

## DATA CODING IN QUALITATIVE RESEARCH

Coding social data (for example, text, images, talk, interactions) is sometimes derided as tedious, but if you think of it as a kind of detective work, it can be intriguing, exciting, and very valuable to the research process.

The purposes of coding are partly **data reduction** (to help the researcher get a handle on large amounts of data by distilling along key **themes**), partly organization (to act as a 'finding aid' for researchers sorting through data), and partly a substantive process of data exploration, analysis, and theory-building. Further, different researchers use coding for different reasons depending on their goals and epistemologies; sometimes coding is used in an exploratory, inductive way such as in **grounded theory** in which the purpose is to generate theories from empirical data, while other times coding is used to support a theory or hypothesis in a more deductive manner. Several approaches are discussed here, with pointers on how to organize and begin the coding aspect of a research project.

### **Types of Codes and Coding**

One common type of coding is **content analysis**, which is essentially a *quantitative* technique and by no means represents the full extent of coding for qualitative research. Content analysis can be done by 'hand' or by computer (see Chapter 15 for a discussion), but either way it is a system of identifying terms, phrases, or actions that appear in a document, audio recording, or video and then counting how many times they appear and in what context. For example, a researcher might be interested in how many times the word 'democracy' is used in newspaper articles from a particular country or in how many and what type of places are portrayed in a television program. Frequently in content analysis, sampling is used in similar ways to quantitative analysis of populations; perhaps only front-page newspaper stories are included in the analysis, or a television program is sampled for five minutes out of each hour. Similarly, researchers using content analysis typically subject their coded findings to standard statistical analysis to determine frequencies, correlations, variations, and so on. There are many good guidebooks and instructions for conducting content analysis, including some available on the Internet (see, for example, Krippendorff 2003; Neuendorf 2001).

While content analysis is a frequently used type of coding, the primary focus of this chapter is on qualitative approaches to coding. However, one of the basic principles of content analysis has broad implications for all coding activity and is thus worth exploring further. It is the notion that there are both 'manifest' and 'latent' messages contained in the material (for example, text, images, video). **Manifest messages** are those that are blatant and obvious—they then generate manifest *codes*. For example, if I (as a feminist geographer) were performing content analysis on a set of international newspapers and the term 'sex worker' appeared with some level of frequency, I would take that as a code and proceed to scan subsequent materials for it.

In much of ethnographic work in which researchers use coding qualitatively, the ideas of manifest and latent codes have parallels in 'descriptive' and 'analytic' codes. **Descriptive codes** are similar to manifest codes: they reflect themes or patterns that are obvious on the surface or are stated directly by research subjects. Descriptive codes can be thought of as category labels because they often answer 'who, what, where, when, and how' types of question. Examples of descriptive codes that might interest geographic researchers include demographic categories (male, female, young, elderly), site categories (home, school, work, public space), or even scale identifiers (local, regional, national, global).

One special type of descriptive code is called ***in vivo* codes**; they are descriptive codes that come directly from the statements of subjects or are common phrases found in the texts being examined (Strauss and Corbin 1990). For example, if interviews were done with elderly women and they repeatedly mentioned concern with crime in their neighbourhoods, 'crime' would become an *in vivo* descriptive code—the term is used by and describes something important to the subjects. *In vivo* codes are a good way to get started in coding, particularly in projects that are designed to be inductive or exploratory.

Ethnographers also develop **analytic codes** to code text (or other forms of data) that reflect a *theme* the researcher is interested in or one that has already become important in the project. Analytic codes typically dig deeper into the processes and context of phrases or actions. For example, it might become apparent that the elderly women mentioned above were especially afraid of young men and boys whom they perceived as threatening while walking down the street, and therefore an analytic code called 'fear of young males in public space' could be developed. This code might then be applied to

the rest of the data to identify other instances of fear, perceptions of young men, and experiences in public spaces.

Often, descriptive codes bring about analytic codes by revealing some important theme or pattern in the data or by allowing a connection to be made (for example, crime, fear of youth in public), while other times analytic codes are in place from the beginning of the coding process because they are embedded in the research questions. For instance, if we were interested from the start in how elderly women navigate urban spaces, their personal mobility and impediments to mobility would be themes reflected in the analytic codes from the project's very beginning. The recursive strength of coding lies in its being open to new and unexpected connections, which can sometimes generate the most important insights.