4 Extension of the feature model

Firewood

Kin terms served as an exemplary domain for investigation of the way semantic features can be related to short-term memory, analogy, polysemy, similarity ratings, and other cognitive phenomena. Necessarily, these findings depend on being able to determine the features of the terms. The techniques described in the preceding chapters for determining the features of kin terms had a long history of development. Out of years of research on kin terms came the techniques developed by Lounsbury, Goodenough, Romney, Hammel, Atkins and others which gave a high degree of precision to the feature analysis of kin terms; although there might be disputes about certain complex configurations, in general a high degree of consensus was obtained in the analysis of kin term systems.

Kin terms, however, are only one small terminological domain in the lexicon of any language. What about all the other domains which did not have the paradigmatic character of kinship terminologies? *The question quickly arose as to whether the same kind of feature model could be applied to other domains of cultural meaning.* The feature model might be quite important theoretically, but if it could only be applied to kin terms it would be of very limited use. Most of these other domains, however, had no long history of semantic analysis. To get at the feature structure of these new domains new methods had to be developed.

One of the early anthropological studies to try explicitly to identify features outside of the domain of kin terms was carried out by Duane Metzger and Gerald Williams (1966). Working in Tzeltal, a Mayan language spoken in the highlands of Chiapas, Mexico, they attempted to analyze the salient features of *firewood*. Since there has been criticism of Metzger and Williams for concentrating on a domain which seems to have little political or social importance, it is worth noting that whatever firewood means to us, it is not a trivial topic for Tzeltal speaking horticulturalists. Much of the land has been stripped of timber, resulting in a daily arduous search for firewood. Knowledge of the kinds of timber which make good firewood is a necessary part of any adult Tzeltal

speaker's stock of knowledge. To criticize Metzger and Williams for studying firewood rather than things which seem more relevant to us is itself a kind of ethnocentrism in which *our* concerns about their lives are given much greater weight than *their* concerns about their lives.

Since the techniques for the analysis of kin terms in which the kin types for each term are used to find contrasting features have no clear analog in studying how people categorize firewood, Metzger and Williams developed a new set of techniques to uncover features. What they did was to adapt a kind of structural approach to the analysis of linguistic forms which utilizes *frames* and *slots*. This approach had been part of what was then called *structural linguistics* developed by Bloomfield and Zelig Harris, although the root ideas go back at least to Saussure and the Prague school. Basically, the idea is that one can get at the meaning of items through the way items are distributed in different environments.[1] The environment, or frame, consists of a well-formed series of items – say a phrase like "___ *is a kind of fruit*" which consists of a syntactically well-formed series of words which make a true sentence when filled in by the right kinds of items. Thus "*A pear* is a kind of fruit" makes a true sentence, while "*A flowerpot* is a kind of fruit" makes a false sentence.

Although the idea of *frame* and *slot* were basic to Metzger and Williams' approach, they found that it was often possible to transform frames into questions. Thus the frame "___ is a kind of fruit" can be transformed into a direct query "What kinds of fruits are there?" and obtain the same kinds of responses. Metzger and Williams say:

> Of basic importance in the formulation of frames is the use of bilingual informants who assist us in formulating relevant questions in the native language, the central problem being one of finding out from informants what things there are to be asked about, what relevant things may be asked about them, and what are the significant answers to be anticipated. (1966:390)

In the firewood example, Metzger and Williams began with the question of whether or not there is a category similar to the English category of *firewood*. They found that there was such a category, which they established by the following series of queries:

Q: How are they named, the things of mother earth in all the world?
A: There are many kinds.

Q: What is the name of the first kind?
A: There are people.

Q: What is the name of the second kind?
A: There are animals.

[1] The idea that meaning could be uncovered by analyzing word distribution was an ideological position in the late forties and early fifties. This position collapsed with the advent of Chomsky and the generative-transformational revolution.

Table 4.1. *Evaluation of firewood*

Good	Poor
Hard wood	Soft wood
Burns strongly	Burns quickly
Dries rapidly	dries slowly
Its fire is hot	Its fire is only a little hot

Q: What is the name of (another of) a third kind?
A: There are "trees-and-plants".

Q: What is the name of (another of) a fourth kind.
A: There are no more (of a fourth different kind). (1966:391[2])

The same queries can be used to find out the kinds of people in the world, or the kinds of animals, or the kinds of trees-and-plants. When asked to name the the kinds trees-and-plants, informants respond with the Tzeltal terms for trees, shrubs, grasses, and vines. When asked what are the name of the kinds of trees, informants give a long list of trees including oaks. Repeating the same process for just oaks, informants are able to produce a list of a number of specific kinds of oak.

This kind of repeated query technique makes a highly formalized way of eliciting a taxonomy. Of course, such a taxonomy could be elicited in a more informal manner, but a part of Metzger and Williams' goal was to develop a process which was *interpretation free* – what they got was what you saw.

Next Metzger and Williams developed queries which asked for the *uses* for the items they have elicited. With respect to trees some of the uses are as wood for houses, axe handles, benches, etc., and finally firewood. About firewood, Tzeltal speakers say that "it serves us (in that) we put it in the fire." Using this somewhat lengthy procedure, Metzger and Williams are able to show where the category of "firewood" fits into the total Tzeltal classification of things – "firewood" is the use of certain trees which along with "people" and "animals" constitute the major kinds of living things.

Having established firewood as a natural category among the Tzeltal, Metzger and Williams go on to use the same kind of frame/query approach to discover the salient features of firewood. Informants know in detail which kinds of trees make the best and worst kinds of firewood, and readily produce such lists if presented with the appropriate query. Next Metzger and Williams develop two queries to uncover the features that make for good or bad firewood, which are glossed "Why is ___ good (as firewood)" and "Why is ___ not good (as firewood)." Responses to these frames produce four major dimensions (see Table 4.1).

These qualities of hardness, burning quality, drying quality, and heat are the

[2] In the original article both Tzeltal phrases and their English glosses are given.

Table 4.2. *Consequences of burning certain varieties of wood*

Because they say that __	if we burn __ in the fire
	our own bones will be burned
	our chickens will die
	we will be seized by insanity
	we will not have children
	our children will get epilepsy
	a snake will come to our house

major properties used in the evaluation of "firewood." However, for a few special kinds of wood the reasons given for why the wood is not to be burnt are quite different (see Table 4.2):

It appears that the burning of firewood has more to it than the ordinary pragmatics of heat production. In their article Metzger and Williams go on to discuss how firewood is prepared, cut into different size lengths, how the lengths are measured (in axe handles), how firewood relates to kindling and to charcoal. For each of these points they discuss the formulation of appropriate queries and the kind of responses made by informants.

The firewood example presented here illustrates well the general query technique developed by Metzger and Williams. Using the same set of techniques, they worked out a precise ethnographic description of the role of the curer in Tenejapa (a Tzeltal speaking municipio) (1963b), and a detailed description of the wedding ritual used by Tenejapa Ladinos (1963a). M. Black and Metzger presented a programmatic study of law, using as informants a second year Stanford law school student and a number of Tzeltal speaking Tenejapans (1965). Black also produced a detailed Ojibwa taxonomy of "living things" (1987).

The major impact of the work of Metzger, Williams, and Black in the mid-1960s was methodological – the basic thesis was that informant knowledge could be obtained in a reliable and replicable way, without imposing the ethnographer's conceptual framework on the elicitation procedure. Everything was put out in the open. The reader was confronted with the ethnographer's data. The attempt was to find out what the informant thinks in a highly *explicit* and *formalized* manner. In all of this, psychological considerations played a negligible part; the major goal was straightforward ethnographic description. Metzger and his associates were interested in cultural knowledge – something which they often found to be detailed, systematic, and sometimes surprisingly different than one might have expected from the perspective of American culture.

While the techniques of Metzger and his associates did not become a part of standard anthropological ethnographic technique, some of the general ideas about frames and queries were adopted by other anthropologists. For example, James Spradley's book *The Ethnographic Interview* uses the query approach to

teach students how to do ethnographic research, as does Michael Agar's *The Professional Stranger.*

What did emerge most clearly from this work was the idea that in any culture there were certain appropriate questions which could be asked about a particular cultural category, and these question elicited kinds of *relationships* which informants understood to hold between cultural categories. For example, in the firewood paper, Metzger and Williams developed queries which asked about the "kind of" relationship, "use" relationships, and "evaluative attributes" of goodness and badness.

Semantic networks

One important effect of Metzger and Williams' work was to open up the general problem of understanding native thought from the analysis of simple features to an analysis of the relationships between categories. As Frake said in his paper "Notes on Queries in Ethnography":

> By presenting an inquiry (e.g. "What kind of tree is that?"), a native inquirer seeks to restrict the appropriate responses to a given set of responses (e.g., tree names), so that the selection of a particular response . . . conveys information significant to the inquirer. A description organized by linked queries and responses is simultaneously a program for finding out information, a program which can be replicated and tested by the reader of the description . . . Furthermore, the topic of a given query will be a response to some other query, making it possible to produce lists of utterances *interlinked* as topics and responses of specified queries. A pair of queries which interlink utterances as mutual topics and responses constitute the basic unit of these procedures, a unit we will call an *interlinkage*. (Italics in the original) (1964:134)

The ethnographic task that Frake uses to illustrate the way in which cultural knowledge can be modeled as an interlinked set of categories is the process of manufacturing *gasi*, a fermented drink used by the Eastern Subanun of the Philippines. In manufacturing *gasi* a starchy mash is reduced to sugar by certain fungi grown with yeast in cakes of rice flower. The fungi reduce the starch to sugars upon which the yeast can then act to bring about fermentation. Frake focuses his description of the yeast cakes and the selection of "spices" to add to the other yeast ingredients.

In trying to work out how the different items of "spices," "yeast," and "beer" (*gasi*) are interrelated, Frake found that he needed a series of linked queries. These queries were:

use: What is __ used for?
kind: What kind of __ is it?
what: What is __ (a kind of)?
ingredient: What is that ingredient of X?
part: What (separated) part of __ is it?
source: What does __ come from?

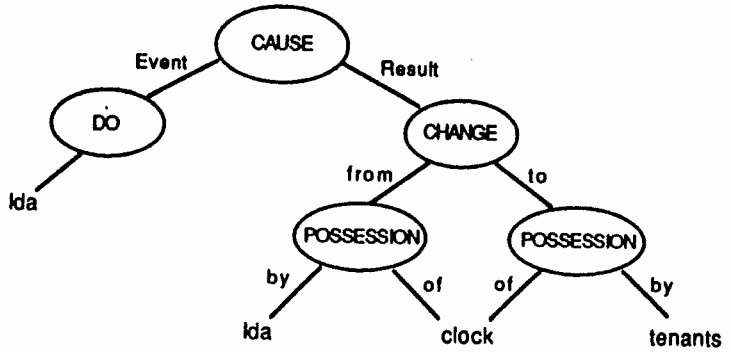

Figure 4.1 Example semantic network model (adapted from Genter 1978)

Frake points out that some of these queries are reciprocals of each other. Thus "yeast" is *used for* making "beer" and "beer" has an *ingredient* which is "yeast." Frake then uses these queries to analyze the relationship between the different items that go into making *gasi*. In carrying out this analysis, he found that some terms had a number of different senses, and that these different senses could be demonstrated by the kinds of relations each sense has to other items.

A similar semantic network model was developed in artificial intelligence by Ross Quillian, Allan Collins, and others in the late 1960s. In Quillian-like models, nodes of the network stood for objects, while lines stood for relationships of various kinds, especially subset/superset relations. This type of model was used in psychology primarily as a representation of the structure of long-term memory, and served as the postulated structure for a number of cognitive processing experiments (Collins and Quillian 1969). A modified version of this network model was developed by the LNR group at UCSD to represent complex word meanings (Peter Lindsay, Donald Norman, and David Rumelhart). Figure 4.1 presents a diagram of the feature structure of sentence "Ida gave her tenants a clock" using the LNR representational format (see Figure 4.1).

In anthropology, Oswald Werner extensively developed the use of semantic network models. In his book *Systematic Fieldwork*, written with G. Mark Schoepfle, the authors lay out a step by step method of doing ethnography using a small number of basic relationships as the building blocks for constructing networks. One of Werner's most extensive examples of a network analysis is *The Anatomical Atlas of the Navajo*, written in collaboration with K. Y. Begishe, M. A. Austin-Garrison, and J. Werner.

As a method, semantic network analysis has the capacity of creating an encyclopedia-like description of a culture. Basically, the networks consist of linked propositions. The problem with such networks is that everything is connected directly or indirectly to everything else, so there is no clear way of grasping

what is important and what is peripheral. Instead of having a small number of features which reveal what is most salient to the native, one ends up with a great encyclopedia in which everything is as important as everything else. Of course, if you want to find out particular information about some cultural object, a good encyclopedia is the place to go.

Finding salient features through similarity judgments

In Chapter 3 the point was made that similarity judgments reflect the sharing of features. Terms which share many features will generally be judged to be more similar than terms which share few features. This "effect of shared features" was used in Chapter 3 to decide between alternative feature analyses of kin terms. It is also possible to use the "effect of shared features" on similarity judgments in cases in which no prior analysis has been made. To take a simple example, let us consider color terms. Without any prior analysis of the feature dimensions of color terms, it is possible to give a number of respondents a set of color terms and have them judge how similar every pair of terms is. Using the mean similarity judgment for each pair for the sample of respondents, it is then possible to analyze these scores to discover something about the structure of these terms.

One of the most widely used multidimensional statistical programs is KYST. The KYST program is the joint effort of a number of people, including Joseph Kruskal, Forrest Young, Roger Shepard, and Warren Torgerson.[3] What the KYST program does is to try to place objects in a spatial configuration in which the computer generated distances between objects corresponds as closely as possible to the original similarity scores between items. That is, objects which are similar to each other should be close together, while objects that are dissimilar should be far apart. For example, if one inputs the distances between various cities in the the United States to the KYST program and asks for a two dimensional output, the result will be a plot which looks just like the placement of cities on a map of the US. However, when the data is noisy, or when the similarities come from a cognitive structure which has many dimensions, the program will still produce, if asked, a two dimensional output, but the distances shown on the output will no longer correspond as closely to the similarity scores. The program gives a figure, called "stress," which indicates how closely the original input scores correspond to the output distances.

To return to the analysis of color terms: in a study by Samuel Fillenbaum and Amnon Rapoport (1971) in which twenty-six English speaking respondents ranked all possible pairs of fifteen color terms by degree of dissimilarity, the resulting output is presented in Figure 4.2. The stress for two dimensions was low; 0.085, which means the fit between the input scores and the output

[3] For a general introduction to multidimensional scaling, see Kruskal and Wish 1978.

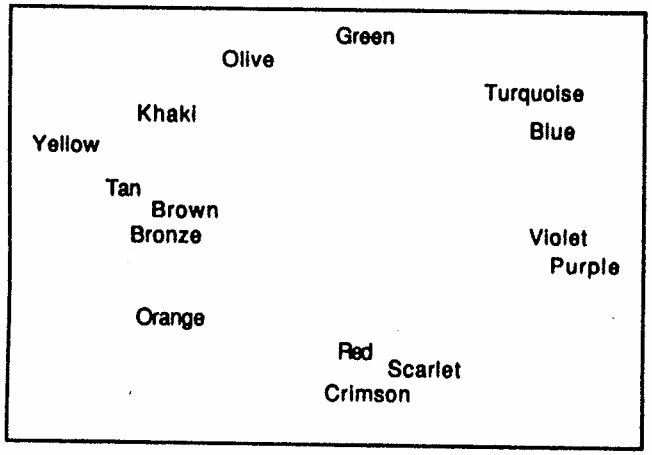

Figure 4.2 Two dimensional KYST representation of color terms (Adapted from Fillenbaum and Rapoport 1971)

distances is very good. This contrasts with what happens if the program tries to order the scores along a single dimension (instead of two dimensions); then the stress score rises to 0.339. The overall result corresponds closely to the normal image of the color wheel, with *red* and *green* on one axis and *blue* and *yellow* on the other axis.

In this color term example the *red/green* dimension and the *yellow/blue* dimension are not simple dichotomous values like *male/female* or *direct/collateral* in the analysis of kin terms. Instead, terms can have a graded position along a dimension. *Orange*, for example, is positioned towards the *red* end of the *red/green* dimension, but not as far out as *red* itself. *Orange* is also on the *yellow* side of the *yellow/blue* dimension, but again not as far out as *yellow* itself. One of the advantages of multidimensional scaling analyses is that they have the capacity of representing fine gradations along a continuum.

Given the ease and effectiveness of multidimensional scaling analysis, it is not surprising that this technique has been used frequently. Anthropologists, psychologists, and sociologists have all used multidimensional scaling of similarity ratings to try to discover the major dimensions underlying a variety of domains. Some of the domains which have been analyzed include American occupational terms (Burton 1972), American-English personality trait terms (Rosenberg and Sedlak 1972), nation states (Wish, Deutsch, and Biener 1972), American concepts of success and failure (Romney, Smith, Freeman, Kagan, and Klein 1979), Maasi personality terms (Kirk and Burton 1977), A'ara personality terms (White 1980), American interpersonal relationships (Wish 1976), interpersonal behavior traits (White 1980), fish shapes (Boster and Johnson 1989), and emotion terms (Russell, Lewicka, and Nilt 1989).

One can think of multidimensional scaling as a kind of sophisticated feature extractor. However, multidimensional scaling techniques generally do not get at the fine detail of minor feature differences. What these techniques usually do is to find a small number of the highly salient dimensions that apply to the majority of terms. The result is a smoothed out mapping of items in which general similarity relations are preserved but specific differences are lost. For example, in Figure 4.2, the color terms *bronze*, *tan*, and *brown* are all grouped together between *yellow* and *orange*. The differences between these three terms – the metallic sheen of *bronze* and the lighter shade of *tan* compared to *brown* – are lost in the two dimensional representation.

While at first the major use of multidimensional scaling was simply to explore the salient feature characteristics of different cultural domains, as the field became more sophisticated investigators began to use multidimensional scaling to test specific hypotheses. Kirk and Burton (1977), for example, were able to show that the salient features of role terms changed for Maasi males as they went through culturally marked life stages. Boster and Johnson (1989) were interested in comparing the knowledge of expert and novice fishermen about fish, and used multidimensional scaling to show these differences. Other studies, such as Russell, Lewicka, and Nilt (1989) and White (1980) were aimed at uncovering cross-cultural universals in the salient feature characteristics of emotion terms and interpersonal behavior traits.

Psychological reality again

It is interesting that some of the same questions that arose concerning feature analyses of kin terms arise again with respect to the results of multidimensional scaling. That is, given certain scaling results, did these results have any psychological validity or were they just some kind of computer output with no real psychological implications? Or, to put the question another way, what difference did it make that respondents rated some things as more similar than other things? After all, the multidimensional scaling results are just a summarization of a large number of similarity judgments, and cannot be expected to have any more psychological meaning than the original judgments.

Several kinds of experiments were undertaken to answer this question. Using a multidimensional scaling analysis of similarity ratings for common animals (*cat*, *dog*, *rabbit*, *monkey*, *lion*, *giraffe*, etc.) Rumelhart and Abrahamson (1973) developed a task using analogical reasoning to see if scaling results could predict respondent choices. The scaling results for common animals, carried out by Henley (1969), had an interpretable two dimensional solution. Figure 4.3 presents the two dimensional Henley analysis. The horizontal dimension is clearly *size*, with *elephant* on the far right and and the other animals in approximate order of size ending with *mouse* on the far left. The vertical dimension is *ferocity*, with *leopard*, *tiger*, *and lion* on the bottom of the

diagram, and *cow* and *sheep* at the top of the diagram. For this test Rumelhart and Abrahamson first chose two animals, and then measured the direction and distance between these two animals. They then chose a third animal, and drew a line from the third animal in the same direction and distance as the line between the first two animals. Whatever animal came closest to the end of this line should be the best choice for an analogy. They gave respondents the following type of test:

CAT: LEOPARD :: MONKEY : ____
A. ANTELOPE
B. BEAVER
C. GORILLA
D. TIGER

Here the most appropriate answer is *GORILLA*, which is approximately the same distance and direction from *LEOPARD* as *MONKEY* is from *CAT*. The results of a series of such tasks supported Rumelhart and Abrahamson's hypothesis; a mean of 70% of the respondents gave the predicted animal as their first choice. It seems likely that respondents use the same kind of analogy searching procedure for this task described in Chapter 3. Given the A:B :: C: __ format, if the second (B) animal of the first pair is somewhat larger and somewhat more ferocious than the first (A), then select as an answer an animal which is equally larger and equally more ferocious than the first animal of the second pair (C).

Another simple test of the psychological validity of multidimensional scaling results for similarity judgments was developed by A. K. Romney (1989). The task consists of asking respondents to list all the common animals they can think of. There is a general finding in psychology that when respondents are asked to list items in free recall, the stronger the association between the items, the more likely they will be recalled together (Bousefield 1953). Using a statistic for the analysis of path lengths, it has been found that respondents are much more likely to list together animals that are close together in the scaling analysis. For example, a respondent might produce a list like *deer*, *sheep*, *cows*, *goats*, and *horses*, but would be very unlikely to produce a list like *gorilla*, *dog*, *camel*, *rat*, and *elephant*. This "clustering in recall" effect is quite robust and replicable (Romney 1989).

Another task to test the psychological implications of scaling of similarity judgments was developed by Nerlove and Romney (1976) using reaction time in a forced choice format. Respondents were shown a series of 70 slides with three animal terms in the following format:

HORSE

SHEEP ZEBRA

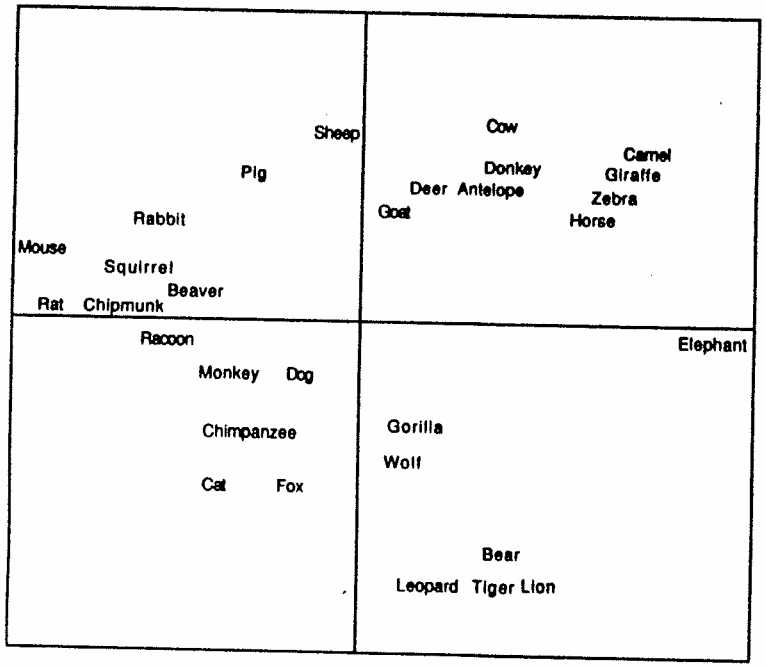

Figure 4.3 Two dimensional representation of judged similarity among common animals (adapted from Henley 1969)

The task for the respondent was to choose among the bottom two terms the term which is most like the top term. Respondents were told to make their choices as quickly as possible, and the response time it took the respondent to answer was recorded. In the example above *zebra* was chosen with a mean reaction time of 2.98 seconds. However, given the following triad

CAMEL

TIGER LION

the mean reaction time was much longer – 4.15 seconds. In their analysis of the data Nerlove and Romney found that two factors had a major effect on reaction time. First, respondents were quicker if the distance between the left and right comparison animals was large (e.g., the distance between *zebra* and *sheep* is much larger than the distance between *tiger* and *lion*). Second, respondents were quicker if there was a large difference between the distance of the top animal to one of the comparison animals versus the distance to the other comparison animal (e.g., *horse* is much closer to *zebra* than to *sheep*, but *tiger* and *lion* are just about the same distance from *camel*). Similar results for the domains of birds and fruits have been obtained by Hutchinson and Lockhead (1977).

In summary, multidimensional scaling of similarity ratings has certain advantages over traditional semantic feature analysis. Traditional semantic feature analysis works best when a few features partition a relatively large number of terms, as in a complete paradigm. But most domains have much more loosely defined contrast sets than kin terms and pronouns. For example, there are a huge number of features, physical and behavioral, which differentiate a *camel* from a *tiger* or a *pig* from a *gorilla*. Similarity judgments present the respondent with the task of finding those features which are most general and most salient in structuring a domain, since the respondent has to consider the ways in which all the different pairs of items in a domain are alike as well as they ways they are different.

Thus scaling analyses of similarity judgments tend to result in the identification of a few salient features which structure an entire domain. However, there is also a drawback here. In some analyses, the feature dimensions are relatively obvious. In the Henley animal analysis, for example, the dimensions of *size* and *ferocity* account nicely for the placement of most of the animals. However, in many analyses it is not obvious which features account for the way in which items are placed in the space. There have been a number of arguments, for example, about the dimensions which structure the domain of personality trait terms. Various techniques, such as having ratings of potential dimensions made on all terms, then finding the vector which correlates most highly with the dimensional ratings, can be useful in helping to select feature dimensions.[4]

Another problem involved in scaling analyses of similarity judgments is the fact that the number of pairs of items increases geometrically with the number of items (for n items, there are n * (n-1) / 2 pairs of items). Thus 10 items generate only 45 pairs, while 50 items generate 1,225 pairs. If the task consists of judging on a similarity scale every pair of items, large domains require too many judgments to be practical. The triads test – the easiest similarity task to administer to non-literate respondents – is even more time consuming than using all possible pairs, since the number of triads increases with the number of items even more rapidly than the number of pairs (for n items, there are n * (n-1) * (n-2) / 6 triads). There are ways around this problem using incomplete balanced block designs,[5] but there is always a concomitant loss of information in using such designs, as well as the difficulty of finding the right design. Another technique is to use a *pile sort* method, in which a respondent is given a stack of cards with one of the items written on each card, and then asked to sort the items into piles on the basis of similarity. One measure of similarity derived from the pile sorting technique is simply to count the number of times respondents place pairs of items in the same pile; the more times any pair of

[4] For an example of this technique, see Rosenberg and Sedlak 1972.

[5] In an incomplete balanced block design only a special selection of pairs or triads are used. Combinations are selected so that all items occur the same number of times. See, for example, Burton and Nerlove 1976.

items is placed in the same pile, the more similar the two items. It has been found that obtaining reliable measures with the pile sorting technique requires a relatively large number of respondents; fifty respondents is usually a minimum. Respondents can do this task with as many as a hundred items, although pile sorting a hundred items is a daunting task. A number of techniques for collecting similarity judgments are described and illustrated in Susan Weller and A. K. Romney's book *Systematic Data Collection*.

Item by feature matrices

So far we have dealt exclusively with multidimensional analyses of *similarity* judgments. However, it is possible to use multidimensional scaling techniques to analyze other kinds of data. For example, consider beliefs about illness. A normal American knows a number of disease terms – *colds*, *mumps*, *cancer*, *ulcers*, etc. A normal American also knows a number of things about these diseases – that some *bring on fever*, that some *can be caught from other people*, etc. The things that Americans know about diseases can be called *disease features*. Given a list of commonly known disease terms and a list of commonly known disease features, respondents can be asked to judge which features go with which disease terms. The result is an *item by attribute matrix* in which both disease features and disease terms can be analyzed together.[6]

A study by D'Andrade, Quinn, Nerlove, and Romney (1972) used item by attribute matrices to attempt to investigate standard American and Ladino Mexican categorizations of disease. We turned to this type of analysis because attempts to construct standard taxonomies of disease terms resulted in shallow, non-exclusive structures (i.e., the same diseases were said to belong to more than one superordinate category), while attempts to carry out standard semantic feature analyses resulted in confused attempts on the part of our informants to use the technical language of medicine.

The first step of the study was to collect a suitable list of disease terms. Lists of common diseases were collected and then a small number of informants asked to judge which terms were most common and least ambiguous. Table 4.3 presents a list of the final selection of disease terms.

To collect lists of features relevant to disease states a series of statements made by informants about diseases were collected from informal ethnographic interviews about beliefs about illness. These statements were then transformed into belief-frames. For example, a common statement was "you can catch colds from other people." This was transformed into the belief-frame "you can catch ___ from other people."

[6] Some of the earliest analyses of item by property matrices were carried out by Volney Stefflre in marketing studies; see for example, Stefflre 1972.

Table 4.3. *American disease terms*

1. Appendicitis	2. Bronchitis	3. Cancer
4. Chicken pox	5. Colds	6. Dental cavities
7. Epilepsy	8. Gonorrhea	9. Heart attack
10. Influenza	11. Laryngitis	12. Leukemia
13. Malaria	14. Measles	15. Mononucleosis
16. Mumps	17. Pneumonia	18. Poison ivy
19. Polio	20. Psychosis	21. Rheumatism
22. Smallpox	23. Strep throat	24. Stroke
25. Syphilis	26. Tonsilitis	27. Tuberculosis
28. Typhoid fever	29. Ulcers	30. Whooping cough

Table 4.4. *American disease belief-frames*

1. You can catch ___ from other people.
2. ___ is caused by germs.
3. Most people catch ___ in bad weather.
4. ___ comes from being emotionally upset.
5. ___ runs in the family.
6. When you are overtired, your resistance to ___ is lowered.
7. ___ can not be cured.
8. ___ has to run its course.
9. ___ should be treated by miracle drugs.
10. ___ gets better by itself.
11. ___ is serious.
12. ___ is a fatal disease.
13. You never really get over ___.
14. ___ is a crippling disease.
15. You can have ___ and not know it.
16. ___ spreads through your whole system.
17. ___ is contagious.
18. If a woman comes down with ___ during her pregnancy it harms her child.
19. Feeling generally run-down is a sign of ___.
20. ___ affects the heart.
21. Your skin breaks out with ___.
22. Runny nose is a sign of ___.
23. Sore throat comes with ___.
24. ___ brings on fever.
25. Once you've had ___ you can't get it again.
26. ___ is a children's disease.
27. Most people get ___ at some time or other.
28. Some people have a tendency to get ___.
29. It is safer to have ___ as a child and get it over with.
30. ___ is a sign of old age.

In both Mexico and the United States, the interviews with informants produced large numbers of sentences about diseases. Once a large corpus of belief-frames had been constructed, informants were asked to select the most basic, unambiguous, and general frames for describing diseases. Table 4.4 presents the final selection of American disease belief-frames.

The American sample consisted of ten monolingual English speaking Stanford undergraduates. In a pretest it was found that American informants preferred to answer on a scale rather than give simple "yes–no" answers. A five point scale was developed for the final test form and administered using the following format:

You can catch ___ from other people.

	definitely	probably	possibly	probably not	definitely not
1. Appendicitis :	_____ :	_____ :	_____ :	_____ :	_____ :
2. Bronchitis :	_____ :	_____ :	_____ :	_____ :	_____ :
etc.					

The Mexican informants preferred a simpler scale, with responses restricted to "yes" (*si*), "at times" (*a veces si, a veces no*), or "no" (*no*) The Mexican disease terms were selected the same way the terms for the American sample were selected. However, for the disease frames a small number of the American frames were translated into Spanish to facilitate comparisons between the two systems despite the semantic awkwardness of some of the translations.

Each American informant was asked to fill in an answer for all 30 disease terms for all 30 belief-frames, making a total of 900 judgments. The scale responses were averaged across respondents, and the mean for each belief-frame across the 30 diseases calculated. To simplify the data the responses were dichotomized by using the mean score for each belief-frame as the basis for deciding whether a score was to be considered a *yes* or a *no*. The results for the American sample are presented in Table 4.5. Each column of the matrix gives the dichotomized "yes" (x) or "no" (a blank) responses for a particular belief-frame, and each row of the matrix gives the responses for a particular disease. The rows and columns have been grouped on the basis of separate cluster analyses of disease terms and belief-frames.[7]

A special method of multidimensional scaling was used to analyze these data. The scaling program used, MDSCAL, was an early version of KYST. Using a special "missing data" option, it was possible to put into the program simultaneously both the belief-frame by belief-frame Pearson correlation matrix and the disease term by disease term Pearson correlation matrix. The result produced a representation in which both the disease terms and the belief are mapped into the same metric space. A two dimensional representation was found to fit the original data reasonably well, and to have a clear interpretation. The MDSCAL plot for both disease terms and belief-frames is presented in Figure 4.4. The abbreviations for belief-frames are in capital letters while disease terms are in small letters. Interpretation of this figure is fairly straightforward. There is a clear *seriousness* dimension running from right to left, with the most serious diseases and relevant belief-frames on the right hand

[7] The cluster analyses used the "average linkage" method – see Aldenderfer and Blashfield 1984:40.

Table 4.5. *American sample, dichotomized data for disease terms and belief-frames*

		WTFINR ehetoe arvsss toeeei harl s rt f t 3 23 24 10 22 6	CCGD aoer tnru ctmg hass g 1 17 2 9	OSCM nahc cfis eelt cd hr 25 29 26 27	SRPC puro rneu edgr aons dwae 16 19 18 8	SFHS eaek rtai iarn olt u 11 12 20 21	INNCTOEN neorelmo hvcindot eeupdatk rrrpegnn ieineio 5 13 7 14 28 30 4 15
Tonsilitis	26	x x x x	x	x x x x			x
Appendicitis	1	x		x x x		x x	x
Poison ivy	18	x	x x		x		
Strep throat	23	x x x x x	x x x x x	x x x	x x x		
Laryngitis	11	x x x x x	x x x x x	x x x	x x x		x
Whooping cough	30	x x x x x x	x x x x x	x x x	x		
Influenza	10	x x x x x x	x x x x x	x	x x x x		x
Pneumonia	17	x x x x x x	x x x x x	x	x x x x	x	
Mononucleosis	15	x x x x x x	x x x x x		x x x x	x x	
Colds	5	x x x x x x	x x x x	x	x x x x	x	
Bronchitis	2	x x x x x x	x x x x	x	x x	x	
Chicken pox	4	x x x	x x x	x x x	x x	x	
Measles	14	x x x x	x x x x	x x x x	x x x		
Mumps	16	x x x	x x x x	x x x x	x	x x	
Smallpox	22	x x	x x x x	x x x x	x x x		
Typhoid fever	28	x x	x x x x	x x x x	x x	x x	
Gonorrhea	8	x	x x x x	x x	x	x	x
Syphilis	25	x	x x x x	x x	x x x	x x	x x
Polio	19		x x x x	x	x x x	x x x	x
Leukemia	12		x x x	x x	x x x	x x x	x x
Tuberculosis	27	x	x x x x	x x	x x x x	x x x	x x x
Malaria	13	x	x x x	x	x x x	x x x	x x

Table 4.5. *contd*

		WTFINR eheteoe arvsssi toeeeei harlsf rtft 3 23 24 10 22 6	CCGD aoer tnru ctmg hass g 1 17 2 9	OSCM naho cfis eelt cd hr 25 29 26 27	SRPC puro rneu edgr aoms dwae 16 19 18 8	SFHS eaek rtai iarn olt u 11 12 20 21	INNCTOEN neorelmo hvcindot eeupdatk rrrpegnn i e l n e l 5 13 7 14 28 30 4 15
Stroke	24				x x x		x x x x x x x x
Heart attack	9	x			x x	x x x	x x x x x x x
Cancer	3	x			x x x	x x x x	x x x x x x x
Psychosis	20	x				x	x x x x x x x x
Ulcers	29						x x x x x x
Epilepsy	7			x		x	x x x x x x
Dental cavities	6			x			x x x x
Rheumatism	21	x	x		x x x		x x x x x x

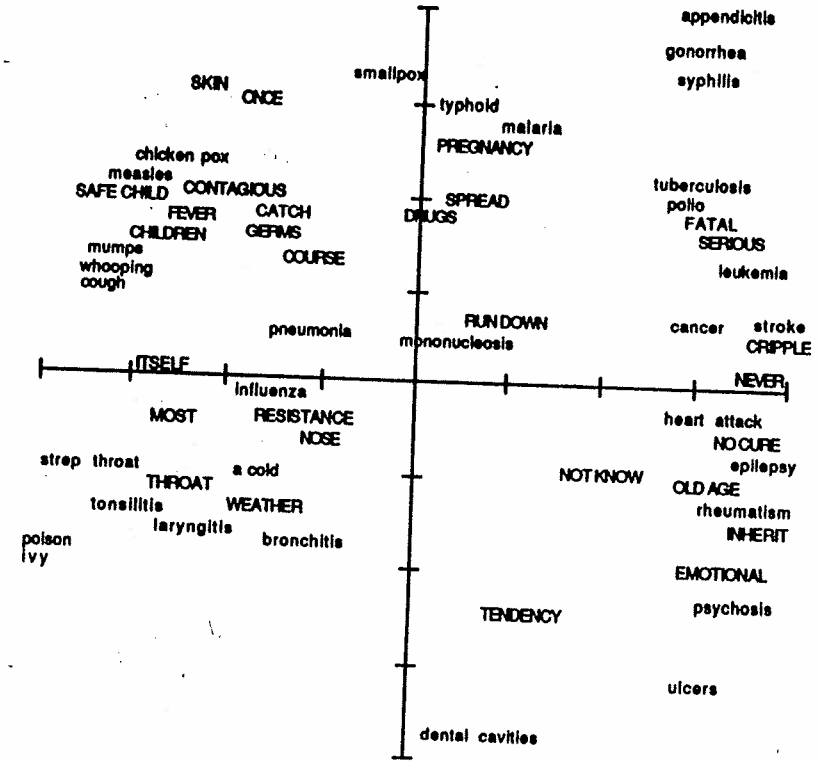

Figure 4.4 MDSCAL plot for American belief-frames and disease terms

side of the space and the minor and self-limiting illnesses and relevant belief-frames on the left.

The second dimension runs from the top left of the plot to the bottom right, and concerns *contagion*. At the bottom right are the old age, inherited, or emotion caused illness and belief-frames. The entire left side of the plot (with the exception of *poison ivy*) contains the simpler contagious diseases, such as *colds*, *influenza*, *chicken pox*, etc., along with belief-frames concerning the throat, resistance, germs, and diseases which are safer to have as a child and can only be gotten once. In the top right (with the exception of *appendicitis*) there are the serious systemic contagious diseases such as *syphilis*, *typhoid*, and *malaria*. At the far right are the crippling, non-contagious and often incurable diseases, such as *cancer*, *heart attack*, and *stroke*.

The Mexican data showed quite a different configuration than the American data. The MDSCAL results show a *hot–cold* dimension, marked most clearly by the kind of medicine one should take. For example, one should take a hot medicine when the disease involves getting wet, having a cold body or chills,

and pains in the bones that go with illnesses which involve internal congestive, respiratory, and pulmonary complaints, as well as when there are problems with the milk of a nursing mother. The second dimension has the appearance of a *contagion* dimension with *measles*, *smallpox*, *colds*, and *whooping cough*, along with belief-frames concerning microbes, contagion, and children's diseases on one side of the dimension and illnesses which involve witchcraft, anger, stomach ache, diarrhea, and having fits on the other side of the space. It is interesting that the range of what is considered contagious appears in a different form in the Mexican data, with dysenteries and respiratory infections not near the "contagious" diseases and related belief-frames.

Another belief-frame by disease term matrix, using slightly different disease terms and belief-frames, was collected by James C. Young (1978) from a Spanish speaking Mexican-Tarascan town. Young used cluster analysis to determine both groupings of disease terms and groupings of belief-frames. To identify the most important belief-frames, Young developed a measure to determine which belief-frames best define particular disease clusters. The belief-frames which best define a particular disease cluster are those in which most of the illnesses in that cluster have that property, while most of the other illnesses in other clusters do not have that property. On the basis of the results of this measure, Young concluded that a major dimension for the Tarascans involved the distinction between *externally caused* illnesses, which result from contact with hazardous agents in the environment like weather, versus *internally caused* illnesses, which result from internally initiated imbalances, including *hot* versus *cold* diet imbalances.

As a general technique, the analysis of "item by feature" matrices appears to do a reasonably good job of identifying the salient features of a domain. However, for cognitive anthropology a distinct shift in the overall goals of analysis occurred as a result of using these kinds of data. In the feature or componential analysis of kin term paradigms, the overall goal was to identify the *criterial attributes* for each term – the *defining* features or properties of each term. In the multidimensional scaling analyses, whether MDSCAL or KYST or clustering, the goal was to find the most *general and salient features* which organize the field of items. However, the most general and salient features of a domain are not necessarily the criterial attributes of the objects in the domain. For example, the feature or property of being a *serious* illness is not a criterial attribute of any of the American diseases. *Cancer* is certainly considered a *serious* illness, but *seriousness* is not a criterial attribute of something being a "cancer." Nor does the fact that *colds* are *contagious* make something a "cold." Yet *contagion* and *seriousness* are two of the most salient features of illnesses for Americans.

The point is that a shift in the technique of analysis led to a shift in what was being looked for. *The shift from componential analyses of distinctive features to the analyses of belief-frame matrices resulted in a shift from trying to find features which could be used to define terms to trying to find features which people respond to most strongly.* Features which are not criterial but which are salient are often called *connotative* features (in contrast to *denotative* features, which consist of criterial attributes). What these new ways of collecting and analyzing data did was make apparent that sometimes connotative features of the items in a domain are of more interest to people than denotative or defining features. In some cases – as in disease terms – the denotative features or criterial attributes are not even known to most folk. It is the medical doctor who knows how to diagnose diseases, not the ordinary person. The categories of concern to most Americans with respect to illness are the conditions which bring about certain diseases and the conditions that these diseases produce.

Under certain conditions an attribute by item matrix data will produce results which correspond closely to direct similarity ratings. For example, in the case of the disease term and belief-frame data, a sample of fifteen American undergraduates were asked to make seven point similarity scale ratings between all pairs of the thirty disease terms listed above. These similarity scale ratings correlated highly (.87) with the number of belief-frame judgments shared by the pairs of disease terms. The reason that these similarity ratings are highly correlated with the degree to which two diseases share belief-frames seems to be because in both cases the same features are involved; that is, in making similarity ratings the rater makes judgments about the similarity of disease terms on the basis of the same features – *contagion*, *seriousness*, *fever*, etc. – that have been used to construct the belief-frames. Of course, if the belief-frames had been *irrelevant* to the features that Americans use to make similarity judgments about diseases, then the number of belief-frames shared by pairs of disease terms would not be correlated with the raters' similarity judgments for these pairs.

Rating correlations based on feature overlap

Ratings are a standard part of the technology of social science. They are used in public opinion attitude polls, in personality tests, and in a vast range of psychological experiments. An ordinary set of character trait rating scales is presented below:

PRESIDENT BUSH

extremely friendly	|	|	|	|	|	|	|	|	|	|	extremely unfriendly
	+4	+3	+2	+1	0	−1	−2	−3	−4	
extremely sociable	|	|	|	|	|	|	|	|	|	|	extremely unsociable
	+4	+3	+2	+1	0	−1	−2	−3	−4	
extremely inventive	|	|	|	|	|	|	|	|	|	|	extremely uninventive
	+4	+3	+2	+1	0	−1	−2	−3	−4	
extremely clever	|	|	|	|	|	|	|	|	|	|	extremely dumb
	+4	+3	+2	+1	0	−1	−2	−3	−4	

Table 4.6. *Character trait ratings of six notables*

	Bush	Einstein	Monroe	Starr	Nixon	Nancy Regan
Friendly/unfriendly	4	2	3	1	–3	–1
Sociable/unsociable	4	2	2	2	–1	2
Inventive/uninventive	–1	4	1	3	2	–2
Clever/dumb	2	4	–1	3	3	1

Table 4.7. *Pearson correlations for four character trait across six notables*

	Friendliness	Sociability	Inventiveness	Cleverness
Friendliness	1.00	.86	–.03	–.26
Sociability	.86	1.00	–.37	–.21
Inventiveness	–.03	–.37	1.00	.58
Cleverness	–.26	–.21	.58	1.00

The instructions would tell the respondent to place a check in the slot that corresponds best to their judgment of how "friendly" or "unfriendly" President Bush is. Typically ratings use adjectives (*friendly, unfriendly*) with adverbial modifiers (*extremely*) to indicate "strength" or "intensity" associated with some number scale. Below is a table of ratings from one informant for four scales (*friendliness, sociability, inventiveness*, and *cleverness*) for six well-known persons (George Bush, Albert Einstein, Marilyn Monroe, Ringo Starr, Richard Nixon, and Nancy Reagan). The result is a special form of *item by attribute* matrix, with persons as the objects and trait ratings as the attributes.

Early in the history of psychology such ratings were used in an attempt to measure individual differences in personality. A standard research technique is to find a social group in which all the members know each other relatively well, and then have every individual rate every other individual in the group on a series of adjective scales. The averaged ratings would then give a fairly good idea of the salient characteristics of each of the members of the group. A variant of this technique is to have a series of individuals rated by a small number of psychologists after an interview or after observing the person in the performance of some sort of task.

An issue that arose early in this kind of research involves the question of the degree to which personality adjectives measure underlying dimensions of personality *or simply reflect the degree of feature overlap between terms.* Consider the four adjectives rating scales presented in Table 4.6. On the basis of feature overlap we would expect the *friendly–unfriendly* scale to correlate highly with the *sociable–unsociable* scale; that is, that someone who is given a high rating on *friendliness* will also be given a high score on *sociability.* We expect this because both scales are similar in *meaning* – that

Table 4.8. *Factor analysis of four character traits across six notables*

	Factor 1	Factor 2
Friendliness	.97	–.03
Sociability	.94	–.03
Inventiveness	–.08	.90
Cleverness	–.14	.86
Variance accounted for	46%	40%

is, the features used to identify someone as *friendly* are similar to the features that identify someone as *sociable.* Both involve features concerning interaction with people in which the interaction is characteristically emotionally pleasant, and the rated person characteristically shows an interest in and liking for other persons. The features of these two terms are not identical – somebody might involve themselves in a lot of social gatherings but not really act very warm to others, and so be rated as *very sociable*, but only *slightly friendly.* However, while we would not expect the correlation between *friendliness* and *sociability* to be perfect, we would expect on the basis of feature overlap that people who are high on one scale are also high on the other scale.

Similarly, we would expect the *cleverness* scale and the *inventiveness* scale to be positively correlated because of the feature overlap between these two scales; both traits involve being able to accomplish difficult intellectual tasks. However, in general we would not expect the *friendliness* scale to correlate with either the *inventiveness* or *cleverness* scale since *friendliness* has few if any features outside of general "goodness" which overlap with being *inventive* or being *clever.* We would also expect the same pattern with the *sociability* scale. Table 4.7 presents the correlations of the four scales across the six persons for the data in Table 4.8.

Of course, these correlation coefficients would change somewhat if a different set of persons were to be rated, or if the ratings were made by a different rater. For this sample we see that *friendliness* and *sociability* are highly correlated at +.86, and that *inventiveness* and *cleverness* are correlated +.58. This is generally what we would expect on the basis of feature overlap – *two traits that have overlapping features will be positively correlated because similar discriminations are used in making both ratings.* For this sample there are relatively small negative correlations between the "social" traits and the "intelligence" traits, which means that for this sample the persons who are *friendly* and *sociable* have a tendency *not* to be *inventive* and *clever.* One would expect that if ratings were made over a larger set of persons that these correlations would move closer to zero, since there is no apparent feature overlap –

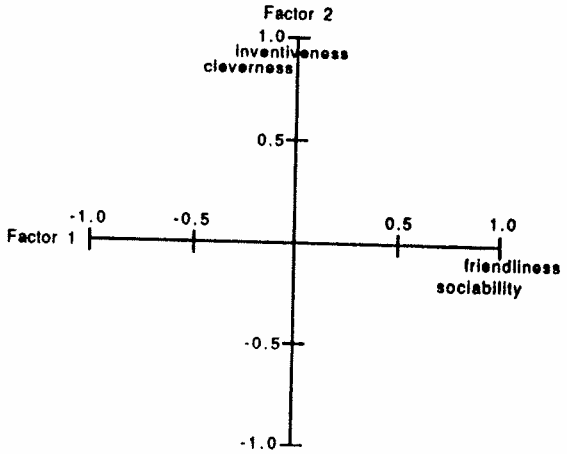

Figure 4.5 Factor plot for four character traits

positive or negative – between these "social traits" and these "intelligence" traits.

A factor analysis of these four traits yields two factors with eigenvalues greater than one.[8] The varimax rotated factor loadings are presented in Table 4.8. The factor loadings for the character traits show a clear pattern in which the social traits have high loadings on the first factor and the intelligence traits have high loadings on the second factor.

These results can be diagrammed in a two dimensional space, each dimension representing a factor. This diagram is presented in Figure 4.5.

The number of traits and ratings in the example above is too small to serve as a real test of our hypothesis. But if more extensive factor analyses of character trait ratings can be duplicated with just similarity judgments, this would indicate that such results are due primarily to feature overlap rather than to the co-variation of distinct attributes in the *external* world such as, for example, the correlation between the heights of fathers and sons.[9]

Over the past twenty years psychologists have developed a general consensus that there are five factors – called the "big five" – which account for most of the ratings using character traits (Digman 1990). Many kinds of raters have been used; teachers rating students, officer candidates rating each other, college

[8] Factor analysis (in this case actually principal components analysis from SYSTAT) is a method of multivariate analysis in which the original matrix of correlations is resolved into a smaller number of orthogonal determinants. Like multidimensional scaling programs such as KYST, the resulting factor loadings can be given a spatial representation.

[9] It is not being argued that *all* the correlations between ratings can be accounted for by feature overlap. Ideas about traits that go together, such as being "loud" and being "aggressive," can also be a source of correlational structure, whether or not these traits do, in fact, co-vary. For an exploration of this hypothesis with respect to personality traits, see Gara and Rosenberg 1981.

Table 4.9. *Factor loadings for twenty character scales*

Factor name and scale	Factor loadings F1	F2	F3	F4	F5
1. Extroversion					
Talkative–silent	**90**	02	–02	04	00
Frank, open-secretive	**78**	–08	07	–03	07
Adventurous–cautious	**79**	15	–20	32	01
Sociable–reclusive	**86**	01	–18	–01	–02
2. Agreeableness					
Goodnatured–irritable	17	**80**	17	12	07
Not jealous–jealous	–10	**64**	20	49	07
Mild, gentle–headstrong	–20	**80**	27	19	10
Cooperative–negativistic	33	**74**	28	13	11
3. Conscientiousness					
Fussy, tidy–careless	–33	–08	**66**	–35	20
Responsible–undependable	–03	32	**86**	08	18
Scrupulous–unscrupulous	–30	–08	**66**	–35	20
Persevering–quitting, fickle	–05	28	**74**	12	27
4. Emotional stability					
Poised–nervous, tense	01	56	15	**61**	05
Calm–anxious	06	21	–10	**82**	–07
Composed–excitable	13	06	16	**71**	24
Not hypochondriacal–hypochondriacal	21	27	00	**65**	–09
5. Culture					
Artistically sensitive–insensitive	–04	08	39	–10	**75**
Intellectual–unreflective, narrow	–04	05	47	04	**74**
Polished, refined–crude, boorish	15	25	53	16	**46**
Imaginative–simple, direct	12	19	03	10	**68**

students rating each other, clinical staff rating trainees, etc. A variety of different trait adjectives have been used in a variety of rating formats. The results have been remarkably consistent, each study finding essentially the same five factors. Many psychologists consider these five factors to constitute the five major *dimensions* of human personality. Table 4.9 presents Warren Norman's 1963 factor analysis of these five dimensions which have served as a standard for other analyses.

In 1965 I attempted to test the hypothesis that these five factors were the result of feature overlap. I had a small sample of ten undergraduates rate the degree of similarity in meaning of all possible pairs of the negative poles (*silent, secretive, cautious*, etc.) for each of the Norman twenty bipolar scales.[10] The results produced five factors with almost exactly the same composition as the Norman analysis. A more elegant result was obtained by Milton Hakel (1974) who had a large number of respondents make co-occurrence judgments

[10] D'Andrade 1965. At the time multidimensional scaling programs were unavailable, so I treated the rows of the similarity matrix as subjects and the columns as variables to compute a correlation matrix which was then factor analyzed.

Table 4.10. *Multidimensional scaling of similarity ratings for Norman character traits*

Factor name and scale	Factor loadings F1	F2	F3	F4	F5
1. Extroversion					
Talkative	**84**	–35	21	13	–15
Frank, open	**92**	11	–24	–06	–31
Adventurous	**60**	–51	48	–40	–38
Sociable	**80**	–59	20	–29	09
2. Agreeableness					
Goodnatured	–33	**92**	04	17	06
Not jealous	13	**84**	–01	40	08
Mild, gentle	09	**81**	43	13	15
Cooperative	–61	**65**	31	31	16
3. Conscientiousness					
Fussy, tidy	05	–63	**65**	–49	09
Responsible	–23	36	**84**	03	06
Scrupulous	–12	–26	**89**	–14	07
Persevering	–06	–09	**80**	38	19
4. Emotional stability					
Poised	–12	21	34	**82**	06
Calm	05	24	–08	**87**	–37
Composed	12	–45	42	**70**	–20
Not hypochondriacal	–33	08	–23	**85**	–06
5. Culture					
Artistically sensitive	–08	–46	30	–26	**80**
Intellectual	–48	10	47	26	**75**
Polished, refined	–53	**60**	54	–19	28
Imaginative	07	13	–11	–07	**103**

("Suppose a person is ____ – how likely is it that he is also ____?") on both sides of the twenty bipolar scales, resulting in a forty by forty matrix. A total of 480 respondents each judged 100 pairs of descriptors in a balanced block design in which no pair was repeated on the same questionnaire and each pair was represented thirty times across all questionnaires. The mean score for each cell was computed and the matrix analyzed with MDSCAL. Hakel's results for just the positive side of each scale are presented in Table 4.10. Out of the twenty positive trait terms, only one does not have its highest loadings on the factor predicted from the Norman analysis. The structural similarity of Hakel's and Norman's results is overwhelming.

The fact that the similarity ratings give the same dimensional results as a correlation matrix of traits does not mean that these dimensions or factors are worthless, or that the specific ratings are worthless. There is considerable evidence that personality trait ratings are both reasonably reliable and valid

[11] See Wiggins 1973 for a review of these issues.

descriptors of behavior.[11] For example, the fact that ratings about someone's "talkativeness," "frankness," "adventurousness," and "sociability" are likely to be positively correlated does not mean that these ratings are inaccurate, only that to some extent *they all measure the same thing*. The issue being contested here is that the correlations reveal something about *real dimensions of personality* rather than *a linguistic fact that these traits all overlap in meaning*.

Dean Peabody summarizes these studies as follows:

> Wiggins (1973) made a useful distinction between *external* structure (based on judgments about the traits of people) and *internal* structure (based on judgments about the relations between traits). The radical attack on personality, initiated by D'Andrade (1965) essentially claims that external structure is simply a projection of the internal structure of the judges. D'Andrade obtained judgments about internal structure (similarity of meaning) for Norman's 20 scales and claimed to find the same Big Five factors as in studies of external structure. Such a correspondence between external and internal structure is also widely accepted by defenders of personality, who argue that this correspondence does not preclude the validity of personality assessment. (1987:67)

How, then, do "defenders of personality" explain the high degree of correspondence between "external" ratings and "internal" judgments of similarity? Basically, they argue that the external structures exist – that is, that personality is really composed of these five major dimensions – and that people have observed and identified these dimensions and have over time encoded them as salient features of trait terms (Goldberg 1982). This could be true. Data from other cultures is needed to resolve the issue.

A number of other studies were carried out in which external ratings of various sorts were reproduced by similarity ratings. Richard Shweder showed that a factor analytic classification of interpersonal behavior, developed by F. Bales, could be replicated solely from similarity judgments (1972). Shweder also replicated with similarity judgments the Alpha factor of the Minnesota Multiphasic Personality Inventory (MMPI), a questionnaire used to diagnose forms of mental illness (1977). I replicated the Leary Grid organization of interpersonal behavior (1965), as well as the four dimensional structure of Overall's Brief Psychiatric Rating Scale (Shweder and D'Andrade 1980).

Memory based rating

All of the examples so far involve ratings in which raters judge on some kind of scale how strongly some target person displays some particular characteristic. A related type of data is often collected by psychologists and others interested in interpersonal behavior. In this kind of study the investigator begins with a classification of behavior acts. A group of people are observed and each time they perform one of these behavior acts, that act is counted. At the end of the observation period every member of the group can be characterized by a frequency count for how many times he or she performed each of the behavior

acts in the classification system. Using the frequencies, the different behavior acts can then be correlated across persons and analyzed in various ways.

In a number of studies, not only were actual behaviors observed and counted, but *retrospective* data was also collected. The retrospective data consisted of the observer's *memory* of what happened. Typically the observers are asked to indicate on a scale running from *very frequently* to *not at all* how often each member of the observed group performed each of the behavior acts. These recalled frequencies can then be used to correlate the different kinds of behavior acts across persons in the same way the actual frequencies are used.

What is interesting is that typically the correlations for the immediately observed behaviors correspond only weakly to the correlations for the recalled behaviors (D'Andrade 1974). However, the correlations for the recalled behaviors correspond *strongly* to the judged ratings of the *similarity* of the different behavior acts. Across seven different tests of intermatrix correspondence taken from a variety of studies, Shweder (1982) found that similarity ratings and recalled behavior correlations show a mean r of +.75, while recalled behavior correlations and immediately observed behavior correlations show a mean r of only +.25 and immediately observed behavior correlations and recalled behavior correlations show a mean r of only +.26.

All this correlating of correlations gets somewhat confusing. Most simply, what this data shows is that, at least for the case of classifications of specific kinds of behavior (e.g. *informs, questions, explains, jokes*, etc.), people's memories do not reflect accurately what kinds of acts go together. What memory based ratings show is that what people remember as going together are the kinds of behavior they judge to be similar. Humans show a systematic distortion in their memories. They falsely recall "what goes with what" based on "what is like what." This effect has been demonstrated across a wide range of kinds of materials, not just behavior frequencies (for example, see Chapman 1967). Overall, these results throw doubt on a broad class of retrospectively based research data.[12]

While the seriousness of the problem of feature overlap and systematic memory distortion for personality assessment is still a controversial issue, the results reported above are much happier for the study of the organization of *cultural* features. That is, the evidence is that similarity judgments, item by property matrices, and external ratings all typically uncover cognitively shared salient features. Cognitively shared salient features are an interesting part of a society's culture. From an anthropological viewpoint, much of the research by psychologists on dimensions of personality traits and dimensions of interpersonal behavior can be seen as an investigation of western culture.

Consider, for example, the Norman "big five" dimensions of personality. If

[12] See Dawes and Pearson 1991 for an excellent review of the problems involved in retrospective questionnaire data.

we look at these dimensions as culturally meaningful ways of organizing our knowledge about people, we see that these dimensions presuppose a certain way of dividing up the individual's world. First, almost all trait terms are *evaluative* – either positively or negatively. This is a cultural system which is oriented towards evaluating how well people do certain things rather than just presenting neutral description. For example, the second dimension, called *Agreeableness*, consists almost entirely of traits which involve *interpersonal behavior*; they describe different ways of treating other people well or badly. The term "agreeableness" is simply one of the more general ways of describing a good way of getting along with others. A more accurate but less "personality" oriented way of describing this second dimension would be to call it the "evaluation of interpersonal behavior." Similarly, the third dimension, *Conscientiousness*, consists of evaluations of how well people do *work* – "responsibly," "with perseverance," "carelessly," etc. The fifth dimension, called *Culture* by Norman and *Intellect* by others, involves the evaluation of the person with respect to basic ideational systems of a society, such as knowledge and art. Only the first and the fourth factor are defined by qualities internal to the person. The fourth factor, *Emotional Stability*, focuses on the rich and varied emotional communication system carried by facial expression, tone of voice, and body language. It seems reasonable that evaluation of the quality of emotions a person communicates would be a salient feature of the individual in a world in which important social relationships, such as marriage and friendship, are based primarily on affective rewards. Finally, the first dimension, called *Surgency* or *Extroversion*, appears to be primarily a measure of *level* of social participation and expressiveness.[13]

Thus the "big five" dimensions of personality, viewed anthropologically, divide the world into *work, interpersonal relations, idea systems (knowledge and art), emotions*, and *general activity*. Evaluation of people in terms of these domains makes sense in modern industrial societies. To date, the evidence, while somewhat sparse, indicates that these same dimensions are found in Germany, Japan, and Israel (Digman 1990). However, Geoffrey White, working in an A'ara speaking Solomon Island society, found what appears to be quite a different system of organization of personality traits (1978). These Solomon Islanders are primarily subsistence horticulturalists. Kinship and political leadership constitute the major institutional systems of their society. Working in the A'ara language, White collected thirty-seven common personality descriptors. Using a modified sorting method in which informants are asked for each term to choose the five other terms that are most like it in meaning, twenty-five male adults rated all terms. Using a median method,[14] White cluster analyzed the thirty-seven trait terms. The results are presented in

[13] This argument is made in more detail in D'Andrade 1985.

[14] Also sometimes called U-statistic clustering; see D'Andrade 1978.

Figure 4.6 using English glosses for the A'ara trait terms. The terms in italics have been added to emphasize what appear to be the most salient contrasts between the clusters. Basically, the A'ara system has two main dimensions; *good* vs. *bad* ways of behaving and *leader-like* vs. *follower-like* ways of behaving. This makes four clusters – *bad followers*, *bad leaders*, *good leaders*, and *good followers*.[15]

The overall argument here is that in a society with different social institutions, the *salient features* of personality trait terms are organized in a different way. White (1978) points out that the pattern found for the A'ara seems general for Melanesian societies in which leadership is highly valued, but where there are often also strong egalitarian values which conflict with domineering styles of leadership, leading to ambivalence about powerful leaders. A'ara personality terms certainly do *not* have an organization of terms like Norman's factors, throwing doubt on the cross-cultural universality of the "big five."

The material presented in this chapter has shown that similarity judgments, item by attribute matrices and external ratings all can be used to uncover general, salient cognitive features which people use in structuring a domain. This gives the cultural analyst a kit bag of techniques which can be used to investigate any particular topic. If there are not too many items, similarity judgments are a direct and simple way of finding out how the items are organized. To investigate simultaneously the organization of features and items, the analysis of an item by attribute matrix is preferable. If there is a large number of items and the domain is not one which has a clear item by attribute organization, external ratings are likely to be most effective method.[16]

The categorization of interpersonal relationships is a good example of a domain which is most appropriately investigated by means of external ratings. There are a great number of statements that people make about interpersonal relationships, making it difficult to use direct similarity judgments. Since these statements about interpersonal relationships are themselves attribute-like (e.g. *I feel at ease around* ___) it is possible to use external ratings of actual relationships to form a data matrix.

As part of a project to study the way Americans evaluate themselves and each other, I undertook an investigation in 1982 of the way in which Americans categorize relationships. The first step was to elicit statements which individuals normally use in describing their relationships. The statements were taken from a series of informal tape recorded interviews in which informants were queried about their family, friends, lovers, acquaintances, and enemies. From the transcripts a series of 300 statements were abstracted which had been used to characterize relationships. A small sample of informants were asked to

[15] White (1978) carried out a KYST multidimensional scaling of these data which strongly supports this dimensional interpretation.

[16] There are also a number ways of combining these three techniques; see Wish, Deutsch, and Biener 1972 for an example in which similarity ratings and external ratings are used together.

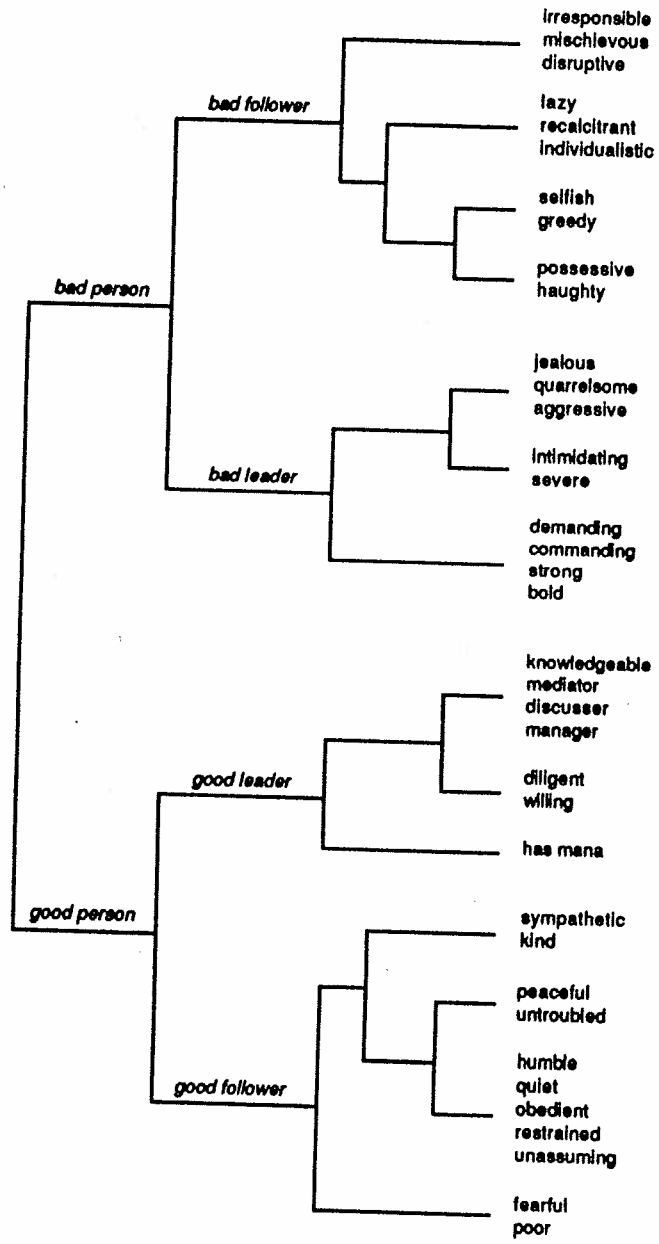

Figure 4.6 Hierarchical clustering of A'ara personality descriptors

choose "the most unambiguous statements which were also most diagnostic about the character of relationship being described." Fifty-eight statements were finally selected for analysis.

The second step was to have ratings made on these fifty-eight statements about specific relationships by a reasonably large sample of respondents. A total of fifty-two undergraduate respondents rated people they knew well. Each respondent was asked to select twenty persons with whom they had significant relationships, making sure that family/non-family, liked/not-liked, and male/female targets were included in the selection. For each person selected, the respondent was asked to rate on a seven-point scale how well each of the fifty-eight statements applied to their relationship with this person. All pairs of statements were then correlated with each other across relationships and respondents. The resulting correlation matrix was then cluster analyzed using the median method. The result is presented in Figure 4.7. The terms in capital letters have been added as interpretations of the clusters.

The clustering of propositions about interpersonal relationships divides neatly into three general groupings; *negative* relationships, *unbalanced* relationships in which the other person wants more from the rater than rater wishes to give, and *positive* relationships. The positive relationships have three distinct sub-clusters; one concerning relations of *mutual caring* and *closeness*, one involving the *dependability*, *support*, and *honesty* of the person being rated, and one concerning *romantic* relationships.

This clustering is notable for several reasons. First, the strong dislike of being controlled by others and being treated as an inferior person, often noted in studies of American national character,[17] comes through strongly. The opposite of this kind of domination is found in the sub-clusters of mutual caring and mutual sharing, in which an egalitarian relationship is maintained at the same time each person in the relationship both gives and receives material and emotional resources. The problem of inequality also arises in the clusters which involve one partner needing more than the other, making the relationship insecure. Overall, the strong anti-authoritarian and pro-egalitarian configuration of the clusters is impressive.

Applied research

The extended feature model has been used in applied research for both marketing and political polling. Volney Stefflre, for example, has developed a set of techniques by which predictions about preferences for different brands of commercial products can be made from the judged features of these brands (1972). The purpose of this research is not simply to predict product preference, but to help businesses develop new products which can compete successfully

[17] See, for example, Geoffrey Gorer's insightful study, *The American People*.

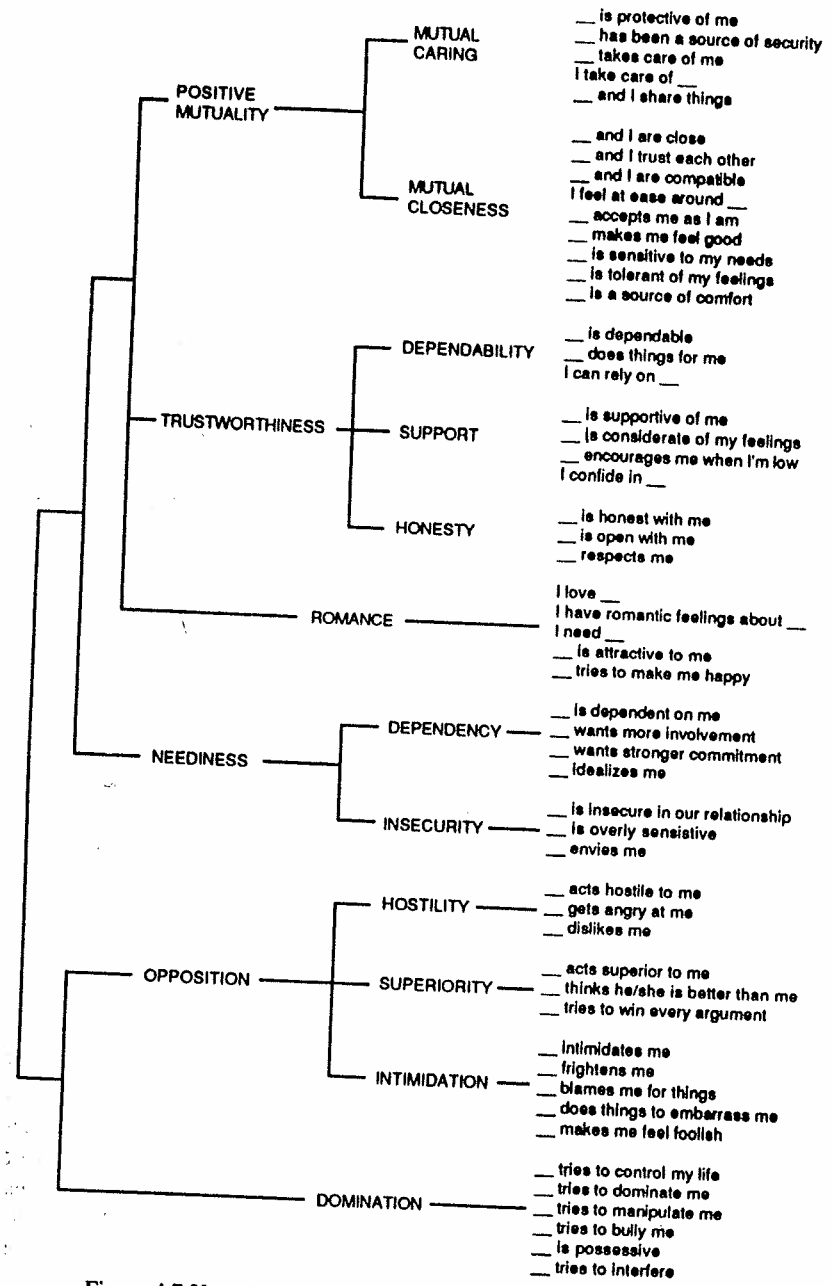

Figure 4.7 U-statistic hierarchical cluster analysis of ratings of propositions about interpersonal relationships

with their competitors' products but not their own. For example, Stefflre was employed by a coffee company to develop a feature profile for a new brand of coffee which would compete in the market with other competing companies' coffee brands. Based on various rating tasks, Stefflre developed a profile of a *light, clean, friendly,* and *mild* coffee. (One does not normally think of *friendly* as a coffee flavor, but apparently consumers use this term often in judging which coffees they like.) By varying bean selection and roasting processes, after a number of trials a new coffee was created which had these features of taste. This coffee did reasonably well in the market, taking away purchases from the seven brands with which it was expected to compete. The Pearson correlation between the predicted and obtained amounts which the new coffee drew from each of the other seven brands was .94 in a survey of performance in a national market.

Stefflre has pointed out that the relation between features and similarity judgments is not uniquely fixed; it depends on the context of judgment. With respect to *making* coffee, two instant coffee brands might be judged as quite similar to each other. However, with respect to *taste*, these same two coffees might be judged as quite different. The point is that features are judged with respect to the aims and goals of the individual, and as a result a small shift in aims and goals can make for large shifts in feature weights.

Stefflre has also pointed out that it is generally true that there is considerable agreement between individuals from the same culture in "what is like what," but considerable variability in "what is liked." Or, to put this another way, *preference judgments are more variable than similarity judgments.* Almost everybody will agree that chicken tastes more like turkey than steak, yet there is considerable variation with respect to judgments about which one tastes best. However, to predict preferences, one can expect that a person will like those things which are most similar to other things that person likes. A person who strongly prefers turkey to steak is likely to prefer chicken to hamburger.

In retrospect

During the middle and late sixties there was a large increase in techniques aimed at uncovering salient features. Many of these techniques were built on the number crunching power of the computer. Various multidimensional scaling programs made it possible to uncover the feature organization of a domain using new kinds of data, such as similarity judgments and external ratings. Psychologists were also interested in the feature model as a way of understanding the basic processes of categorization and added greatly to this new technology. However, in the development and application of these techniques of analysis, there was a shift in the model of feature analysis itself. Not only was there a shift away from the identification of defining or denotative features to the identification of whatever features were most salient to respondents, there was also a shift from the dichotomous features of most lexical analyses to a more continuous dimensional type of representation. Further, in the earlier work, much of the point of such analyses was an economical definition of the items in the domain. The goal of definition *per se* drops out of the later work, and the new emphasis is on finding which, out of the many features that potentially apply to a set of items, are the most salient and frequently used. The basic task was no longer to find out how particular terms are defined, but rather to discover the most general categories people use to understand their world.