Miles Myers

PROTOTYPE THEORY AND HOLISTIC SCORING

The issues in the assessment of writing keep changing. At one time, the primary issue was the multiple-choice test versus the writing sample, but when scores of holistic assessments were shown to have excellent reliability, writing samples were generally accepted as the most appropriate data for writing assessments. The primary issues now are which writing samples should be collected (what topics or prompts, how many samples), and how the assessment should be anchored. It is the second question which I want to discuss.

There are those who argue that the best way to begin anchoring a group of readers is to give them a list of features—coherence, clear introduction, and so forth. Others argue that the best way to begin anchoring a group of readers is to have them score unmarked samples of the various score categories—1, 2, 3, and so forth. The argument is entirely a matter of emphasis because those who emphasize features use samples as supplementary information, and those who emphasize samples usually recommend that feature analysis be a follow-up activity to holistic scoring. At bottom, the issue is how do people form categories, with a list of features or with a sample.

A number of linguists have recently begun to challenge the traditional use of a checklist of semantic features to determine the necessary and sufficient conditions for set-membership in the category designated by a word. Kay and McDaniels (1978),[1] for instance, have shown that there are no necessary and sufficient conditions of *redness*, but rather red is a gradient quality whose category boundaries are fuzzy. George Lakoff (1972),[2] Charles Fillmore (1976),[3] Kay, and others argue that checklist semantics should be replaced by prototype semantics in which word meaning is determined by a central prototype or typical use of the word, not the yes/no category boundaries of a list of features. One example of the problem is Fillmore's question, "How old is a bachelor? Sixteen? Twenty-one? Thirty?"

The work of Eleanor Rosch has provided additional empirical evidence that people use prototypes, not a list of features, to define categories. Rosch (1977)[4] had people compose sentences with category names like bird, fruit, and vehicle. She then replaced the category name (for instance, bird) with a member of the category (for instance, robin, eagle, ostrich) and asked people to rate how sensible the resulting sentences were. The prototypical bird is something like a robin, producing sensible results more often than chicken or ostrich or numerous other choices.

Rosch and Mervis (1975)[5] also found that when people are asked to list the important features of a given member of a category, the list turned out not to match exactly the important features of another member of the same category. Rosch and Mervis asked people to list the characteristic attributes of twenty different kinds of fruit, and for these fruits, from apples to olives, there was no feature list common to all.

In prototype theory, then, a category is defined by a prototype or central tendency member which contains the maximal number of features in common with members of its own set and a minimal number of features in common

with members of other sets. In writing assessment, this means that the anchor papers for the one category are like the robin of the bird category. There are papers which are like chickens. That is, they are not exactly like the anchor prototype but they are close enough. Defining score categories by a list of features can lead readers to expect scoring to be a matter of yes or no, not more or less.

The more-or-less quality of paper scoring does not mean that features are ignored. Prototypes can be used to describe cases in which some features of the prototype for the category are more heavily weighted than others. But the more-or-less quality of paper scoring means the scoring is an approximation, not an adding up of features. Zadeh (1973), [6] the father of fuzzy set theory, has argued that in humanistic systems, people reason by approximations of fuzzy sets, not by precise, quantitative terms. Says Zadeh, "...as the complexity of the system increases, our ability to make precise yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance (or relevance) become mutually exclusive characteristics" (Zadeh, 1973, p. 28).

In summary, then, from linguistics (Fillmore, Lakoff, Kay), from psychology (Rosch, Mervis), and from philosophy (Zadeh), there is good evidence that the categories people use in their daily problem solving do not have precise boundaries and are defined first by matching to prototypes, not a list of features. Holistic scoring which emphasizes the anchor papers, not a list of features, is following the same set of principles.

[1]   Paul Kay and Chad D. McDaniel, "The Linguistic Significance of the Meaning of Basic Color Terms," *Language* 54:3 (1973).

[2]   George Lakoff, "Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts," in *CLS* 8, 1972.

[3]   Charles J. Fillmore, "Topics in Lexical Semantics," In R. Cole (Ed.) *Current Issues in Linguistic Theory*, Bloomington, 1976, pp. 76-138.

[4]   E. Rosch, "Human Categorization," In N. Warren (Ed.), *Advances in Cross-Cultural Psychology (Vol. 1)*. London: Academic Press, 1977.

[5]   E. Rosch, and C.B. Mervis, "Family Resemblances: Studies in the Internal Structure of Categories," *Cognitive Psychology*, 7 (1975) 573-605.

[6]   L.A. Zadeh, "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," *IEEE Transactions on Systems, Man and Cybernetics*, 3:1 (1973) 28-44.

Miles Myers is Administrative Director of the Bay Area Writing Project, School of Education, University of California, Berkeley; he has just completed an NIE study on the teaching of writing in the secondary schools.