# 2

# Concepts and Coding

## Linking Concepts and Data

Many analyses of qualitative data begin with the identification of key themes and patterns. This, in turn, often depends on processes of coding data. The segmenting and coding of data are often taken-for-granted parts of the qualitative research process. All researchers need to be able to organize, manage, and retrieve the most meaningful bits of our data. The usual way of going about this is by assigning tags or labels to the data, based on our concepts. Essentially, what we are doing in these instances is condensing the bulk of our data sets into analyzable units by creating categories with and from our data. This process is usually referred to as coding, although that can imply a rather mechanistic process. We prefer to think in terms of generating concepts from and with our data, using coding as a means of achieving this.

We stress here that although coding may be part of the process of analysis, it should not be thought of as the analysis in itself. In other words, coding should not be seen as a substitute for analysis. It would be as much a mistake to think that coding is an activity that is universally

understood across the qualitative (or indeed quantitative) research spectrum. Rather, the term *coding* encompasses a variety of approaches to and ways of organizing qualitative data. As parts of an analytical process, however, attaching codes to data and generating concepts have important functions in enabling us rigorously to review what our data are saying.

The analytic procedures that underpin coding procedures establish links of various sorts. First, codings link different segments or instances in the data. We bring those fragments of data together to create categories of data that we define as having some common property or element. We define them as being about or relating to some particular topic or theme. The coding thus links all those data fragments to a particular idea or concept. As we will see, such concepts are in turn related to one another. Codes, data categories, and concepts are thus related closely to one another. The important analytic work lies in establishing and thinking about such linkages, not in the mundane processes of coding. The importance of the work lies in how we use the codings and concepts, not in whether we use computer software to record them or rely on manual ways of marking and manipulating the data.

Important analytic work also lies in the identification of relevant concepts. We use the data to think with, in order to generate ideas that are thoroughly and precisely related to our data. Coding can be thought about as a way of relating our data to our ideas about those data. Because codes are thus links between locations in the data and sets of concepts or ideas, they are in that sense heuristic devices. Coding reflects our analytic ideas, but one should not confuse coding itself with the analytic work of developing conceptual schemes. As Seidel and Kelle (1995, p. 52) note, "codes represent the decisive link between the original 'raw data,' that is, the textual material such as interview transcripts or fieldnotes, on the one hand and the researcher's theoretical concepts on the other."

In practice, coding can be thought of as a range of approaches that aid the organization, retrieval, and interpretation of data. Miles and Huberman (1994) suggest that coding constitutes the "stuff of analysis" (p. 56), allowing one to "differentiate and combine the data you have retrieved and the reflections you make about this information" (p. 56). They argue that coding is a process that enables the researcher to identify meaningful data and set the stage for interpreting and drawing conclusions. They describe codes as

> tags or labels for assigning units of meaning to the descriptive or inferential information compiled during a study. Codes usually are attached to "chunks" of varying size—words, phrases, sentences or whole paragraphs, connected or unconnected to a specific setting. They can take the form of a straightforward category label or a more complex one (e.g. metaphor). (Miles & Huberman, 1994, p. 56)

They go on to say how they see codes being used to retrieve and organize data:

> The organizing part will entail some system for categorizing the various chunks, so the researcher can quickly find, pull out and cluster the segments relating to a particular research question, hypothesis, construct or theme. (Miles & Huberman, 1994, p. 57)

Later in this chapter, we provide some examples, drawn from the data on anthropology students and their mentors, illustrating how codes can be assigned to chunks of data and how we can then use these codes to generate concepts and themes. Before we do so, however, it might be useful to review some of the different ways in which coding can be approached.

On the one hand, coding can be thought about in terms of data simplification or reduction. If the codes are kept to a general level and their number relatively small, then the data are reduced to their bare bones, stripped down to a simple general form. This coding approach can be compared directly to simple forms of content analysis (Krippendorf, 1980). The addition of simple, broad analytic categories or codes can thus be used to reduce the data to manageable proportions. Here the analyst is concerned primarily with the identification of a simple conceptual schema. The main goal of such coding is to facilitate the retrieval of data segments categorized under the same codes. Coding in this context is essentially a process of indexing the data texts, whether they be fieldnotes, interview transcripts, or other documents. Data are reduced to equivalence classes and categories. The qualitative analyst will thus be able to retrieve chunks or segments of textual data that share a common code. Such code-and-retrieve procedures can be used to treat the data in quasi-quantitative ways by, for example, aggregating instances, mapping their incidence, and measuring the relative incidence of different codes. Such coding and retrieving can be implemented in a

variety of manual styles. Texts can be marked up physically with marginal keywords or code words, different colors can be used to mark or highlight the texts, and index cards can be used to cross-reference instances to numbered pages or paragraphs in the data.

Such a code-and-retrieve procedure also has been implemented using a number of computer software packages. There are now a number of applications designed specifically for the analysis of qualitative data, some of which we introduce in more detail in Chapter 7. Many of the contemporary programs incorporate code-and-retrieve functions. Such data-handling procedures also can be accomplished to varying degrees by the general cut-and-paste functions of word-processing software (see Stanley & Temple, in press).

Coding and retrieving is the procedure most often associated with coding as an analytic strategy. The role of coding in such a conceptualization is to undertake three kinds of operations, according to Seidel and Kelle (1995, pp. 55-56): (a) noticing relevant phenomena, (b) collecting examples of those phenomena, and (c) analyzing those phenomena in order to find commonalities, differences, patterns, and structures. Seidel and Kelle are clear that even when coding is used to reduce data, codes are heuristic devices. In this sense, coding qualitative data differs from quantitative analysis, for we are not merely counting. Rather, we are attaching codes as a way of identifying and reordering data, allowing the data to be thought about in new and different ways. Coding is the mechanics of a more subtle process of having ideas and using concepts about the data. It can be viewed as

> nothing more than a preparation for this process which is based on a careful inspection and analysis of raw data (that is segments of text) and on their comparison for the sake of identifying patterns and structure. (Seidel & Kelle, 1995, p. 58)

As well as data simplification and reduction, coding can be conceptualized as data complication. Coding need not be viewed simply as reducing data to some general, common denominators. Rather, it can be used to expand, transform, and reconceptualize data, opening up more diverse analytical possibilities. We go on to say more about this later in the chapter, but it is important to recognize at the outset that in the hands of commentators such as Anselm Strauss, coding can refer to

a different kind of orientation toward one's data from that implied by data reduction. The general analytic approach here is not to simplify the data but to open them up in order to interrogate them further, to try to identify and speculate about further features. Such data complication is not used to retrieve and to aggregate instances to a restricted number of categories; rather, it is intended to expand the conceptual frameworks and dimensions for analysis. Coding here is actually about going beyond the data, thinking creatively with the data, asking the data questions, and generating theories and frameworks.

In practice, coding usually is a mixture of data reduction and data complication. Coding generally is used to break up and segment the data into simpler, general categories *and* is used to expand and tease out the data, in order to formulate new questions and levels of interpretation. One should try to ensure that coding does not lose more than is gained. It is especially important to avoid the use of coding merely to apply simple and deterministic labels to the data. Data reduction or simplification of that sort is not the main analytic purpose of qualitative coding. Coding should be thought of as essentially heuristic, providing ways of interacting with and thinking about the data. Those processes of reflection are more important ultimately than the precise procedures and representations that are employed.

Seidel and Kelle (1995, p. 58) capture this by saying that "codes do not serve primarily as denominators of certain phenomena but as *heuristic devices* for discovery." This is apparent whichever model or approach to coding is adopted. Take, for example, the approach of Tesch (1990). Tesch describes qualitative analysis in terms of decontextualization and recontextualization. Decontextualizing data involves segmenting portions of data and slicing up the data set. Tesch defines segmenting as dividing data into portions that are comprehensible by themselves and large enough to be meaningful. Decontextualization means separating data extracts from their original context while retaining meaning. Segmented data are then organized and sorted as part of a process of recontextualization. Tesch (1990) suggests that the first step of sorting

> consists of tagging text segments with information about the category of the organizing system into which it belongs (or several categories if the segment is relevant to more than one). Many researchers call this process "coding." (p. 121)

Tesch (1990) suggests that once data segments have been coded, they are still not ready for interpretation. Drawing on the work of Marton (1986), who argues that each quotation has two contexts—the one from which it was taken and the "pool of meaning" to which it belongs—Tesch suggests that an organizing system for data is based on developing pools of meaning. Concepts are identified or constructed from prior material, theoretical frameworks, research questions, or the data themselves. The segmented data are coded according to those organizing categories and then re-sorted, again according to those categories. Data segments are reassembled or recontextualized. Coding as part of this process aids the recontextualization of data, giving a new context for data segments. In this way, Tesch regards coding as a means of providing new contexts for viewing and analyzing data. Decontextualizing and recontextualizing help to reduce and then expand the data in new forms and with new organizing principles. To put it another way, segmenting and coding data enable the researcher to think about and with the data.

Strauss (1987) provides perhaps the best example of using coding to complicate and expand qualitative data. We discuss the work of Strauss and his collaborators in more detail later in this chapter. For now, it is important to note that Strauss advocates coding as an essential analytical procedure. He argues that qualitative researchers must learn to code well and easily. Strauss also is keen to stress that coding is often misunderstood to be a simple and unproblematic procedure. The argument here is that coding is much more than simply giving categories to data; it is also about conceptualizing the data, raising questions, providing provisional answers about the relationships among and within the data, and discovering the data. Strauss argues that coding should be used to open up the inquiry and move toward interpretation. Coding is thus about breaking the data apart in analytically relevant ways in order to lead toward further questions about the data. To paraphrase Strauss (1987), coding can be viewed as a way toward the excitement and inevitable payoff of grounded conceptualization.

We can get ideas for coding from a variety of sources that are not mutually exclusive. We can start with a simple framework for coding based on what we as researchers are interested in. Reading through data extracts, one might discover particular events, key words, processes, or characters that capture the essence of the piece. Alternatively, one might code the data extracts using a code list created prior to reading the data. Miles and Huberman (1994) suggest that this method, of creating a "start

list" of codes prior to reading the data or even prior to the fieldwork, is a useful way of beginning to code. These codes or categories can come from a variety of sources. For example, we can start from our theoretical or conceptual frameworks—coding data according to key concepts and theoretical ideas. We might have hypotheses that could be used to select code words to identify segments of the data, in order to test or modify those ideas. Equally, we could start with preselected codes taken from our reading in the general area, or a comparative area, or previous studies. Key variables and concepts can be derived from the research literature.

Another way of beginning to code is to start from the foreshadowed research question that inspired the research project. One might begin with the data and categorize them in a more inductive fashion, starting with the local categories of the actors or informants themselves. We can thus categorize the data more in accordance with the indigenous terms and categories of the culture or the individual informants. We will try to illustrate such thought processes when we explore some of our own data later in this chapter.

It is worth stressing here that codes are organizing principles that are not set in stone. They are our own creations, in that we identify and select them ourselves. They are tools to think with. They can be expanded, changed, or scrapped altogether as our ideas develop through repeated interactions with the data. Starting to create categories is a way of beginning to read and think about the data in a systematic and organized way.

What to code, or what categories to create, will always partly depend on the intent of your data analysis. Strauss (1987) makes the distinction between sociologically constructed codes and in vivo codes. The latter refer to codes that derive from the terms and the language used by social actors in the field, or in the course of interviews. The systematic use of in vivo codes can be used to develop a "bottom up" approach to the derivation of categories from the content of the data. Initial coding, then, should help us to identify themes, patterns, events, and actions that are of interest to the researcher and that provide a means of organizing data sets. Coding can be more or less complex, depending on the level of analysis.

## The Coding Process

The segmenting of data using codes or categories, as we have indicated, can be achieved in a variety of ways, through the application of a

variety of analytical strategies. For the purposes of illustration, we have taken an extract from an interview with an established academic in anthropology. The interview extract focuses on the question of what makes a good doctorate. Set out below is the interview extract, to which we have attached a number of coding categories.

*Extract From Interview With Dr. Fitton (Kingford University)*

*Odette Parry:* What do you think makes a good PhD? — Good PhD

*Dr. Fitton:* I think PhDs should show a substantial contribution to research, but I don't think that necessarily means innovation for innovation's sake. — Good PhD; contribution; originality

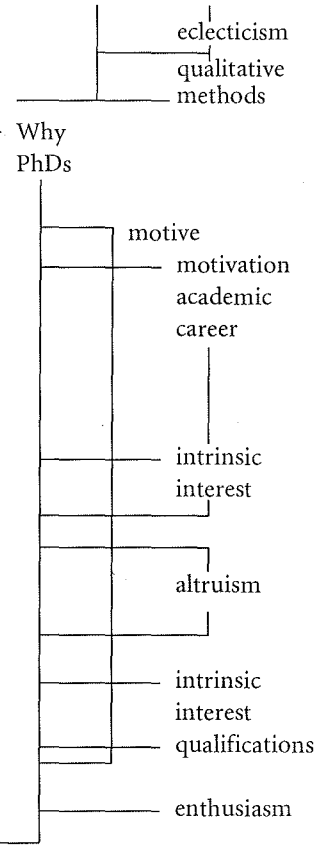
I personally would want to favour a PhD which showed a very sound knowledge of theoretical positions, an ability to sort out those positions and put forward something in a logical, coherent, structured fashion. I'd favour someone who was able to do that over someone who has studied something that no one had thought of studying before, and you're encouraging something that is peripheral, marginal, not necessarily of significance. So I think that what I would look for is a very sound acquaintance with theoretical work, an ability to sort it out, and take it further—have a sufficient substantial commentary on that work, combined in the case of anthropology with fieldwork, and showing that the fieldwork had been done in a way which shows empathy with the people you'd studied, and that the fieldwork and the theoretical part had been merged together. — Ability; theory; logic; originality; theory; fieldwork; theory

Quite a tall order. I'd look for a "feel" about the work, I wouldn't have a list of guiding points, because I don't think you can do that—they are too different. It has — indeterminate knowledge

been said that the strength of anthropology is its eclecticism, it relies on qualitative analysis rather than quantitative.

*Odette Parry:* This is really a general question. Why do you think people do anthropology PhDs?

*Dr. Fitton:* In some cases it's the obvious reason that doing a PhD will hopefully lead to the first rung of the academic track. My own motivation was not that clear. I was surprised when I did get a job at the end of it, but to further an interest I wanted to take as far as I could. I expect most people doing a PhD are doing it to further an interest they have. There seems to be a trend towards PhDs written to do with development, so you could say that a concern for other societies is another factor. So it's not just a selfish endeavour. I can think of one student I've had, the interest in doing a PhD wasn't there, there was an external push, she was expected to get high qualifications. And because her heart wasn't in it, she didn't have the necessary enthusiasm and drive for it.




As shown here, with a relatively simple approach to the process of coding, different levels of complexity can be explored. What we have done here, in fact, is summarize a series of successive decisions about the data and its categorization. When they are superimposed in one display, it easily can look as if they were all derived simultaneously, from the same set of interests and concerns. That is not the case, and here we try to indicate some of the decisions that have gone into such a coding outcome.

At the simplest level, the data can be reduced to two possible generic categories: "a good PhD" and "why people do a PhD." These reflect directly the questions that Odette Parry asked and reflect two of the substantive problems that the research team brought to the data-collection

exercise. The research was intended to examine some of the personal and institutional factors that influence the PhD process, including views on the PhD itself as a means of academic socialization. Clearly, coding the data according to those themes adds nothing initially to our understanding of the data. It is essentially a data-reduction task. Segmenting and coding the data in that particular way would at least allow us to characterize what each stretch of the interview was about in terms of general thematic content, in this instance relating directly to the topics of the interview elicitations and responses. Such wide, generic categories would facilitate the retrieval of different segments of data that deal with descriptions of good PhDs and academics' speculations as to doctoral candidates' motivations.

The application of these and equivalent codes, reflecting substantive research questions, would be one basic way of organizing the data. Such procedures have considerable practical value. The nature of qualitative data means that data relating to one particular topic are not found neatly bundled together at exactly the same spot in each interview (and fieldnotes usually have even less predictable organization). The ability to locate stretches of data that, at least ostensibly, are "about" the same thing is a valuable aspect of data management. Such coding, therefore, can be a useful preliminary to more detailed analysis. We could proceed a little further in the same vein. Another way of thinking about this particular data extract would be to "code" in terms of specific abilities or competencies identified by the informant. That is, as we have indicated, the first half of the interview extract definitely describes a number of abilities or attributes linked to the production of a good PhD. Coded as "ability," this data segment could be used to search for and compare other aspects of the data set where graduate students' abilities are referred to. For example, PhD supervisors might in other contexts refer to the abilities required for supervising doctoral work or for being a successful academic more generally. Likewise, doctoral students in interviews might also describe their experience in terms of the abilities they brought to their studies, developed over the course of their research, failed to develop, and so forth. Each category of actor might refer to their own competence and that of other actors. As is clear, these very general categories promote the reordering of the data in accordance with preliminary ideas or concepts. They are not necessarily the final ways in which the data can be examined and explored. This first approach to the

data, however, does little or nothing toward complicating the questions we can ask about them. If we are going to use coding to generate more interesting and complex ideas about our materials, then we need to do something more. Most fundamentally, we need to think more about how we interact with the data.

Coding at such a very general level is a first step toward organizing the data into meaningful categories. It can be seen as the first level of coding. As we demonstrate in the data extract above, coding also can be thought about in a more complex way. Using a *good PhD,* or *ability,* or *why a PhD* as initial codes or categories, a number of subcategories can be generated and used to segment the data. In this extract, we have identified a number of such categories and attached codes accordingly. Some of those more detailed codes come more or less directly from the informant's words, such as *enthusiasm.* Others are our summary glosses of what the informant seems to be referring to or describing at a particular point in the text. For example, we have glossed one of the motivating factors as *altruism,* not the actual word used by the informant but used by us to capture descriptions of motivation based on or ascribed to the desire to help others or a commitment to another culture. Other codes reflect more directly *our* conceptual interests. For example, we have categorized one segment as referring to *indeterminate knowledge.* Dr. Fitton is talking about evaluating a PhD with reference to its "feel" and talks about a lack of specific guidelines. Our identification of the segment in this way, and our decision to code as we have, reflects our own interest in pursuing how anthropologists talk about their own knowledge and the knowledge of other anthropologists, past and present. Here, therefore, we note an appeal to an apparently indeterminate criterion in evaluating doctoral students' work. In coding it as we have, we can collate all the instances where similar appeals to indeterminate knowledge have been made, talked about, or denied. We can inspect those instances further to examine the varieties of indeterminate knowledge and their sources, how indeterminate knowledge is evaluated, how it is learned, and so on.

These more detailed subcategories can be represented by a single code attached to a discrete segment of the data. Subcategories also can overlap one another. Codes and their segments can be nested or embedded within one another, can overlap, and can intersect. The same subcategory can be applied several times in a single unit of data (such as an

interview), and the same segment can have more than one code attached to it. In the data extract above, the dense nesting and overlapping of the more detailed codings can be seen. These dense patternings are quite characteristic of code maps of qualitative data. After all, ordinary social action, including conversational talk, does not present itself to us in neatly bounded packages. When we segment the data by attaching codes, we often reflect how topics run into one another and how there may be multiple issues to concern ourselves with simultaneously.

Coding is never a mechanistic activity. Because of our selection of a data extract, we hope that the outcomes are fairly transparent, but they are by no means automatic. We need to *decide,* for example, not only what aspects of the data to tag with codes but also what levels of generality or detail to go into. As we have indicated in the extract above, we have identified and defined three levels of generality. The most general categories are two: what makes a good PhD and the motivation to undertake a PhD. In this instance—and by no means is this always true—the categories correspond with the thrust of questioning in the interviews and were part of the agenda followed in the semistructured interviews themselves. We have here identified an intermediate category, having to do with the abilities or competencies required for the successful completion of a PhD. The third and most specific level of category breaks down those more general themes into more specific and detailed codes. As can be seen from how we have marked those codes on the extract, they also relate to stretches of the interview of different lengths. The more detailed, specific codes are embedded within the longer, more general ones. This is a feature of dense, detailed coding that becomes especially important in the computer-assisted handling of qualitative data, which we discuss in Chapter 7. The identification of codes and decisions about levels of detail are far from straightforward. As Weaver and Atkinson (1994, p. 32) put it:

> [I]f we decide to delineate a number a number of general, inclusive categories, much of the text will be coded with a single code (or conjunction of codes). The advantage of this strategy is that it should maximize the usefulness of the codes; they are likely to be applied to enough segments to justify the purpose of recontextualization. However, it may also have several disadvantages. First, since so much text will be coded with the same category, there might be difficulty in locating particular episodes significant to analysis; a likely scenario is that the researcher will have to siphon through reams of

irrelevant data, despite recontextualization. Second, coding may be too crude, and this might make the analysis seem rather vague, lacking detail, or the exploratory avenues of analysis being superficially restricted.

On the other hand, coding schemes that are too detailed can be equally problematic:

> [I]f we decide to define a large number of categories, with fairly exclusive meaning, the problems are reversed. Coding will be more detailed and intricate, and there will be a greater differentiation of segments accordingly. However, if the segmentation of text is too intricate, in that specific categories are attached to very small segments of text, important contextual information may be lost, and thereby some of the segments' meanings. (Weaver & Atkinson, 1994, p. 32)

As a consequence, Weaver and Atkinson made an explicit decision to include codes of different degrees of generality so that data retrieval could be undertaken at different levels.

We can develop and illustrate decision-making processes, as well as the significance of different levels of coding, with another extract from the interview data. As will be seen, it deals with a set of issues similar to those identified from our first extract.

*Extract From Interview With Dr. Throstle (Southersham University)*

***Odette Parry:*** What sorts of skills do you think it imparts, the actual process of doing a PhD?

***Dr. Throstle:*** It's a very big question. Again I think in anthropology you learn a whole lot of things that you don't normally learn in a PhD, which is partly to do with fieldwork. That trains you to carry on on your own both academically and personally, it's social skills training of a very exacting kind.

. . .

I think one of the peculiarities of anthropology is that unlike most other disciplines, certainly in the social sciences, you're dealing—unlike history for example—you don't start from one body of documentation and convert it into another kind of body of documentation, you start with people's

lives and conversations with them and you have to turn that into an academic text.

Which is why it takes such a long time and why it's so difficult, because these two things are miles and miles apart. And it's very common, I think, for graduate students when they come back from the field to react against what they're doing, to feel that what they're writing is somehow a betrayal or it falls far short of the relationships they had when they were in the field. Writing a rather dull piece of academic work somehow feels like a betrayal.

Nevertheless most students learn to do that, and in the process learn an immense amount not only about the people they study, but also writing skills and how to produce a high-level academic text.

I think one of the problems of fieldwork, if you're away for a long time—and most anthropologists are—you lose touch to some extent with your academic and your home culture, and then you have to get back into it, and it's often a slow process when you come back.

It would be easy to treat this extract in much the same way as we did the first extract, and to deal with it initially in terms of Odette Parry's elicitation. We could thus relate it primarily to the kinds of skills that the question asked about. Below, therefore, we have categorized the data extract accordingly. It will be seen that we allowed ourselves to stick closely to the overarching theme implied in the question about skills.

| | |
|---|---|
| *Odette Parry:* What sorts of skills do you think it imparts, the actual process of doing a PhD? | skills from PhD |
| *Dr. Throstle:* It's a very big question. Again I think in anthropology you learn a whole lot of things that you don't normally learn in a PhD, which is partly to do with fieldwork. That trains you to carry on on your own both academically and personally, it's social skills training of a very exacting kind. . . . | academic<br>personal<br>social |

| | |
|---|---|
| I think one of the peculiarities of anthropology is that unlike most other disciplines, certainly in the social sciences, you're dealing—unlike history for example—you don't start from one body of documentation and convert it into another kind of body of documentation, you start with people's lives and conversations with them and you have to turn that into an academic text. | writing |
| Which is why it takes such a long time and why it's so difficult, because these two things are miles and miles apart. And it's very common, I think, for graduate students when they come back from the field to react against what they're doing, to feel that what they're writing is somehow a betrayal or it falls far short of the relationships they had when they were in the field. Writing a rather dull piece of academic work somehow feels like a betrayal. | writing |
| Nevertheless most students learn to do that, and in the process learn an immense amount not only about the people they study, but also writing skills and how to produce a high-level academic text. | writing |
| I think one of the problems of fieldwork, if you're away for a long time—and most anthropologists are—you lose touch to some extent with your academic and your home culture, and then you have to get back into it and it's often a slow process when you come back. | |

We can see here that so far we have produced a very thin and flat set of categories. They reproduce only the bare bones of skills and do not appear to do justice to the dense descriptive language of this particular academic. Another way of approaching the task, therefore, is to pay much closer attention to the categories of expression that the informant actually uses. Rather than using the interview extract as an extended reply to our one question, therefore, we pay much closer attention to the content of the talk. Here, therefore, we approach the data once more, and now try to identify themes that reflect the informant's views more closely. It is not necessary for us to use precisely the same words to index,

or code, those themes; we are interested in exploring them and linking them with other data segments, not only in labeling them.

| | |
|---|---|
| *Odette Parry:* What sorts of skills do you think it imparts, the actual process of doing a PhD? | skills from PhD |
| *Dr. Throstle:* It's a very big question. Again I think in anthropology you learn a whole lot of things that you don't normally learn in a PhD, which is partly to do with fieldwork. That trains you to carry on on your own both academically and personally, it's social skills training of a very exacting kind. . . . | fieldwork<br>normal PhD<br>academic independence<br>personal independence<br>social skills<br>exacting |
| I think one of the peculiarities of anthropology is that unlike most other disciplines, certainly in the social sciences, you're dealing—unlike history for example—you don't start from one body of documentation and convert it into another kind of body of documentation, you start with people's lives and conversations with them and you have to turn that into an academic text. | peculiarity<br>disciplines<br>social science<br>history<br>people<br>lives<br>conversation<br>academic text |
| Which is why it takes such a long time and why it's so difficult, because these two things are miles and miles apart. And it's very common, I think, for graduate students when they come back from the field to react against what they're doing, to feel that what they're writing is somehow a betrayal or it falls far short of the relationships they had when they were in the field. Writing a rather dull piece of academic work somehow feels like a betrayal. | time<br>difficult<br>difference<br>return from field<br>writing<br>relationships in the field<br>dull<br>academic work<br>betrayal |
| Nevertheless most students learn to do that, and in the process learn an immense amount not only about the people they study, but also writing skills and how to produce a high-level academic text. | people<br>writing<br>high-level<br>academic text |

| | |
|---|---|
| I think one of the problems of fieldwork, if you're away for a long time—and most anthropologists are—you lose touch to some extent with your academic and your home culture, and then you have to get back into it, and it's often a slow process when you come back. | fieldwork<br>time<br>absence<br>academic culture<br>home culture<br>return |

In indexing the data in this kind of way, we can start to develop a much denser set of themes and categories. We can start to glimpse some of the recurrent preoccupations of this particular anthropologist, and we can use such categorizations to build systematic comparisons and contrasts with the views expressed by other faculty members. In looking at the data in this way, moreover, we can start to identify some further themes and issues.

Having begun by staying close to the informant's own categories, we can start to see how they might be categorized further, possibly in relation to linking categories of our own devising. Take, for example, the first paragraph of the informant's reply. It will be seen that here she is prefacing a response about "skills" by saying that there is something distinctive about anthropology as a discipline that makes it different from others, certainly among the social sciences. This helps us to identify a superordinate category, which we might identify as *the distinctiveness of anthropology*. If we identify that as a category and code the data accordingly, then we can use that to search for other data extracts in which the anthropologists express what is special and distinctive about their subject.

There is also a potentially intriguing theme to be constructed from this anthropologist's words. Writing is identified as a major academic skill to be acquired, but that thought is expressed in a particular, and striking, way. Dr. Throstle talks of the difference between the academic text that the student must prepare and the people who were studied. People's lives and conversations have to be turned into text, and students find the essential difference between the social and the textual to be a problem. Not only is this a very interesting comment about anthropology, but it also connects with our previous category in two ways. First, it is offered as an example of how anthropology differs from other disciplines (the requirement of constructing a body of text out of lives and conversations). Second, it picks up on the significance of difference in a new way.

Let us look ahead in the data once more before commenting further. The informant talks about the importance of fieldwork. Fieldwork is introduced as one of the distinctive aspects of the anthropology PhD at the beginning of the extract. The topic reappears later. Toward the end, for example, fieldwork is contrasted with coming back, and there is a contrast between fieldwork, on one hand, and academic culture and home culture on the other. Again we see Dr. Throstle describing things in terms of difference, in this case differences between cultures. We can see now that there is a potential superordinate category relating to difference that relates to all these aspects of the interview extract. The differences that are described are different in content, but they all seem to relate to a coherent set of underlying issues: the distinctiveness of anthropology, the significance of fieldwork, and the separation of fieldwork from other aspects of the anthropologist's life. We should note that this theme, which relates closely to the content of the data, is clearly one that we have constructed. It is also one that takes us toward concepts of a more analytic, even theoretical, relevance. We have thus moved our coding process from identifying categories that remain close to the original data to those that imply much broader analytic issues. We may therefore return to the data once more and apply a further set of codes.

*Odette Parry:* What sorts of skills do you think it imparts, the actual process of doing a PhD?

*Dr. Throstle:* It's a very big question. Again I think in anthropology you learn a whole lot of things that you don't normally learn in a PhD, which is partly to do with fieldwork. That trains you to carry on on your own both academically and personally, it's social skills training of a very exacting kind. . . .

distinctiveness of anthropology

contrastive rhetoric

significance of fieldwork

I think one of the peculiarities of anthropology is that unlike most other disciplines, certainly in the social sciences, you're dealing—unlike history for example—you don't start from one body of documentation and convert it into another kind of body of documentation,

contrast anthropology with other disciplines

difference between the field and academic texts

| | |
|---|---|
| you start with people's lives and conversations with them and you have to turn that into an academic text. | |
| Which is why it takes such a long time and why it's so difficult, because these two things are miles and miles apart. And it's very common, I think, for graduate students when they come back from the field to react against what they're doing, to feel that what they're writing is somehow a betrayal or it falls far short of the relationships they had when they were in the field. Writing a rather dull piece of academic work somehow feels like a betrayal. | difference between field and academic texts<br>metaphor: betrayal |
| Nevertheless most students learn to do that, and in the process learn an immense amount not only about the people they study, but also writing skills and how to produce a high-level academic text. | |
| I think one of the problems of fieldwork, if you're away for a long time—and most anthropologists are—you lose touch to some extent with your academic and your home culture, and then you have to get back into it, and it's often a slow process when you come back. | difference between the field and academic culture<br>difference between the field and home culture |

In attaching codes in this way we have indicated some key generic issues. That work has been done as more than a mechanistic exercise in segmenting the data. It reflects a series of readings and re-readings of the data, in which the details of the interview and our own emergent concerns interact. We may also note while dealing with the interview text at this level that the repeated references to differences and distances suggest two further issues. In identifying these categories, we have also started to move from a consideration of the content of the anthropologist's talk in order to pay some attention to its form. We can note that his or her descriptions draw on a recurrent pattern, a series of contrasts: anthropology/other disciplines, fieldwork/home, and people's lives/academic

texts. In Chapter 4, we will consider in more detail how we can examine such formal properties, and we will discuss this particular property again when we consider *contrastive rhetoric* (Hargreaves, 1984) as one device that actors use in producing their accounts of the social world.

In considering form and content here, we have noted one further aspect of this academic's talk. We have identified the graphic way in which the distance between field and text is expressed. We are told that student anthropologists may feel that their work is a *betrayal.* Again, we will return to discuss the exploration of such figures of speech in Chapter 4, but we note that it may prove useful to identify figures of speech such as metaphors during the process of coding. We have therefore added two codes to the extract above, one identifying the use of contrastive rhetoric, the other identifying the location of this particular metaphor.

We have been through this one extract from our data and applied to it different, complementary sets of codes. For the sake of clarity, we have displayed the data and the codes separately. In the course of repeated examinations of the data for the purposes of a comprehensive analysis, we would not necessarily have recourse to such separation. We can think of the successive passes at the data as resulting in overlays of different codes, reflecting different levels of specificity or generality as well as reflecting different sets of analytic themes. These different approaches to the data could result in different physical disaggregations of the data: physically cutting up different copies or cutting and pasting segments into different files with the word processor. If we were using computer software to perform these tasks, we would be able to retrieve the coded segments by using different codes or combinations of codes to identify them.

Our illustrative example is not an exhaustive treatment of the data. Any other analyst could conceptualize them in different ways. The point is not to search for the "right" set of codes but to recognize them for what they are: links between particular segments of data and the categories we want to use in order to conceptualize those segments.

## Beyond Coding and Toward Interpretation

Coding qualitative data enables the researcher to recognize and re-contextualize data, allowing a fresh view of what is there. Because coding

inevitably involves the reading and re-reading of data and making selections from the data, it involves interpreting the data set. However, a key issue is what to do with data once they have been selected, cut up, fragmented, coded, and categorized. The move from coding to interpretation is a crucial one, as Wolcott (1994) suggests. Interpretation involves the transcendence of "factual" data and cautious analysis of what is to be made of them.

Once coding is achieved, the data have to be interrogated (Delamont, 1992) and systematically explored to generate meaning. There is a case that coding, while reorganizing data, also involves a certain amount of information loss. To some extent, that depends on how thorough and detailed the coding has been. The data loss is much greater if one does not move from the process of coding to an exploration of how codes and categories relate to the original data, to other data, to theoretical ideas, and so forth. Ian Dey (1993) argues that categorizing enables one to think about the data in a new way. This is only the case if we move beyond the codes, categories, and data bits back to what the "whole" picture is or may be.

The move from coding to interpretation has a number of discrete levels. First, the coded data need to be retrieved. This essentially means that recontextualized data need to be displayed in such a way that they can be read easily. The data bits that relate to a particular code or category need to be presented together in order for the researcher to explore the composition of each coded set. Huberman and Miles (1994) argue that data display is a key element of the analytical process. This can be achieved by organizing all the data under a particular code physically in the same place; by producing diagrams, matrices, and maps of the code; or by using a retrieval function on a microcomputing program. Whichever way it is done, the idea is that the codes or categories and the data need to be in such a form that they are accessible both for reading and for exploring.

Second, the move from coding to interpretation involves playing with and exploring the codes and categories that were created. Dey (1993) provides many ideas about how you can go about doing this. He suggests that once data are displayed in a coded form, the categories can be retrieved, split into subcategories, spliced, and linked together. Essentially, the codes and categories you have selected should be used to make pathways through the data. It is worth remembering here that such codes are *not* cast in stone. As you chose and selected them, so you can abandon, change, re-sort, rename, and so on. Similarly, once you are in

a position to look at all the data across the codes, you should not be tempted to ignore incidents, events, individuals, or chunks of data that do not "fit" into the codes. The exceptions, misfits, and "negative" findings should be seen as having as much importance to the process of coding as do the easily coded data.

This leads to a further level of the process of moving from coding to interpretation, that is, the transformation of the coded data into meaningful data. Here, the emphasis is on what to look for in the codes and categories. Delamont (1992) suggests that one should be looking for patterns, themes, and regularities as well as contrasts, paradoxes, and irregularities. One then can move toward generalizing and theorizing from the data. The emphasis on the "negative" exceptions as well as the "positive" patterns remains crucial. Huberman and Miles (1994) work from a similar continuum. They suggest no less than 13 "tactics" for generating meaning or data transformation. These move from descriptive through explanatory tactics. At the one end of the continuum are things such as noting patterns of themes, the "counting" of phenomena occurring from the data, and comparing and contrasting the data sets. At the opposite end are the moves toward generalizing, noting and questioning the relations between variables, and finding conceptual and theoretical coherence in the data.

We already have illustrated how codes can represent categories of different sorts. Some of those already imply interpretive frameworks and link data segments to emergent concepts. Hence, we do not need always to think of coding first and theorizing afterward. Our decision making implies analytic ideas at every stage of the coding process. Furthermore, we already have seen that our codings can imply systematic relationships among categories and concepts. These relationships can form one basis for the development of interpretations. For example, if we return to the second of our data extracts, we can see that our codings start to suggest possible relationships. The different levels of coding suggest that some categories may subsume a number of others. This is so for several different codings. For example, Dr. Fitton's responses in the first extract suggest the following kinds of relationships:

| | |
|---|---|
| Student motivation | Academic career |
| | Intrinsic interest |
| | Altruistic commitment |

| | |
|---|---|
| Good PhDs | Originality<br>Theoretical knowledge<br>Logical coherence<br>Quality of fieldwork |

In a different way, we have seen how Dr. Throstle's coded comments can be linked through a common use of difference and distance, which could linked in the coding scheme through a code of *contrast.* In other words, our codes not only establish linkages between data and concepts; they also can start to map out dimensions within conceptual categories and to establish superordinate links among concepts. We can summarize this set of more abstract ideas as follows:

| | |
|---|---|
| Difference | Home culture/Field<br>Field/Academic culture<br>People's lives/Academic texts<br>Anthropology/Other disciplines |

We will suggest in Chapter 6 that although there is more to it than linking codes, the establishment of ordered relationships between codes and concepts is a significant starting point for reflection and for theory building from qualitative data.

Strauss (1987) develops the use of coding as part of the process of interpretation and analysis. That is, Strauss links the initial process of coding (which he refers to as *open coding* and which is essentially what we did with the interview extract with which we exemplified our earlier discussion) to a more refined process of using categories to generate broader conceptual frameworks. Strauss identifies a set of procedures that allows initial categories to be elaborated and developed.

We are not going to follow Strauss's own model closely. We do not recapitulate all the steps and analytic strategies that he identifies or has codified in his work with Corbin (Strauss & Corbin, 1990). Indeed, it is worth noting in this context that Glaser (1992) accuses Strauss of taking the general inspiration and strategic approach of "grounded theory" and transforming them into unduly prescriptive recipes for analysis. Nevertheless, the general approach to coding that is to be found in Strauss's methodological writing is valuable in encouraging the researcher to move beyond local codings to generate ideas and broader conceptual frameworks.

In essence, Strauss's approach encourages us to go beyond the essentially "summary" approach to coding in which the data are simply reduced to a limited set of categories. Strauss does not encourage us simply to index the data, as it were, or to use the code words merely to mark and retrieve segments of data. On the contrary, his general approach exhorts us to expand on rather than to reduce the data, to take categories and exhaust their full analytic potential. One point is to use our codings and categories to think with and not to remain anchored in the data (notes, transcripts, etc.) alone.

From this general perspective, the process of coding is about asking oneself questions about the data. Those questions help to develop lines of speculation and hypothesis formation. In accordance with more general principles of grounded theory, they may also direct further data collection strategies. In the course of open and axial coding, then, one takes as a topic a "phenomenon" (in Strauss's terms) and attempts to identify its dimensions, its consequences, and its relationships with other phenomena.

For example, if we return to our previous data extract, we start to think more creatively about some of the themes that we have begun to identify in our coding. Consider once more the same interview extract (see pp. 38-39). When we first coded that, we were able to identify stretches of the transcript that seemed to refer to theoretical sophistication, analytic value, originality, fieldwork, indeterminate knowledge, contribution to the discipline, and empathy with the people studied. These were all in an informant's response to the question of what makes a good PhD, and all those separate code words, and the fragments they refer to, all seem to bear on one major phenomenon, that is, the essence of what it takes to produce a successful PhD in social anthropology. We can, therefore, start to think about the various possible "dimensions" of such abilities or competencies.

In this one account, we find the established anthropologist constructing his own characterization of how the competencies are (or should be) combined in one ideal-typical doctoral candidate. In thinking more about this phenomenon, we need to ask ourselves, and to ask of the data, what sorts of abilities or competencies are recognized and described. Our first informant has given us the abilities to theorize, to organize and structure coherent arguments, to empathize with the people studied, to make an original contribution to the discipline, to do fieldwork, and to

integrate fieldwork and theory. Further, we note that these abilities are generally characterized as indeterminate qualities rather than being rendered explicit. In looking at other interviews, we note that other categories and codings bear on the same general phenomenon. They are expressed in different ways, for example, the abilities to turn other people's lives and conversations into academic text, to do scholarship, to be original, to be critical, to add to human understanding, to put a new slant on what's been written before, and to contribute to the ongoing debates. Successful anthropologists should also have a distinctive approach and exhibit openness, humility, and the ability to reflect upon themselves. The open coding of a series of interviews would therefore give us a wide range of dimensions of "ability" or "competence" or "quality" (the general phenomenon is very general indeed).

We can also go on to think about some further features. For example, under "axial" coding, Strauss recommends thinking about such features as *consequences.* Following that line of reasoning, we can go beyond the data immediately at hand and ask ourselves such further questions as "What happens if . . .?" For example, we can ask "What happens if there isn't any fieldwork?" That question can inform further inspection of the data, further data collection as the research develops, or both. Inspection of our anthropology data suggests that there may be different answers to that question, depending on the antecedents or causes of fieldwork being absent (another of Strauss's features of axial coding). For example, fieldwork may not take place because the graduate student has decided to pursue "library research" (a term used by our informants). Further exploration of consequences and antecedents shows that library research is often discouraged, and the absence of fieldwork, though deliberate, may be stigmatizing. On the other hand, fieldwork may be incomplete because of conditions beyond the student's control (such as civil disorder). This is not stigmatizing, though it may prove to be a major handicap and is unfortunate. Such observations again help us to open up questions about the significance of fieldwork in the construction of a professional career and an academic identity.

In the same vein, we can ask ourselves what the consequences are for the writing and organizing of the thesis. What are the consequences of not having the ability to integrate theory and fieldwork, or of not making a "contribution" to the area of specialization? Such a reflection perhaps would take us to other aspects of the data. We might look for academic

supervisors' accounts of how PhD theses get examined and what aspects of organization and style are rewarded or penalized. We also need to explore the students' accounts of their writing, looking for such things as their strategies for writing, their experiences of writer's block, and their strategies for writing. We need to examine our data (or, ideally, collect further data) on how our anthropologists recognize and reward "a contribution" to the discipline or to a particular debate within it. In procedural terms, this means cross-reference to the "open" codings identified elsewhere in the data. In intellectual terms, it means using the various elements in the data to pursue lines of interrogation and speculation—moving between data and codings to explore and expand on key analytic themes.

We might also follow others of Strauss's analytic strategies in coding—thinking about conditions and antecedents, perhaps. We can ask ourselves, for example, about the antecedents of fieldwork. This approach would probably lead us to think about and to search out data concerning the necessary preconditions (organizational, personal, academic) to embarking on fieldwork. We would then find ourselves asking—and interrogating the data—about when fieldwork is possible and who gets to decide about it (where, under what circumstances, with what intellectual preparation). We would also start to explore some of the more mundane antecedents and conditions, the practical tasks and relationships that go into the practical work of fieldwork. We might thus start to generate ideas and themes that lead us to think seriously and systematically about academic and practical dimensions of anthropological fieldwork. One would hope to be generating themes that facilitated comparative thinking and exploration, for example, contrasting with other travelers to exotic parts, such as war correspondents, travel writers, workers in international transport, or others who live on a long-term basis away from their home base and creature comforts.

## Conclusion

In this chapter, we have explored some of the rationales for coding qualitative data and introduced some of the different approaches to be found in the methodological literature and in practice. This discussion has not been intended as a comprehensive cookbook as to how to

perform coding. We have preferred to discuss contrasting and complementary strategies rather than prescriptively recommending a single orthodoxy. We also have tried to suggest from time to time that "coding," however it may be conceptualized and carried out, is not the final word on qualitative data analysis.

The segmentation of field data and retrieval of marked data segments is a valuable resource in the management of qualitative data. It is an established approach that in recent years has been reinforced by the development of microcomputing strategies (Weaver & Atkinson, 1994), many of which essentially recapitulate the same logic of data handling. They substitute rapid and comprehensive searching supported by software for the uncertain and slow process of manual searching and filing. In and of themselves, however, such procedures by no means exhaust either the possibilities of the data or, therefore, the possibilities of data exploration. In particular, it should be apparent that the fragmentation of data, dependent on code-and-retrieve approaches, or what Tesch (1990) called the decontextualization of data, does little to preserve formal features of those data.

Our interview informants may tell us long and complicated accounts and reminiscences. When we chop them up into separate coded segments, we are in danger of losing the sense that they are accounts. We lose sight, if we are not careful, of the fact that they are often couched in terms of stories—as narratives—or that they have other formal properties in terms of their discourse structure. Segmenting and coding may be an important, even an indispensable, part of the research process, but it is not the whole story. Consequently, we turn to a consideration of narrative analysis in the chapter that follows. We will see how one can look at the same sort of data from a fresh perspective, paying due attention to its more formal properties and being sensitive to the *storied* quality of many qualitative data.

## Suggestions for Further Reading

Dey, I. (1993) *Qualitative data analysis: A user friendly guide for social scientists.* London: Routledge and Kegan Paul.

*An unusually detailed description of the processes of categorizing and coding qualitative data. Illustrated with data extracts from Woody Allen and Victoria Wood scripts.*

*Addressed to the management of coded data on computer. Dey is the developer of the software application called Hyperresearch.*

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

*An exceptionally systematic and formal approach to the classification and representation of qualitative data. A uniquely rich source for analytic ideas, with a strong visual emphasis.*

Strauss, A. L. (1987). *Qualitative analysis for social scientists.* Cambridge, UK: Cambridge University Press.

*An unusually detailed account of Strauss's own approach to data analysis. Illustrated throughout with extracts from Strauss's postgraduate seminars, giving a particularly privileged insight into his preferred approach.*

Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory, procedures, and techniques.* Newbury Park, CA: Sage.

*An accessible introduction to coding procedures, following Strauss's approach to grounded theorizing. Goes beyond coding itself to discuss the development of ideas and theoretical frameworks.*

Tesch, R. (1990). *Qualitative research: Analysis types and software tools.* London: Falmer.

*Combines a general discussion of analytic strategies with a survey of computer software packages for qualitative data management and analysis.*