JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 7, 589–592 (1968)

A Method for Developing Comprehensive Categories of Meaning[1]

IRIS C. ROTBERG

The Johns Hopkins University, Baltimore, Maryland 21218

This paper describes a method for developing more comprehensive categories of meaning than those found by existing procedures. The method extends Deese's procedure (1962) for classifying groups of words similar in associative meaning. It is based upon the assumption that the common free associates to a set of related stimuli are mediating responses and are part of the associative structure of the set. To determine words related to the factors in the original set, succeeding groups of free associates are readministered as stimuli. The entire matrix of words obtained in this way is then factor analyzed. Data were presented comparing (a) factors derived from the basic stimulus set with (b) the factors for the same set combined with the response given to it. Results indicated that factors for the large matrix were more tightly organized or included additional closely related words. These factors, therefore, should be more representative of factors occurring when the words are embedded in varying stimulus sets; this stability would increase the effectiveness of relating associative structure to other language data. In addition, a study of mediators given frequently by children and their links with the basic stimuli indicated relationships not evident from administration of students' responses alone.

The method reported in this paper extends Deese's procedure (1962) for classifying groups of words similar in associative meaning. Deese developed a measure of similarity in associative meaning between two stimuli based on the distribution of responses common to both stimuli. Sets of stimulus words highly related in associative meaning were constructed, and words within these sets classified by factor-analytic procedures. In Deese's method, the words are chosen because the *E* has some reason to believe they are similar in meaning. This approach demonstrates the associative distribution of the stimulus words and their factor structure. However, it provides no way for determining the other words that are highly related to the set that was chosen.

The present method is based upon the assumption that the common free associates to a set of related stimuli are mediating responses and are part of the associative structure of the set. To determine those words related to the set that was chosen, succeeding groups of free associates are readministered as stimuli. The entire matrix of words obtained in this way is then factor-analyzed, and the way in which each word in the initial list is linked with the subsequent responses to the word set is examined. The clusters obtained for any group of *S*s do not include only those words which the *E* considers relevant, but are modified by *S*s' sequential responses.

It is hypothesized that by including successive responses, more meaningful factors can be found than those derived from small stimulus sets. When a single set is administered to *S*s, groups of words within it usually share some meaning with each other, but the particular words linked vary depending on the set in which they are included. Data will be presented comparing the factors obtained for a word set with the factors for the same set combined with the free associates given to it. It is expected that factors from the large matrices will include additional closely related words,

[1] This study was supported by Grant GB-1483 from the National Science Foundation. The author wishes to thank Dr. James Deese for his valuable assistance in connection with the research project.

while other words with weak links to the factor will be eliminated. These word clusters, therefore, should be more appropriate for studying variations in meaning structure for different groups of *S*s, linguistic usage, and so forth.

Method

Word Set and Subjects. A 19-word set was used in the present research. The set included *butterfly* and words which were responses to *butterfly* in the Minnesota norms. The *S*s were 50 Johns Hopkins students from elementary psychology classes and 50 kindergarten children.[2] Students were tested by a group-written administration method, and children by an individual-oral method. For the students, responses given by two or more *S*s and which did not appear on the stimulus list were administered to the same *S*s about a week later. Again, all responses given by two or more *S*s and which were not used as stimuli in either of the two previous trials were readministered to the same *S*s. This procedure was repeated for five trials, each approximately one week apart.

Under this procedure, more than 3000 different responses were given by two or more *S*s in the five trials. Since an analysis of a matrix of this size was not possible, the most commonly occurring responses were selected. Those responses given five or more times to one stimulus in any trial were included as stimuli for the following trial in the data analysis. The resulting matrix consisted of 92 stimuli. The number of stimuli in each trial ranged from 17 to 20.

For the children, only responses given by five or more *S*s and which were not previously used as stimuli were presented as stimuli for further associations. At the completion of five trials, each child was also administered those responses which had been given five or more times by the students' group, but not by the children's group.

Data Analysis. Tables of overlap coefficients were constructed for the basic word set and for the entire stimulus set for each group. The Trial 1 matrix indicated the overlap coefficients for the initial 19-word stimulus list; the complete matrix included the overlap coefficients for the stimuli from all five trials. Factor loadings for each matrix were derived from the overlap coefficients.

Results

Factor Structure and Organization. The principal components factor analysis (Harman, 1960) was used to classify stimuli. A separate factor analysis was done for Trial 1, and for the complete matrix. Overlap coefficients were entered as correlation coefficients, with diagonals of unity.[3] A varimax criterion of rotation (Kaiser, 1958; 1959) was used with the number of factors to be rotated in each analysis equal to the number of *eigenvalues* greater than unity. Seven factors were derived from the basic stimulus set administered to students, and six from the set administered to children. Thirty-four factors were extracted from the students' complete stimulus matrix (92 variables); 29 from the children's responses to this matrix (92 variables); and 31 from the children's responses to their own mediators readministered as stimuli (96 variables). Words cited within factors have loadings ranging from .84 to .20 on that factor.[4]

Results indicate that words within Trial-1 factors are recombined with subsequent related responses to form factors in the complete matrix. Sometimes these mediating responses combine with the basic stimuli to form factors which are more inclusive than the original factors. In other instances, mediating responses are substituted for basic stimuli to produce more tightly organized clusters. For example, in the students' data, the Trial-1 factor, *blue*, *sky*, *color*, *yellow*, recombines in the complete matrix into a tightly organized factor, *sky*, *cloud*, *blue*, and into a more general one for *color* and color names; the Trial-1 factor *garden*, *flower*, *nature*, recombines into one related more specifically to *flower* and another related to *tree*; the Trial-1 factor *summer*, *spring*, *sunshine*, *nature*, develops into a factor

[2] The cooperation of Kiddie College Nursery, Ruxton, Maryland, in making children available as *S*s for this experiment is gratefully acknowledged.

[3] An unpublished program developed by Dr. Herbert Clark was used for computing overlap coefficients.

[4] Tables describing these findings are available from the author or from the American Documentation Institute. To obtain a copy, order Document No. 9968 from the Chief, Photoduplication Service, Library of Congress, Washington 25, D.C., Auxiliary Publications Project, remitting $1.25 for 35 mm microfilm or $1.25 for 6 by 8-inch photocopies.

specifying seasons and another related to autumn; and so forth.

The children's results follow a pattern similar to those for students, though specific responses and overlap coefficients differ. However, certain of the children's factors are more general than the corresponding factors for students; a similar finding was reported by Entwisle (1965), who compared associative structure for adults and children. The greater generality is evident also from the pattern of overlaps for the basic stimulus set. These stimuli are linked in associative meaning with a greater number of words by children than by students (using an arbitrary intersection coefficient $\geqslant .05$). Children do not make certain distinctions made by students; in other respects, children's factors correspond closely to those found in the students' data.

In general, children and students given the same stimuli organize them similarly. However, when children's own responses were presented as stimuli, these responses shared common associative meaning with the basic stimuli. Thus, relationships between words were found for this children's group which were not present when students' responses only were included in the matrix. For example, in the children's data, *bug* shares meaning with *bumble bee*, *sting*, *dirt*, *animal; butterfly* with *butter*, *bumble bee*, *sting; insect* with *bumble bee*, *animal*, *lion*, *sting*, *elephant*, *tiger; color* with *crayon* and *blood; summer* with *swim*, *pool*, *swimming*, *play*, *swing*, *dirt; garden* with *dirt*, *dig*, *mud*, *shovel;* and so forth.[4]

Discussion

The purpose of the present method is to develop comprehensive and stable categories of meaning. These categories should be more representative of those occurring when the words are embedded in varying stimulus sets. It is difficult to relate associative structure to other language theories when the specific meaning of a word varies depending on the list in which it is found. For example, a vague category such as the Trial-1 factor, *bees*, *yellow*, *color* and *sunshine*, is one of many possible categories depending on the word list used. There is little reason to believe that each one of these possibilities would relate meaningfully to other language data. However, when these words are given an opportunity to cluster with the responses to them, this category no longer exists. Instead, *bees* clusters with *honey* and *hive; yellow* and *color* cluster with other colors; and *sunshine* becomes linked with *rain; light*, *bright*, *dark*, *shine;* and *sun*, *moon*, *bright*. In general, these categories are more closely related than were the original factors and should cluster regardless of the list in which the words are included.

A study of mediators given frequently by children and their links with the basic stimuli provides motivational and linguistic information not evident from students' responses alone. For example, children's general associative relations between *bug*, *butterfly*, *bees* and *sting* can result in a corresponding undifferentiated anxiety. Moreover, administration of children's mediators as stimuli to adults might uncover associative links present at deep adult levels which were formed during early developmental stages.

In summary, therefore, the method provides a more or less mechanical discovery procedure for devising sets of meaningfully related words. The method permits reaching words that might not be considered in constructing original sets. The way in which these new words cluster with each other and with the basic set can be studied. These meaning units can then more effectively be correlated with word usage than can clusters derived from small stimulus sets. For example, we can determine more accurately whether words within meaning units share common linguistic contexts or are otherwise related in terms of sentence structure or semantics. Or meaning units developed for word sets of clinical interest can be made more realistic by factor analyzing the set in combination with *S*s' responses to it. In addition, if we wish to determine the ways various *S* groups link these words, we can readminister responses within groups. Meaning units for different groups can be compared not only for the basic word set, but for the set combined with each group's responses to it. Thus, we can note the way *S*s group words which are relevant to them, rather than the way they group words within a preplanned set.

References

Deese, J. On the structure of associative meaning. *Psychol. Rev.*, 1962, **69**, 161–175.

Entwisle, D. *Word associations of young children.* Baltimore: The Johns Hopkins Univer. Press, 1965.

Harman, H. H. Factor analysis. In A. Ralston, and H. S. Wilf (Eds.), *Mathematical methods for digital computers.* New York: John Wiley, 1960.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187–200.

Kaiser, H. F. Computer program for varimax rotation in factor analysis. *Educ. psychol. Measmt.*, 1959, **19**, 413–420.

(Received May 9, 1967)