USING ANTHROPAC TO INVESTIGATE A CULTURAL DOMAIN

By *Stephen Borgatti (U of South Carolina)*

With this issue, we begin a series of comments on using ANTHROPAC, a set of programs written by Stephen Borgatti. ANTHROPAC was described in the first issue of CAM. The program is available from Borgatti for $25. Write to Borgatti at the Sociology Dept., University of South Carolina, Columbia, SC 29208. Tel. (803) 777-3123. Bitnet N040016@UNIVSCVM.

Suppose you are interested in the prestige associated with various occupations, tasks, or roles. You are aware that a standard list of occupations has been used "successfully" in a number of industrialized countries, but you are concerned that most of the items on the list would be quite meaningless to the people you study. Conversely, many of the occupations relevant to them are not on the list. So (having read Weller and Romney 1988) you decide to begin by eliciting occupation terms from a small group of respondents, say, 20 people. Depending on the exact nature of your interest, you might ask these people some variant of "So, you and Fulana are weavers. What else do people do around here?"

This procedure yields a list of terms or phrases from each respondent. One thing you notice right away is that the first few respondents you talk to generate most of the terms that are mentioned with any frequency. The last respondents don't generate any new terms, except ones which only they mention. Suspicious (and masochistic) researchers will then interview another 20 people just to see if these idiosyncratic items start getting additional mentions. But they don't. Instead, each new respondent reprises the items that everyone mentions, and then adds a few new ones that no one has mentioned, and so it goes. The more you interview, the longer the set of idiosyncratic items grows.

Formally or informally, you have a theory for all this. You believe that there is a cultural domain of occupations which is shared by all members of the culture. You believe that your elicitation taps that domain and also taps other cognitive processes that generate additional terms: individual respondents, put to the test, use their powerful brains to generate a series of terms that are logically or illogically related to the culturally shared items, but which don't actually belong in the domain.

In any case, you need a way to chop down the list of terms to include (a) only those terms which you believe are truly shared, and (b) few enough terms than you can comfortably manage. To do this, you would like to generate a frequency distribution, sorted from most to least frequent. This is something ANTHROPAC can help with.

Step 1. Enter the data in an ASCII file using the ANTHROPAC full-screen editor. The exact format is given in the manual. Essentially, you just type each term or phrase one to a line, in the order in which they were mentioned. Pound signs (#) are used to separate respondents.

Step 2. Start up ANTHROPAC and choose COLLECTION from the main menu. Then choose FREELIST from the COLLECTION submenu. The program will ask the name of your data file, and will permit you to change the default values of various parameters. One of these is whether to use the SOUNDEX phonetic matching system. If you select this option, then ANTHROPAC will ignore exact spellings of words and concentrate only on how the words would sound in English. In other words, if you spelled "night" in one case and "nite" in another, ANTHROPAC would see those as the same term, and will tally two occurrences of that term rather than one occurrence each of two different terms. This makes entering data much easier because you don't have to contend with different respondents and/or interviewers spelling things slightly differently. Of course, it's only a computer program: chances are it will not make every choice exactly as you would have made it. Consequently, it is programmed to ask you if two items that are below its threshold of similarity are really the same item. A new data file in which SOUNDEX-equivalent terms are combined into one, is one of the outputs of the FREELIST program.

Step 3. The main output of the FREELIST procedure is a sorted list of items and their frequencies. This is the one you use to determine a cutoff. If you're lucky, there is a natural gap between the high frequency, high consensus items, and the low frequency, idiosyncratic items. If not, you make the decision based on other criteria such as convenience and the opinions of informants and colleagues.

Either way though, chances are that things don't go right the first time. Looking through the chart you see that some of the items that got just one mention are simply misspellings of others. That means that (a) they shouldn't be counted as terms in themselves, and (b) the frequency counts for the items they were intended to be are too low.

What you need to do is edit the data file (either the raw or the cleaned one) and run it again. You may need to repeat these three steps several times before getting it just the way you would like.

Once you pass Step 3, you are ready for a little analysis. We take up Step 4 next issue.

References

Weller, Susan C. and A. Kimball Romney
1988 Systematic Data Collection. Newberry Park, CA: Sage Publications.