In search of one set of categories

ROEL POPPING
Department of Sociology, University of Groningen, Grote Rozenstraat 31, NL 9712 TG Groningen, The Netherlands

Abstract. A method is described for transferring the categories as developed by several raters on the basis of the answers given to open-ended questions to one set of categories. This set covers all possible answers in an adequate way.

1. Introduction

Very often data will be collected for which at that moment there is not already a good set of categories available to which they should be assigned. Such data are in general answers to open questions as part of questionnaires, or statements noted on cards that should be taken together. Also entities as used in content analysis can be considered as such.

If a good set of categories has to be developed for such data, one possibility is to set judges to work. Using the data gathered, they develop a set of categories and assign the data to these categories. This way of working is denoted the a-priori method of coding (Montgomery and Crittenden, 1977: 236).

The a-priori method of coding is especially useful if one is looking for a set of categories that has to be used in a main investigation. The method is also to be used when the kind of answers that will be gathered is not (yet) clear. This often holds for studies where content analysis or (participating) observation will be used. The method can also be useful when training judges.

When the judges do their task using the a-priori method of coding, they will in general formulate a different set of categories. The question in this article is how to get from this situation to one set of response categories that contains all relevant information. About the requirements that are to be put to a set of response categories, see Lazarsfeld and Barton (1968: 156–157).

Coding tasks can be divided into three types (Crittenden and Hill, 1971: 1078): A, B1, and B2. “Type A coding tasks require a coder to find a specific answer to a specific question at a given place on an instrument. Type B1 coding tasks involve locating relevant information within a larger context . . . , type B2 coding tasks are those where the coder has not only to
locate relevant information, but also to evaluate the relative importance of two or more possible responses to arrive at a single code” (Montgomery and Crittenden, 1977: 236). A complex coding task might force an investigator to split the original question into sub-questions. The following presentation starts from the situation of one (sub-) question.

2. To one set of categories

When the a-priori method of coding has been used, different sets of categories can be formulated. Therefore a continuation should come on the coding task. Elsewhere three possibilities are distinguished (Popping, 1982):
1. One can look for dimensions that might be found behind the formulated sets of categories;
2. One can invite the next judge to use the categories as developed by the first judge;
3. One can start a discussion between the judges.
The first point might be interesting if statements about health or politics have to be classified; I do not know a concrete elaboration. About the second point it is known from an own experiment that the method does not give so much results. A new judge tends to develop an own set of categories, and very often is not willing to join the set formulated by a previous judge. The third method offers good perspectives, therefore it will be treated below.

The method is described by Montgomery and Crittenden. They distinguish three steps. The first one is the comparison of pairs of codings and the computation of the amount of agreement. The second is “to develop final coding categories and instructions. The responses and coding categories are examined to develop a smaller final set of categories for the item. Reformulation of categories is based on substantive considerations, taking into account the cross-tabulation of the original codes... Coding instructions for the final set of categories are developed with the help of the original coders.

Once the final coding instructions are developed, step 3 is to code the responses by typical procedures” (Montgomery and Crittenden, 1977: 239).

For the second step as formulated by Montgomery and Crittenden (which they have not elaborated any further) two working methods are proposed here that might complement each other:
1. Judges evaluate the categories as developed and the assignments. They redefine or combine categories. By computing the amount of agreement at once after an action (for example combining two categories), one determines whether this agreement increases. If desired one can leave answers that are hard to classify in first instance out of the computations.