The Meaning of Words

Joseph B. Kruskal *Bell Telephone Laboratories, Inc.*

How can we possibly use statistics to capture that elusive thing called "meaning"? Even granting the possibility, won't it take all the romance out of poetry and the charm out of graceful speech?

Well, we *can* study the meaning of words by the orderly methods of statistics. I shall discuss two interesting studies of this kind. But the romance of words, you will discover, is safe from science.

Practical-minded people may ask *why* we should work so hard to pin down meaning. Doesn't language, after thousands of years of natural evolution, serve well enough? Quite simply, no. Any college admissions officer will testify to the difficulty of interpreting teachers' written recommendations, and many colleges use a standard form to reduce this difficulty. Studies like those discussed here may one day improve communication among teachers, students, and colleges. They may even contribute to peace among nations.

Numbers, in other situations, have supplemented descriptive words for a long time, often very helpfully. The use of *inches, pounds,* and *degrees Fahrenheit* certainly has advantages over *fairly short, very light,* and *rather hot.* Once upon a time, *feverish* carried a meaning as vague as *charming* or *surly.* The introduction of the fever thermometer has been a boon to human health.

In subtler areas, such as musical tones, speech sounds, and intelligence, numbers have wide use today. Color provides a striking example. Many people imagine colors so subtle as to elude numerical description, yet several numerical systems enjoy routine use. (In most of them, three numbers describe each color.) (See the essay by Vance on determination of numerical color tolerances.) These systems played a vital role in developing color television and improving color photography, and have also aided the paint industry, stage makeup, and other fields. Some artists, too, find this scientific way of describing colors helpful. Nothing about this science reduces the scope of their art, nor your pleasure in looking at their paintings, seeing a color movie, or watching the sun go down.

These older examples, however, differ from the studies to be described here in one important way: the manner in which measurements are taken. Physical devices can measure height, weight, and temperature. Colors are handled by elaborate devices for mixing and dimming light, and a standard test measures intelligence. In the work described below, however, a person acts as a measuring instrument rather than as an object of study. His or her subjective impression, obtained and analyzed in a very careful way, constitutes the measurement.

PERSONALITY TRAITS

A study by Rosenberg, Nelson, and Vivekananthan (1968) deals with the meaning of 64 personality-trait words such as *impulsive, sincere, cautious, irritable,* and *happy*—words that describe people. By using a novel statistical technique the study helps to provide order where order is difficult to find.

Everyone uses these words, and they carry meaning to us all. Furthermore, they have importance. Imagine overhearing someone describe you as "humorless" or as "good-natured," and think how different these two descriptions would make you feel! When these words are used to arrange blind dates, or by teachers writing recommendations, it makes a big difference just which ones are chosen.

Nevertheless, people use these words differently. If two people both describe the same friend, or the same movie character, we would be surprised if they used *exactly* the same words. In psychiatry, the same uncertainty of meaning exists, not only for common words but also for technical terms such as *schizoid* and *autistic.* Long articles are written about what these words should or do mean.

This creates the problem. With words that are used so frequently and carry so much importance, clarification of their meaning may well prove useful—to psychiatrists perhaps, to college admissions officers possibly, to computer dating organizations, and who knows to whom else.

A DETOUR ON MULTIDIMENSIONAL SCALING

Clarify their meaning—but how? Statistics first enters here, providing an approach to the problem, as well as a method of analysis. First, however, let us take stock of the problem more clearly.

1. What do we *want*? An orderly description of personality-trait words, according to their meaning. With success, we will obtain a description as helpful as a city map, which displays roads, bus routes, parks, and so forth, according to their location.
2. What *means* do we have for getting this description? Nothing but the way people use the words. We have nothing to measure with a ruler or a thermometer; we can only observe how people use or respond to the words.

A recently developed statistical method called *multidimensional scaling* has particular value in problems like this one. The fact that Rosenberg and his coauthors were aware of this method helped stimulate their experiment. Statistics provided the approach.

What does multidimensional scaling consist of? It can best be described in three parts: what goes in, what comes out, and what happens in between. What goes in to multidimensional scaling are *similarities* or *dissimilarities* between various objects of one kind. For each pair of objects (in this case, for each pair of words from among the 64 used), a number describes how much alike, or how different, the two objects are. Many methods can provide such numbers. For example, the similarity between two colors may be provided by a person who looks at the two colors and describes, on a scale from 1 to 5, how alike they are. (If a larger number indicates greater likeness we call the numbers "similarities"; in the opposite case, "dissimilarities.") A more elaborate method was used in this study; I'll describe it later.

What comes out of multidimensional scaling is a *map* or *picture*, such as Figure 1. (For the moment ignore the lines in the figure and consider only the configuration of the points.) Briefly, the method places each object in a particular position. The map produced by scaling, though not a real map, shares many characteristics of real ones. For example, we are free to turn it and look at it from any direction we please; it has no particular direction that is truly up. Also we are free to enlarge (or diminish) the map, change the scale, so to speak, though our maps do not really have any scale. Convenience dictates how big we make the map.

What happens between the input and the output of multidimensional scaling? It is simple to say, though hard to do. We place objects on a map with the goal of having objects that are close together be very much alike and having objects that are far apart be very different. In other words, small distances should correspond to small dissimilarities, and large distances to large dissimilarities; or vice versa for similarities. In brief, the goal while constructing the map is a good relationship between the map distances and the input (dissimilarities or similarities).

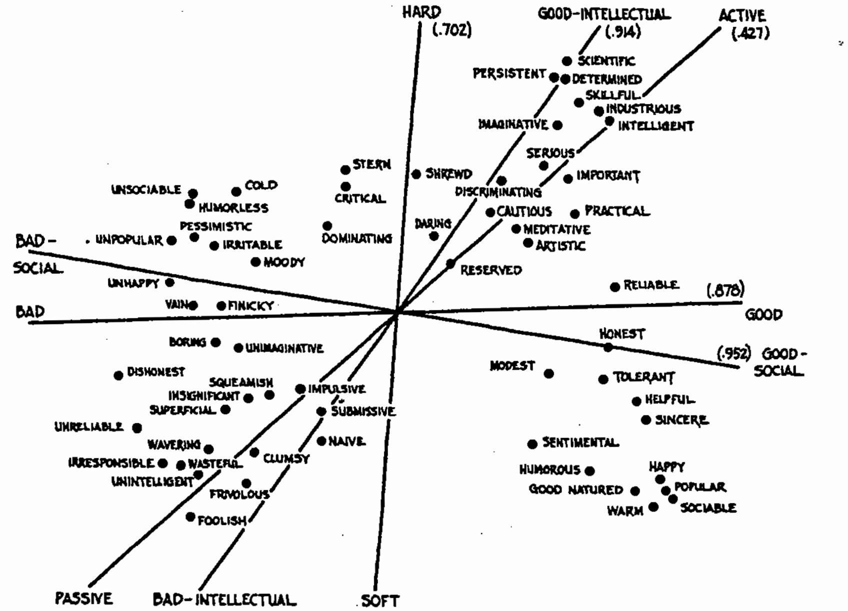

Figure 1 *Multidimensional scaling map with trend axes. Source: Rosenberg, Nelson, and Vivekananthan (1968)*

How good a relationship can be achieved depends on the input. The quality of the relationship has great importance in multidimensional scaling. Where it is good, the map may tell us something useful; where it is very bad, we may as well throw the map away.

The complex process of constructing the map is almost always carried out by computer. The user of the computer program does not need to understand this process any more than the driver of a car needs to understand how its motor works. (Today multidimensional scaling is frequently carried out on ordinary personal computers.)

PERSONALITY TRAITS RESUMED

To obtain the similarities, Rosenberg and his coauthors presented 69 subjects (college students) with slips of paper containing 64 personality-trait words. Each slip contained one word, and each subject received a complete pack of 64 slips. The subjects were asked to put the slips in their packs into roughly 10 groups, so that the words in a group could plausibly describe a single person. (A slip could be put in only one group.) Of several procedures that psychologists use to obtain similarities, this is the most rapid.

By counting how many subjects put two words together, the investigators got an *agreement score* for the two words. For example, 37 of the subjects put

reliable and ***honest*** in the same group, so these two words have an agreement score of 37. Several examples are:

1. ***Reliable*** and ***honest,*** score = 37
2. ***Clumsy*** and ***naive,*** score = 19
3. ***Sentimental*** and ***finicky,*** score = 2
4. ***Good-natured*** and ***irritable,*** score = 1
5. ***Popular*** and ***unsociable,*** score = 0

The agreement scores form a 64-by-64 square array. The 64 agreement scores of a word with itself are of no interest because all equal 69, the number of subjects. The remaining 64 × 64 − 64 = 4,032 values come in pairs of equal values because the agreement score of ***warm*** with ***intelligent,*** for example, is just the same as the agreement score of ***intelligent*** with ***warm.*** Thus only 4,032/2 = 2,016 distinct agreement scores really matter.

The agreement scores are similarities and can be used as the input for multidimensional scaling. In fact, the authors tried this in the earlier stages of their analysis, but the results were not clear. Fortunately the authors spotted the trouble: many of the agreement scores are so small that comparisons among them have little reliability. The scores given above shout that pair 2 is more alike than pair 3, but they only whisper that pair 3 is more alike than pair 4. With more than half the scores under 10, greater accuracy could help a lot. (Incidentally, this sort of trial and improvement often occurs when data are analyzed in a novel way.)

Fortunately, greater accuracy was found by using the data differently. For example, consider any two words; say, ***sentimental*** and ***finicky.*** Let's now use two rows to list their agreement scores with all 64 words:

	Foolish	*Inventive*	*Wavering*	*Submissive*	*Cold*	*Tolerant . . .*
Sentimental	2	6	5	15	3	22 . . .
Finicky	6	2	18	10	9	2 . . .

If two words have similar meanings, then not only should they have a large agreement score, but also ***their agreement scores with a third word should be about the same.*** Thus for two similar words the corresponding numbers in the two rows should not differ greatly. If we measure the difference between these two rows of numbers in some reasonable way, then this secondary (or derived) measure should indicate the dissimilarity (rather than similarity) in meaning of the two words. Following this idea, the dissimilarity between ***sentimental*** and ***finicky*** is formed by adding up the squares of the differences,

$$(2-6)^2+(6-2)^2+(5-18)^2+(15-10)^2+(3-9)^2+(22-2)^2+\ldots.$$

The larger this dissimilarity, the greater the difference between the words.

The application of multidimensional scaling to the dissimilarities[1] gave the map in Figure 1. This map must be examined, interpreted, and assessed before we can feel we have found something useful or illuminating. You will find it profitable to examine Figure 1, to see what groups of words occur together, and to see how meanings change systematically across the figure. If an investigator finds anything sensible at this point, then a preliminary judgment of success may be entered. If the map provides an orderly picture that we did not know about before, that is progress.

PERSONALITY TRAITS—A SECOND STEP

The authors found a good deal of structure in Figure 1:

> Going from the upper-right corner to the lower-left corner, the desirability of traits for intellectual activities appears to decrease systematically. Another systematic change appears to take place as one goes from the upper-left corner to the lower-right corner; in this case the social desirability of the traits increases.

Notice that the upper-right corner includes ***intelligent, skillful,*** and ***scientific,*** while the opposite corner includes ***unintelligent, foolish,*** and ***frivolous.*** The upper-left corner includes ***unsociable, humorless,*** and ***unpopular,*** while the opposite corner includes ***sociable, popular,*** and ***warm.*** A good way to indicate these trends is by axes, as drawn in Figure 1. For each trait, its position ***along*** the axis indicates where it stands with regard to that trend. (To get the position of a word along the axis we drop a perpendicular onto the axis from the plotted position of the word. It makes no difference how far from the axis the word lies.)

This interpretation does not exclude other interpretations. A given direction (such as upper right to lower left) may have several interpretations, and other directions may have meaning also. Not everybody agrees on the best way to interpret the figure.

The authors did not stop with their own subjective interpretation. Subjects different from those who provided the agreement scores rated the 64 traits on five different scales. Each subject dealt only with a single scale. For example, 34 subjects rated each of the traits on a 7-point scale "according to whether a person who exhibited each of the traits would be good or bad in his intellectual activities." A second scale dealt with social activities. Three other scales, namely, good-bad, hard-soft, and active-passive, were also used. (These were chosen from among the semantic differential scales in Osgood, Suci, and Tannenbaum, 1967.)

For each scale, the authors took several steps. They found the median of the subject ratings for each trait. (The median of the ratings is the middle value,

[1]Due to limitations (no longer present) on the computer program, only 60 words were actually included in the multidimensional scaling.

smaller than half the values and larger than half the values.) By using a statistical method called *linear regression,* they found the axis that best matches the median ratings. The five axes shown in Figure 1 resulted from this procedure. The second experiment clearly verifies the two trends described by the authors, as well as displaying some other trends. (There may still be other valid trends.)

It is instructive to check the strength of each trend. In other words, how well do the median ratings on each scale match the positions of the words along the corresponding axis? To make this comparison, we calculate the *correlation coefficient* between the median scores for the word on the scale, and their positions on the axis. This widely used statistical measurement indicates how closely two sets of numbers vary with each other. It always lies between −1 and +1, and +1 indicates perfect agreement in the way the numbers vary. The correlation coefficients, shown in Figure 1, are:

Social good-bad	.95
Intellectual good-bad	.91
Good-bad	.88
Hard-soft	.70
Active-passive	.43

These indicate that the first three scales match the map really quite well, the fourth only fairly well, and the last in a very mediocre way (although .43 is high enough to prove conclusively that there is *some* connection between this scale and the map).

In conclusion, this analysis makes a significant start toward providing a systematic explanation of the meanings of a considerable set of words for personality traits. The authors have constructed a single map that explains how subjects use and understand these words in several different tasks. Of course, the map only explains *part* of the meaning of these words. Other aspects are entirely ignored. Thus *skillful, industrious,* and *intelligent* are very close on the map because they often describe the same person, even though their meanings differ greatly in other respects. Also, the map only explains *aggregate* data, based on many subjects. This washes out and ignores individual variation in the use of these words.

Nevertheless, it is striking and illuminating that a simple map captures an important part of the meaning of a wide variety of words for personality traits. Real progress has been made.

MULTIDIMENSIONAL SCALING—A FURTHER COMMENT

The maps produced by multidimensional scaling need not be ordinary two-dimensional maps. They may be (and often are) three-dimensional, four-dimensional, or even higher-dimensional. As a matter of fact, Rosenberg and his coauthors actually constructed a one-dimensional map, a two-dimensional map (Figure 1), and a three-dimensional map, before deciding which one was the right one. (To explain how the choice was made would take us too far afield.)

I mentioned earlier that the quality of the relationship between map distances and dissimilarities has great importance. If the relationship is very bad, the map might as well be thrown away. A picture of this relationship (for the map of Figure 1) appears as the central plot of Figure 2. This plot contains (60 × 59) /2 = 1,770 points, one for each pair of words. For each pair, the dissimilarity gives the vertical coordinate of a point, and the distance between the two words (taken from Figure 1, just as you would measure it with a ruler) gives the horizontal coordinate. Clearly quite a good relationship connects dissimilarity and distance. The other two plots of Figure 2 give the corresponding pictures, based on one-dimensional and three-dimensional maps.

NATIONS

The names of nations, like other words, bring to mind many associations. Two experiments [one by Wish (1970) using 12 nations and another by Wish, Deutsch, and Biener (1970, 1972) using 21 nations] investigated how people perceive nations and their interrelationships. The most important questions addressed to the subjects concerned how *similar* various nations are. By methods somewhat like those above, and using students at Columbia University as subjects, the authors discovered several interesting facts.

Of the many characteristics of nations that influence their perceived similarity, it is not surprising that the two that emerge as most influential (in both experiments) are political alignment (ranging from "aligned with U.S.A." to "aligned with Russia") and economic development. A more complex characteristic that also displayed importance in the second experiment is "geography, race, and culture," under which the nations break down into four groups: European, Spanish, African, and Asian.

The importance of the characteristics to different subjects did yield surprises, however. In the first experiment, the 18 subjects were divided into doves, moderates, and hawks according to which of five recommendations each selected concerning the war in Vietnam. All six hawks attached greater impor-

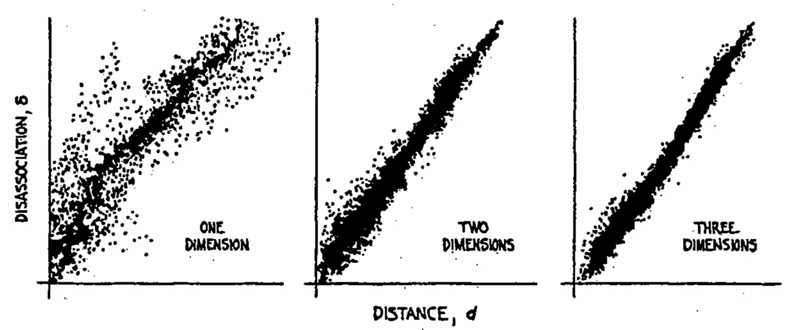

Figure 2 *Relationships between map distances and word dissimilarities. Source: Rosenberg, Nelson, and Vivekananthan (1968)*

tance (in making their similarity judgments) to political alignment, while all seven doves attached greater importance to economic development. Thus political attitude affects even a question like how similar two nations are.

In the second experiment, 75 subjects from eight countries displayed the same systematic effect. Also, females as compared with males and students from underdeveloped countries as compared with students from developed countries attached higher importance to political alignment. The students were also tested for their knowledge about the 21 nations and were then divided into better informed and less informed groups. Within each of these groups, the same effects were displayed, although much more strongly among the less informed group.

PROBLEMS

1. In what way are the measurements used in the personality-traits study different from pounds or inches?
2. What quality was being quantified in preparing the input for multidimensional scaling?
3. Suppose Rosenberg and his coauthors had been studying the meaning of 50 (rather than 64) personality-trait words. How many distinct agreement scores would have then mattered?
4. Consider a Rosenberg-type experiment using only four words with the following agreement scores:

	Faulty	*Friendly*	*Mean*	*Perfect*
Faulty	—	1	4	2
Friendly	1	—	0	4
Mean	4	0	—	1
Perfect	2	4	1	—

   What is the dissimilarity between *friendly* and *perfect*?

5. Refer to Figure 1.
   a. What five words form an extreme good-soft group?
   b. How would you characterize *stern* and *critical*?
6. Consider the locations of *meditative* and *impulsive* in Figure 1.
   a. How do the two words compare on the active/passive axis?
   b. Does the answer in (a) seem counterintuitive? Explain your answer.
7. Were the axes in Figure 1 produced by the multidimensional scaling procedure? Explain your answer.

8. The number .914 (under Good-Intellectual) in Figure 1 is a correlation coefficient. What two measures are being correlated? What can you conclude about these two measures?
9. Examine Figure 2. How is δ measured? What about d?

REFERENCES

Charles E. Osgood, George J. Suci, and Percy Tannenbaum. 1967 (1957). *The Measurement of Meaning.* Urbana, Ill.: University of Illinois.

S. Rosenberg, C. Nelson, and P. S. Vivekananthan. 1968. "A Multidimensional Approach to the Structure of Personality Impressions." *Journal of Personality and Social Psychology* 9:283–294.

Myron Wish. 1970. "Comparisons Among Multidimensional Structures of Nations Based on Different Measures of Subjective Similarity." *General Systems* 15:55–65.

Myron Wish, M. Deutsch, and L. Biener. 1970. 'Differences in Conceptual Structures of Nations: An Exploratory Study." *Journal of Personality and Social Psychology* 16:361–373.

Myron Wish, M. Deutsch, and L. Biener. 1972. "Differences in Perceived Similarity of Nations." In *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences,* A. K. Romney and S. Nerlove, eds. New York: Academic Press.