically, this experiment examined the influence of instructional set on the relationship between affect and frequency estimates, with a distinction between personal and nationwide norms. The experiment design provided for two within-subjects variables--natural language frequency and affect (three categories: positive, neutral, and negative)--and one between-subjects variable--instructions (two categories: personal and Kucera-Francis). Sixty-nine subjects completed the two tests. Results showed that there is a curvilinear relationship between affect and estimates of English word frequency; that subjects believe that positive and negative English words occur equally often, but more often than neutral words; that the influence of affect on frequency estimates was strongest in the personal instructions condition and in the lower frequency range; and that subjects tended to make more conservative judgments in the Kucera-Francis condition than in the personal condition. (JM)

Relationship between Affect and Estimates of English Word Frequency

Margaret W. Matlin

S.U.N.Y. at Geneseo

Margaret W. Matlin

Paper presented at the 83rd Annual Convention of the American Psychological Association, Chicago, August 31, 1975.

Experimental psychology appears to be experiencing a rebirth of interest in frequency estimation. In the majority of these studies, frequency has been experimentally manipulated in the laboratory (Begg, 1974; Hintzman, 1969; Howell, 1973; Underwood et. al., 1971). Some of these experiments have been concerned with the influence of affect on frequency estimates but the exact nature of this influence is unclear. Several studies have shown a linear relationship between affect and estimated frequency (Stang, 1974; Matlin & Stang, 1975a; Matlin & Stone, 1975) while others have demonstrated a curvilinear relationship (Gerard, et al., 1973; Stang, 1974). In other words, pleasant items are consistently judged to have occurred more frequently than neutral items. However, it is unclear, whether unpleasant items are judged to occur more frequently or less frequently than neutral items. This literature is reviewed in Matlin and Stang (1975b).

In contrast to this abundant literature on experimentally-manipulated word frequencies, there is a surprising scarcity of literature on estimates for natural word frequency. We do have evidence that subjects are quite accurate in estimating the frequency with which words appear in the English language (Howes, 1964; Shapiro, 1969), but few studies have explored the factors which may influence these estimates of English word frequency.

One exception is a study by Milburn and Bell (1969). These authors asked subjects to judge English word frequencies on a 0-7 frequency-rating scale. There were 24 English words, which had been previously rated by other subjects as being positive, neutral, or negative. The results showed that subjects judged both positive and negative words as being more frequent than neutral words but not differing in frequency from each other. Thus, these results support the curvilinear relationship between affect and frequency estimation observed by Gerard et. al., (1973) and Stang (1974).

My primary interest in pursuing this paradigm is to investigate the generality of Milburn and Bell's curvilinear relationship, especially in light of the

contradictory findings for experimentally manipulated frequencies. The present experiment is an extension of Milburn and Bell's study, modified by several methodological changes that will be discussed later. As an additional variable, I examined the influence of instructional set on the relationship between affect and frequency estimates. Milburn and Bell had asked subjects to make their judgments on the basis of how frequently they felt the word occurred in the English language or in printed material. These instructions imply to subjects that they must make judgments regarding the incidence of words in the English language on the basis of nationwide norms, rather than on the incidence among words they have encountered personally.

This distinction between personal and nationwide norms is an interesting one because subjects may predict different outcomes for themselves than for the rest of the world. There is abundant evidence, for example, that subjects typically believe that they are happier than average (Matlin and Stang, 1975b). Subjects also believe that their personal happiness ratings are substantially more positive than similar happiness ratings for the national situation (Watts and Free, 1974). It seems likely, then, that subjects may indeed judge positive and negative words as being equally frequent in a nation-wide sample. As Tesser has observed, "A glance at your morning paper suggests that, by and large, people relish communicating bad news," and subjects may give large frequency estimates for negative words on the basis of this knowledge. In contrast, subjects may tend in their own experiences to avoid bad news (Matlin and Stang, 1975b). If this is the case, then instructions for subjects to attend to the frequency of occurrence in their own experience will result in higher estimates for positive words than for negative words.

METHOD

Design

The design of the experiment provided for two within-subjects variables: Natural language frequency (seven frequency categories from Kucera-Francis norms (1967) of approximately logarithmically equal intervals: 1, 2-4, 5-9, 10-24, 25-49, 50-99, 100-249) and Affect (three affect categories: positive, neutral, and negative). The one between-subjects variable was Instructions (Two instruction categories: Personal and Kucera-Francis). Sixty-nine Ss were tested.

Stimulus Material

A list containing 210 words was derived by selecting 10 words from the Kucera-Francis norms (1967) for each category of affect and frequency. The 210 words were then rated by 106 naive Ss, and an average rating was obtained for each word. The final list of 105 words was selected from these rated words, in accordance with the following criteria: 1) The "positive" categories contained words with average ratings of 1.00-2.50, "neutral" between 3.50-4.50, and "negative" between 5.50-7.00. 2) For words in a given frequency category, all three evaluation categories were equated for word length, part of speech, and exact frequency. 3) For words in a given evaluation category, all seven frequency categories were equated for average affect rating. (Without this last proviso, high frequency categories would have had more positive affect ratings.)

The words were typed on two frequency estimation sheets. One of two instruction sheets was stapled to the front. The "Personal Frequency" instructions read:

> In this experiment I want you to make some judgments about English word frequency. Assume that in the last year in all your reading you have read a total of one million words. (Obviously, many words occurred more than once.) This might be in magazines, newspapers, and all types of books, including textbooks. I want you to estimate how many times in the last year you read each of the words that appear on these sheets. To give you some guidelines, imagine that the number of 250 is the maximum number of times you have read any of the words--250 is the largest number you can put down. For those words you have not read in the last year, put down 0. To repeat, put down some number from 0 to 250 in the space in front of each word. Be sure to judge the frequency of the word exactly as you see it on the sheet, not some other form of the word. (For example, if it says CHAIRS, write down the number of times you read the word CHAIRS, not CHAIR.) Glance over all the words before you begin

The "Kucera-Francis Frequency" sheet read:

> In this experiment I want you to make some judgments about English word frequency. A few years ago, a language expert took passages, one million words total, out of some newspapers, and books. (Obviously, many words occurred more than once.) He included magazines, newspapers, and all types of books, including textbooks. Then he counted up how many times each individual word occurred in his whole sample. Now, I want you to estimate how many times each of the words here was found in his sample. To give you some guidelines, imagine that the number 250 is the maximum number of times any of these words appeared in his sample--so 250 is the largest number you can put down. For those words you do not think appeared in the sample, put down 0. To repeat, put down some number from 0 to 250 in the space in front of each word. Be sure to judge the frequency of the word exactly as you see it on the sheet, not some other form of the word. (For example, if it says CHAIRS, write down the number of times he found the word CHAIRS, not CHAIR.) Glance over the words before you begin.

Subjects were allowed 15 minutes to complete the sheets.

RESULTS

Inspection showed that the data were markedly skewed, so the frequency estimates were converted using the formula: $\log_{10}(x + 2)$. An analysis of variance performed on the converted scores demonstrated the main effect of instructions, was not significant ($F = 5.67$, $df = 1\ 67$, $p>.05$) all other factors were significant: ($F = 365.79$, $df = 6/402$, $p<.001$): Affect ($F = 120.27$, $df = 2/134$, $p<.001$): Instructions x Affect ($F = 6.45$, $df = 2/134$, $p<.002$); Frequency x Affect ($F = 40.29$, $df = 12/804$, $p<.001$); Instructions x Frequency ($F = 8.42$, $df = 6/402$, $p<.001$; Instructions x Frequency x Affect ($F = 6.50$, $df = 12/804$, $p<.001$).

Post hoc analyses showed that the relationship between affect and frequency estimates was curvilinear ($F_{\text{quadratic trend}} = 34.52$, $df = 1/134$, $p<.001$). The mean frequency estimates were identical for positive and negative words (1.72). The mean frequency estimate for neutral words (1.52) was markedly lower. These results clearly confirm the findings of Milburn and Bell 91972).

The Instructions x Affect interaction is a primary focus of the experiment (see Figure 1).

Figure 1 about here

Note, however, that the Personal instructional set does not show the higher frequency estimates for positive words than for negative words that we had predicted. Instead, Personal instructions serve to heighten the contrast between neutral words and both other categories. In other words, all subjects estimate that positive and negative words occur more often than neutral words, but the discrepancy is largest when subjects are asked to judge the words in terms of their personal experiences. Subjects are willing to make extreme judgments about their personal experiences. They are more conservative about judgments for national norms.

Figure 2 illustrates the nature of the Frequency x Affect relationship. At

---

Figure 2 about here

---

high frequencies, affect does not influence frequency estimates. At lower frequencies, affect has an increasing influence on frequency estimates. Affect has the largest effect upon estimates in the lowest frequency category.

The Instructions x Frequency interaction is due to a wider spread of means for subjects responding in the Personal condition. The two groups estimate low frequency words similarly. However, subjects in the Kucera-Francis condition gave much more conservative estimates for the high frequency words than did subjects in the Personal condition. As in the Instructions x Affect interaction, subjects seem to be more willing to make extreme judgments about their personal experiences than about national norms.

Finally, let us consider the triple interaction Instructions x Frequency x Affect. Inspection of the means here shows that the Frequency x Affect interaction is stronger in the Personal instructions condition than in the Kucera-Francis instructions condition.

DISCUSSION

The present experiment differs methodologically in several ways from the Milburn and Bell (1972) study: 1) This study examined adjectives, verbs, and nouns while the earlier study used only nouns; 2) Subjects in the present study made magnitude estimates with anchoring points of 0 and 250, while the earlier study used a rating scale; 3) The present sample of words was substantially larger; 4) The present study equated negative and positive words for polarity (deviation from the affectively neutral point), while Milburn and Bell's negative words were more polarized than the positive words. Despite these differences, the Milburn and Bell results are confirmed here.

It is interesting that subjects are more conservative when they make judgments in terms of national norms. This observation bears some resemblance to the observation elsewhere (Begg, 1974) that means for immediate, online judgments show a wide spread as a function of frequency while means for delayed judgments tend to cluster closer to the overall mean. It seems that subjects do not hesitate in supplying extreme judgments for immediate experiences. With uncertainty, resulting either from a delay period or an ambiguous norm, judgments grow conservative.

CONCLUSIONS

There is a curvilinear relationship between affect and estimates of English word frequency; subjects believe that positive and negative English words occur equally often, but more often than neutral words. Furthermore, the influence of affect on frequency estimates was strongest in the personal instructions condition and in the lower frequency range. Finally, subjects tended to make more conservative judgments in the "Kucera-Francis" condition than in the "Personal" condition.

REFERENCES

Begg, I. Estimation of word frequency in continuous and discrete tasks. Journal of Experimental Psychology. 1974, 102, 1046-1052.

Gerard, H. B., Green, D., Hoyt, M. & Conolley, E. S. Influence of affect on exposure frequency estimates. Journal of Personality and Social Psychology, 1973, 28, 151-154.

Hintzman, D. L. Apparent frequency as a function of frequency and the spacing of repetitions. Journal of Experimental Psychology, 1969, 80, 139-145.

Howell, W. C. Representation of frequency in memory. Psychological Bulletin, 1973, 80, 44-53.

Howes, D. On the interpretation of word frequency as a variable affecting speed of recognition. Journal of Experimental Psychology, 1954, 48, 106-112.

Kucera, H. and Francis, W. N. Computational analysis of present-day American English. Providence, R. I.: Brown University Press, 1967.

Matlin, M. W. and Stang, D. J. Some determinants of word frequency estimates. Perceptual and Motor Sills, 1975a, 40, 923-929.

Matlin, M. W. and Stang, D. J. The Pollyanna Principle. Book in preparation, 1975b.

Matlin, M. W. and Stone, M. The effects of evaluation, activity, and potency on frequency estimates. Bulletin of the Psychonomic Society, 1975, 5, 391-392.

Milburn, T. W. and Bell, N. Values and estimation of word frequencies. Psychological Reports, 1972, 31, 47-56.

Shapiro, B. J. The subjective estimation of relative word frequency. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 248-251.

Stang, D. J. Influence of frequency and affect on exposure frequency estimates. Paper presented at the annual meeting of the Eastern Psychological Association, Philadelphia, May, 1974.

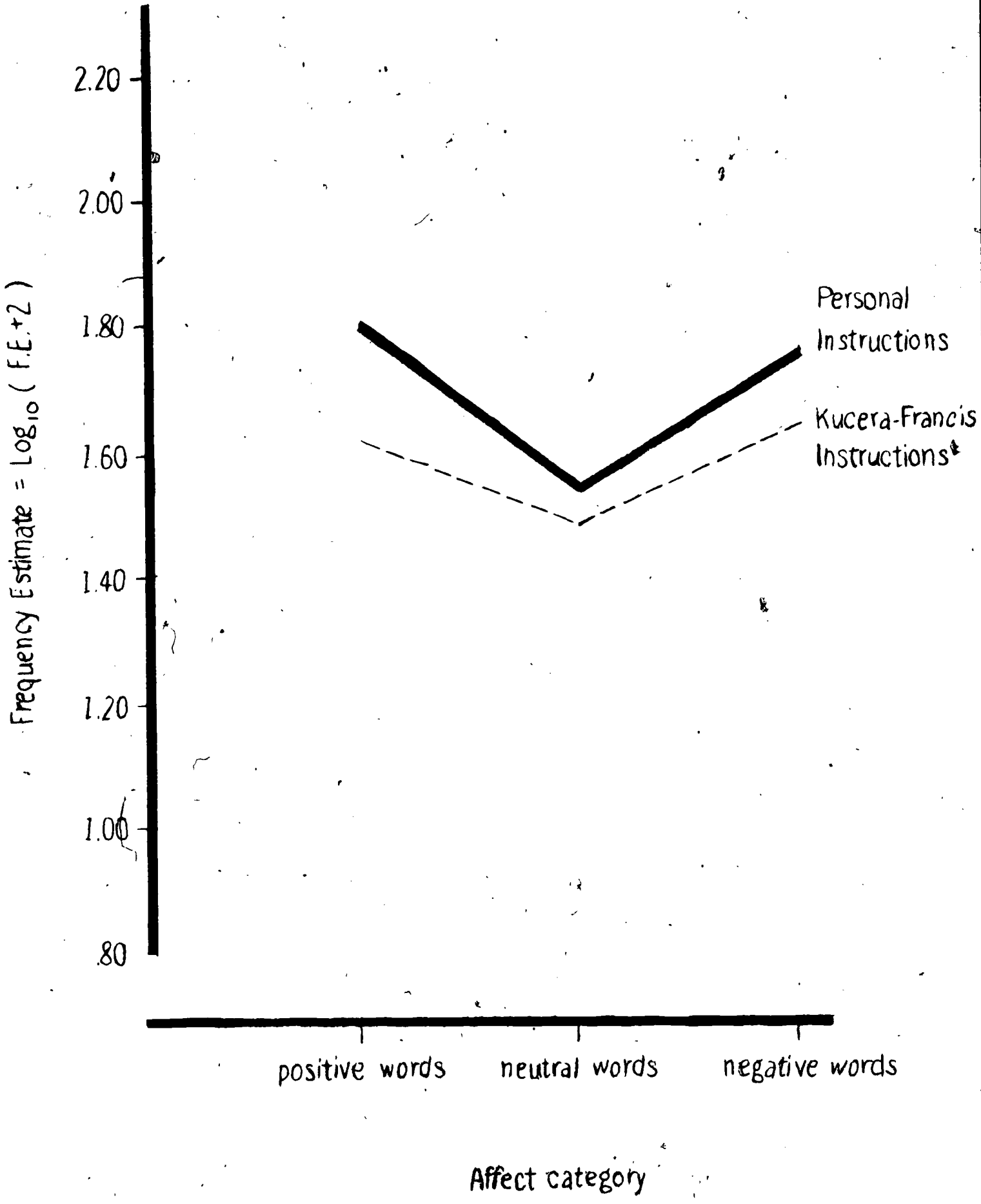
Underwood, B. J., Zimmerman, J. & Freund, J. S. Retention of frequency information with observations on recognition and recall. Journal of Experimental Psychology, 1971, 87, 149-162.

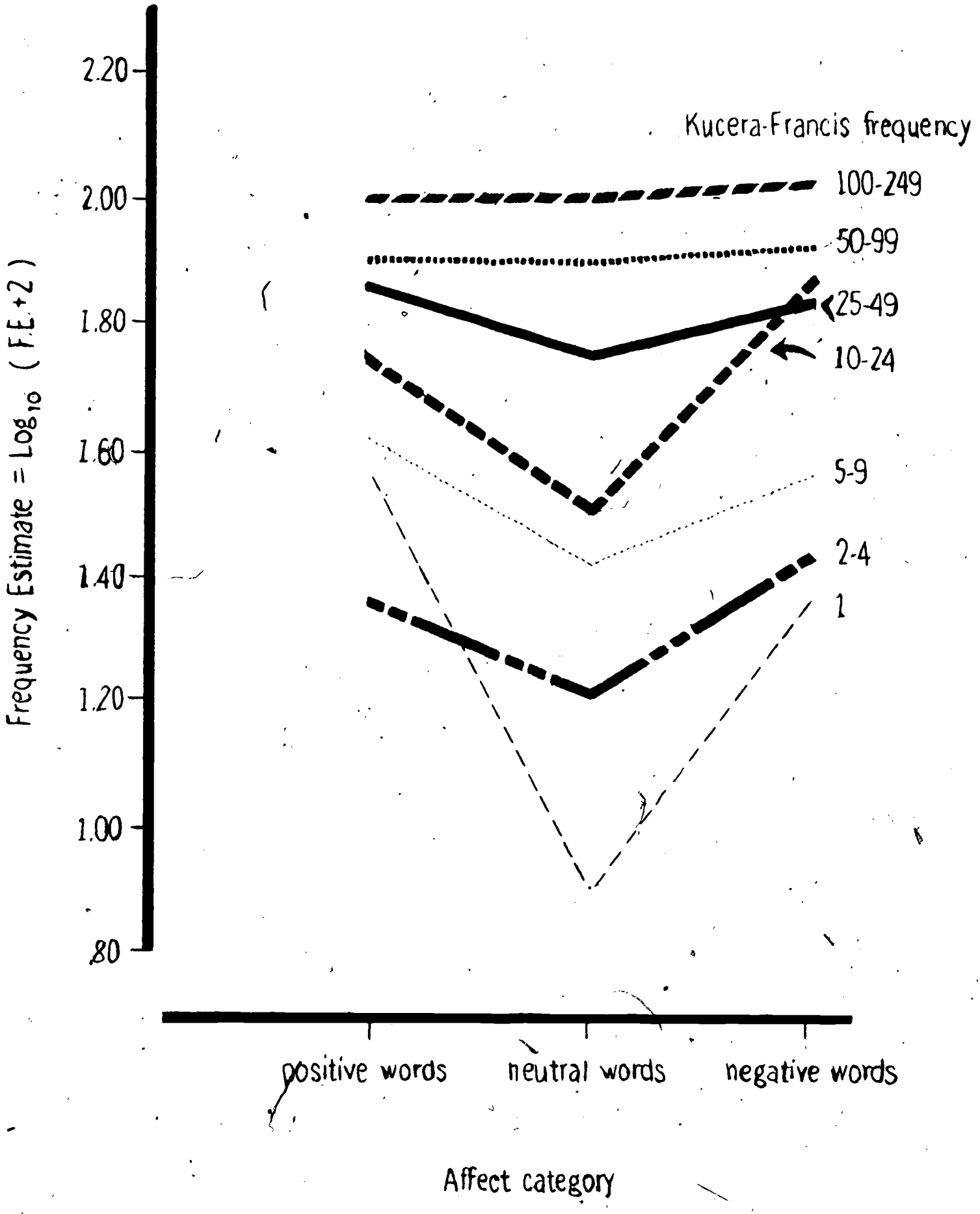
Watts, W. and Free, L. A. State of the Nation 1974. New York: Harper, 1974.

Figure Captions

Figure 1: The relationship between category of affect and frequency estimate for the two Instructions conditions

Figure 2: The relationship between category of affect and frequency estimate for the seven Frequency conditions.


2.20
2.00
1.80
1.60
1.40
1.20
1.00
.80
Frequency Estimate = Log10 (F.E.+2)
Personal Instructions
Kucera-Francis Instructions
positive words
neutral words
negative words
Affect category


Kucera-Francis frequency
100-249
50-99
25-49
10-24
5-9
2-4
1
2.20
2.00
1.80
1.60
1.40
1.20
1.00
.80
Frequency Estimate = Log10 (F.E.+2)
positive words
neutral words
negative words
Affect category