RQ's to Data | Research Questions | Context & Ethics

INTRODUCTION TO RESEARCH METHODS IN EDUCATION

KEITH F PUNCH

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

Qual Data Analysis

9 The Analysis of Qualitative Data

LEARNING OBJECTIVES

After studying this chapter you should be able to:

- Discuss the diversity of approaches in qualitative data analysis
- Explain what is meant by induction, and by levels of abstraction
- Describe coding and memoing
- Summarize the Miles and Huberman approach to qualitative data analysis
- Summarize the grounded theory approach to qualitative data analysis
- Explain the main ideas behind narrative analysis, ethnomethodological and conversation analysis, discourse analysis, semiotic analysis, documentary analysis

In Chapter 7 the diversity of qualitative research was noted. Perhaps nowhere is that diversity more apparent than in approaches to the analysis of qualitative data. Indeed, the term 'data analysis' itself has different meanings among qualitative researchers, and these interpretations lead to different methods of analysis. We begin this chapter by looking at the present day diversity in qualitative analysis. This is followed by a description of some of the main ideas and approaches in qualitative data analysis. The chapter ends with some general advice about writing the analysis of data section in a qualitative research proposal, in view of the multiple methods available.

9.1 DIVERSITY IN QUALITATIVE ANALYSIS

Qualitative research, in education and other areas, concentrates on the study of human behaviour and social life in natural settings. Its richness and complexity mean that there are different ways of analysing social life, and therefore multiple perspectives and practices in the analysis of qualitative data: 'There is variety in techniques because there are different questions to be addressed and different versions of social reality that can be elaborated' (Coffey and Atkinson, 1996: 14). The different techniques are often interconnected, overlapping and complementary, and sometimes mutually exclusive – 'irreconcilable couples' (Miles and Huberman, 1994: 9). But whether complementary or contrasting, there are good reasons for the existence of the many analytic strategies, since any set of qualitative data can be looked at from different perspectives. A repertoire of analytic techniques thus characterizes qualitative research today, and different techniques can be applied to the same body of qualitative data, illuminating different aspects of it (Example 9.1).

EXAMPLE 9.1 DIFFERENT ANALYTIC TECHNIQUES

Feldman (1995) applies the four techniques of ethnomethodology, semiotics, dramaturgical analysis and deconstruction to the one body of qualitative data, drawn from a single study of a university housing office. The different techniques illuminate different aspects of the data.

Despite this variety, some writers have sought to identify the common features of qualitative data analysis. For example, Miles and Huberman (1994: 9) suggest a 'fairly classic set' of six moves common across different types of analysis – these are shown in this book in Appendix 1. Similarly, Tesch (1990: 95–7), while concluding that no characteristics are common to all types of analysis, nevertheless identifies ten principles and practices that hold true for most types of qualitative analysis. But Tesch also identifies no fewer than 26 different approaches to the analysis of qualitative data in her survey.

This variety and diversity in approaches underlines the point that there is no single right way to do qualitative data analysis – no single methodological framework. Much depends on the purposes of the research, and it is important that the proposed method of analysis is carefully considered in planning the research, and is integrated from the start with other parts of the research, rather than being an afterthought. In the expanding literature on qualitative analysis, terms such as 'transforming', 'interpreting' and 'making sense of' qualitative data are prominent, and it is the different ways of doing these things that lead to the diversity in methods of analysis. This diversity is valuable, but scholarly rigour and discipline are also important. In their book *Making Sense of Qualitative Data*, Coffey and Atkinson (1996: 3) stress: 'What links all the approaches is a central concern with transforming and interpreting qualitative data – in a rigorous and scholarly way – in order to capture the complexities of the social worlds we seek to explain.' A similar point about the need for discipline is made by Silverman (1993: 211).

These recent concerns for disciplined methods of analysis echo this well-known quote from some 30 years ago:

> **The most serious and central difficulty in the use of qualitative data is that methods of analysis are not well formulated. For quantitative data, there are clear conventions the researcher can use. But the analyst faced with a bank of qualitative data has very few guidelines for protection against self-delusion, let alone the presentation of unreliable or invalid conclusions to scientific or policy-making audiences. How can we be sure that an 'earthy', 'undeniable', 'serendipitous' finding is not, in fact, wrong? (Miles, 1979: 591)**

Methods for the analysis of data need to be systematic, disciplined and able to be seen (and to be seen through, as in 'transparent') and described. A key

question in assessing any piece of research is: How did the researcher get to these conclusions from these data? If there is no answer to that question – if the method of analysis cannot be described and scrutinized – it is difficult to know what confidence to have in the findings put forward.

All empirical research has to deal with this problem. One strength of quantitative research is that methods for the analysis of its data are well known and transparent. That enables reproducibility in the analysis – a second analyst, working with the same quantitative data and using the same statistical operations as the first, should get the same results.[1] For qualitative research, the relevance of the criterion of reproducibility is a matter of debate in the literature. But there have been great developments in the analysis of qualitative data in the past 30 years,[2] and the concept of the 'audit trail' through the data analysis is now realistic for much qualitative research.[3]

For the individual researcher, this problem comes alive at the point of sitting down in front of the collected qualitative data – perhaps interview transcripts, and/or field notes from observations and discussions, and/or documents. At this point, what, exactly, does the researcher do? Deciding what to do can cause bewilderment, as Feldman's vivid description shows (1995: 1).

Despite this progress, it would be wrong to assume that all developments in qualitative analysis have been directed at this issue. For one thing, there are researchers who would reject the view of knowledge on which the ideas of reproducibility and the audit trail are based – for example, those devoted to a relativist epistemology rooted in a postmodernist and constructivist philosophy (Kelle, 1995). For another, some more recent developments in qualitative analysis have taken the field in quite new directions, where this criterion has seemed both less central and less problematic. This will be seen in later sections of this chapter.

A survey of methods of analysing qualitative data suggests a division of analytic approaches into general and specialized. The next three sections (9.2–9.4) describe some important general approaches to the analysis of qualitative data, which can be applied across a wide variety of education research situations. Section 9.5 deals with the specialized approach of grounded theory analysis, and section 9.6 then overviews some of the other more specialized approaches.

9.2 ANALYTIC INDUCTION

In the search for regularities in the social world, induction is central. Concepts are developed inductively from the data and raised to a higher level of abstraction, and their interrelationships are then traced out. But while induction is central, deduction is needed also, since, as noted in Chapter 7, theory generation involves theory verification as well. This sort of qualitative data analysis is a series of alternating inductive and deductive steps, whereby data-driven inductive hypothesis generation is followed by deductive hypothesis examination, for the purpose of verification (Kelle, 1995).

The fact that much qualitative analysis depends on induction suggests 'analytic induction' as a useful general term. But this term also has a more specific meaning. The method of analytic induction was developed by Znaniecki (1934), and was originally identified with the search for 'universals' in social life.[4] Today, it is often used to refer to the systematic examination of similarities between cases to develop concepts or ideas (Example 9.2). It has been described by, for example, Lindesmith (1968), Cressey (1950, 1971) and Hammersley and Atkinson. This is the description given by Hammersley and Atkinson (1995: 234–5):

1 An initial definition of the phenomenon to be explained is formulated.
2 Some cases of this phenomenon are investigated, documenting potential explanatory features.
3 A hypothetical explanation is framed on the basis of analysis of the data, designed to identify common factors across the cases.
4 Further cases are investigated to test the hypothesis.
5 If the hypothesis does not fit the facts from these new cases, either the hypothesis is reformulated or the phenomenon to be explained is redefined (so that the negative cases are excluded).
6 This procedure of examining cases, reformulating the hypothesis, and/or redefining the phenomenon is continued until new cases continually confirm the validity of the hypothesis, at which point it may be concluded that the hypothesis is correct (though this can never be known with absolute certainty).

EXAMPLE 9.2 ANALYTIC INDUCTION

Bloor (1978) used analytic induction in his study of surgeons; the study is summarized in Silverman (1993).

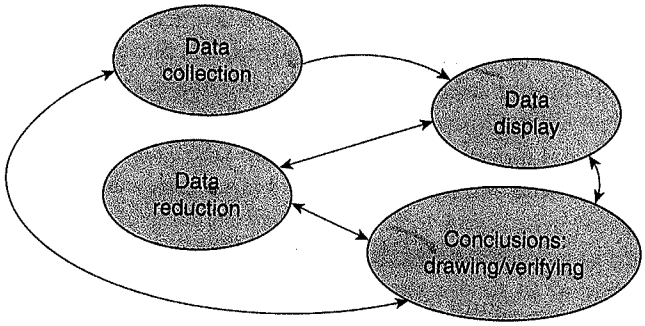
Cressey (1950) used analytic induction to study 'trust violation'.

Lindesmith (1947) used analytic induction of study drug addiction.

9.3 THE MILES AND HUBERMAN FRAMEWORK FOR QUALITATIVE DATA ANALYSIS

Qualitative Data Analysis (1994), by Miles and Huberman, is a comprehensive sourcebook, describing analysis that is directed at tracing out lawful and stable relationships among social phenomena, based on the regularities and sequences

Figure 9.1 Components of data analysis: interactive model



Source: Miles and Huberman, 1994: 12

that link these phenomena (1994: 4). They label their approach 'transcendental realism', and their analysis has three main components:

- data reduction
- data display, and
- drawing and verifying conclusions.

They see these as three concurrent streams or activities, interacting throughout the analysis, as shown in Figure 9.1.

(i) *Data reduction:* Data reduction occurs continually throughout the analysis. It is not something separate from the analysis, it is part of the analysis. In the early stages, it happens through editing, segmenting and summarizing the data. In the middle stages, it happens through coding and memoing, and associated activities such as finding themes, clusters and patterns. In the later stages, it happens through conceptualizing and explaining, since developing abstract concepts is also a way of reducing the data. In the need to reduce data, qualitative analysis is not different from quantitative analysis, and the parallels in conceptual structure are shown in the levels of abstraction diagram in section 9.4 of this chapter. In both quantitative and qualitative analysis, the objective of data reduction is to reduce the data without significant loss of information. In qualitative analysis, an additional important component of not losing information is not to strip the data from their context.

(ii) *Data display:* Data displays organize, compress and assemble information. Because qualitative data are typically voluminous, bulky and dispersed, displays help at all stages in the analysis. Miles and Huberman regard displays as essential, often using the phrase 'You know what you display'. They have no doubt that better displays are a major avenue to valid qualitative analysis (1994: 11). There are many different ways of displaying data – graphs,

charts, networks, diagrams of different types (Venn diagrams, causal models, etc.) – and any way that moves the analysis forward is appropriate. Displays are used at all stages, since they enable data to be organized and summarized, they show what stage the analysis has reached and they are the basis for further analysis. The message is clear: good qualitative analysis involves repeated and iterative displays of data. The same point is made in the grounded theory literature.

(iii) *Drawing and verifying conclusions:* The reasons for reducing and displaying data are to assist in drawing conclusions. While drawing conclusions logically follows reduction and display of data, in fact it takes place more or less concurrently with them. Thus possible conclusions may be noted early in the analysis, but they may be vague and ill-formed at this stage. They are held tentative pending further work, and sharpened during it. They are not finalized until all the data are in, and have been analysed. Conclusions will be in the form of propositions, and once they have been drawn, they need to be verified.

Conclusion drawing and verifying is the third part of this analysis. It involves developing propositions, and is conceptually distinct from the other stages, but again is likely to happen concurrently with them. Miles and Huberman give a list of 13 tactics for drawing meaning and conclusions from displayed data. Since conclusions need also to be verified, they give a second list of 13 tactics for testing and confirming findings. The two lists are shown here in Appendix 1.

This stage in the analysis is the most difficult to describe, because it typically involves a number of different analytical processes, which may be used simultaneously rather than sequentially, and which cut across and combine with each other. In other words, several things are going on at once. This work starts from the point where ordering and integration of the previous analysis is required. After coding and memoing (see sections 9.3.1, 9.3.2), there are many labels, at different levels of abstraction, and piles of memos of various kinds. The aim of this stage is to integrate what has been done into a meaningful and coherent picture of the data. The two lists of tactics give an overview of the activities involved, and, as noted, are shown in Appendix 1.

These three overall components are interwoven and concurrent throughout the data analysis. The first two, data reduction and display, rest mainly on the operations of coding and memoing. In virtually all methods for the analysis of qualitative data, coding and memoing are the two basic operations that get the analysis going. I discuss them here in general terms, and deal with them separately. In practice, they happen together and are closely related.

9.3.1 CODING

Coding is the starting activity in qualitative analysis, and the foundation for what comes later. For analysis directed at discovering regularities in the data, coding is central.

What is coding? Codes are tags, names or labels, and coding is therefore the process of putting tags, names or labels against pieces of the data. The pieces may be individual words, or small or large chunks of the data. The point of assigning labels is to attach meaning to the pieces of data, and these labels serve a number of functions. They index the data, providing a basis for storage and retrieval. The first labels also permit more advanced coding, which enables the summarizing of data by pulling together themes, and by identifying patterns. In view of the volume and complexity of much qualitative data, these early labels become an essential part of subsequent analysis. So basic coding is both the first part of the analysis and part of getting the data ready for subsequent analysis. Advanced coding is the same activity – labelling and categorizing – applied at higher levels of abstraction with the data. The type of coding done – that is, what sorts of labels are attached to the data – depends upon the method of analysis being used.

In the Miles and Huberman approach, there are two main types of codes – descriptive codes and inferential (or pattern) codes. Early labels may be *descriptive* codes, requiring little or no inference beyond the piece of data itself. These are especially valuable in getting the analysis started, and in enabling the researcher to get a 'feel' for the data. Glaser and Strauss use the term 'in vivo' codes in the same way, in grounded theory coding. In Example 9.3 Richards uses the term 'topic coding' in much the same way. First level coding mainly uses these descriptive, low inference codes, which are very useful in summarizing segments of data, and which provide the basis for later higher order coding. Later codes may be more interpretive, requiring some degree of inference beyond the data. Thus second level coding tends to focus on pattern codes. A pattern code is more inferential, a sort of 'meta-code'. *Pattern* codes pull together material into smaller and more meaningful units. A good way to understand pattern codes is by analogy with factor analysis in quantitative research (section 12.6). A factor is a concept at a higher level of abstraction, which brings together less abstract variables. Similarly, a pattern code is a more abstract concept that brings together less abstract, more descriptive codes.

There is the usual range of possibilities, when it comes to bringing codes to the data or finding them in the data. At one end of the continuum we can have prespecified codes or more general coding frameworks. At the other end, we can start coding with no prespecified codes, and let the data suggest initial codes. This decision is not independent of other such decisions concerning research questions, conceptual framework and the structuring of data generally. Nor, as before, does it need to be an either–or decision. Thus, even when guided by an initial coding scheme, we can be alert to other labels and categories suggested by the data. Similarly, we might start with a 'tabula rasa',

derive a first set of codes from the data and then draw on a coding scheme after the initial analysis.

There is another similarity, in this sort of coding of data, with quantitative research. It concerns operational definitions:

> Whether codes are prespecified or developed along the way, clear operational definitions are indispensable, so they can be applied by a single researcher over time and multiple researchers will be thinking about the same phenomena as they code. (Miles and Huberman, 1994: 63)

Operational definitions, in a quantitative context, mean the definition of a variable in terms of the operations necessary to measure it. This quote makes clear the applicability of the concept in this style of qualitative analysis. There must be clear links between data indicators and the conceptual labels (or codes) given to the data. These links enable check coding, and tests of inter-coder reliability in qualitative analysis. They are important in establishing the audit trail through the analysis.

In Example 9.3 Coffey and Atkinson (1996: 33–44) illustrate coding using an interview with an academic anthropologist. Miles and Huberman (1994: 55–72) give coding examples from several different studies and settings, and show some coding frameworks and lists. In the third item in Example 9.3, Richards (2005: 87–8) illustrates another useful general approach to coding, using the terms descriptive coding, topic coding and analytical coding.

EXAMPLE 9.3 CODING

- Coffey and Atkinson (1996: 33–44) illustrate coding using an interview with an academic anthropologist.
- Miles and Huberman (1994: 55–72) give coding examples from several different studies and settings, and show some coding frameworks and lists.
- Richards (2005: 87–8) gives the following example, using the terms descriptive, topic and analytic coding.

Descriptive coding involves coding and storing information about the cases being studied.

Topic coding, the 'hack work' of qualitative research, labels pieces of text according to its subject.

(Continued)

(Continued)

Analytic coding goes further, and is central to qualitative inquiry. It involves both the interpretation of data, and the conceptualizing and theorizing of data.

A passage of text will normally require all three types of coding.

'A man interviewed is discussing the need for community action in the local council elections, in which a schoolteacher is a candidate. This man says that he never listens to gossip about the schoolteacher, it's women's stuff. But he does worry that she is standing for local council, when she is obviously not a responsible person.'

- Descriptive coding: First, store the information about the speaker, perhaps about three attributes: gender, age and job – *male, 45 years and tradesman*
- Topic coding: Now, what topics are being discussed in this passage? *The need for community action and the schoolteacher; perhaps we need to code for her multiple roles.*

In two ways the coding has described the passage: what sort of person offered these ideas, and what they were about.

- Analytic coding: Now, what's going on in the statement about the schoolteacher? *Several themes there that need noting, about patriarchal assumptions, the credibility of 'gossip', the informal networks of women, the authority of the schoolteachers and the interplay of interpersonal and political relations.*

Richards notes that such coding leads the researcher to ask further useful questions: *Did men always deny they gossiped? Are the negative attitudes to the schoolteacher coming mainly from the over-forties? And how do they relate to attitudes to community action?*

Different writers use different terminology to describe the levels and types of coding, and this can produce confusion when reading the literature. Yet both the main ideas involved in coding and the main types of coding have similarities, despite the different terms used. Box 9.1 illustrates this, using the Miles and Huberman, Richards and grounded theory (section 9.5) approaches to coding.

BOX 9.1
Terminology in coding

There are many descriptions of coding in the literature, and considerable variation in terminology. Even the three types of coding described in sections 9.3–9.5 of this chapter use different terms, but the similarities between them are important. Comparing Miles and Huberman (section 9.3.1) and Richards (Example 9.3):

Miles and Huberman	*Richards*
Descriptive codes	Topic codes
Pattern codes	Analytic codes

Descriptive and topic codes focus on identifying and labelling what is in the data. *Pattern and analytic codes* go further, interpreting and/or interconnecting and/or conceptualizing data.

Grounded theory coding (section 9.5) is similar, but more detailed. Thus:

- *In vivo codes* focus on what is in the data
- *Open codes* raise the conceptual level of the data
- *Axial codes* focus on interconnections between open codes
- *Selective codes* raise the conceptual level of the data again

The important point in all of these is that a first level of coding is descriptive, whereas second and higher levels are analytic. These higher levels take the analysis of the data from a descriptive level to a conceptual or theoretical level, and there are several ways of doing this (finding patterns, abstracting-conceptualizing, interpreting, etc.).

In summary, coding is the concrete activity of labelling data, which gets the data analysis under way, and which continues throughout the analysis. Initial coding will typically be descriptive and low-inference, whereas later coding will integrate data by using higher-order concepts. Thus there are two main types of codes – low inference descriptive codes and higher inference pattern or conceptual codes. While coding is central, basic to all analysis, and goes on throughout the analysis, analysis is not only coding. It also involves memoing.

9.3.2 MEMOING

Memoing is the second basic operation of qualitative data analysis, but this does not imply that it is the second stage. The operations are not sequential – memoing begins at the start of the analysis, along with coding.

While the researcher is coding, at whatever level, all sorts of ideas will occur. These become the stuff of memos, which record the ideas. Glaser's definition of a memo is widely used:

> A memo is the theorising write-up of ideas about codes and their relationships as they strike the analyst while coding ... it can be a sentence, a paragraph or a few pages ... it exhausts the analyst's momentary ideation based on data with perhaps a little conceptual elaboration. (Miles and Huberman, 1994: 72; Glaser, 1978: 83–4)

These memos can cover many things. They may be substantive, theoretical, methodological or even personal. When they are substantive and theoretical, these memos may suggest still deeper-level concepts than the coding has so far produced. Thus they may point towards new patterns, and a higher level of pattern coding. They may also elaborate a concept or suggest ways of doing this, or they may relate different concepts to each other. This last type of memo produces propositions.

As with the higher levels of coding, the important thing about substantive and theoretical memos is that *they have conceptual content and are not simply describing the data.* They help the analyst move from the descriptive and empirical to the conceptual level. They are therefore especially important in induction, since they move the analysis towards developing propositions. Memoing links coding with the developing of propositions. It is important in qualitative analysis to balance discipline with creativity, and it is in memoing where creativity comes in. We can think of coding as the systematic and disciplined part of the analysis (though creativity and insight also are needed to see patterns and connections), whereas memoing is the more creative-speculative part of the developing analysis. This speculative part of course needs verification.

Together, coding and memoing provide the building blocks for this style of qualitative analysis. While the initial analysis might be mainly concerned with coding, it is not long before memoing is involved. We have said earlier that the analysis of qualitative data cannot be reduced to rules. But there is one exception to this one rule (Glaser, 1978: 83): *Record all ideas, as they happen, as memos. When an idea occurs during coding, stop the coding and record the idea.* Later, the memos can be indexed for storage and subsequent use. Miles and Huberman (1994: 72–5) show several memos taken from different projects, and Charmaz (2006: 72–95) describes and discusses memoing, and gives several examples of memo-writing.

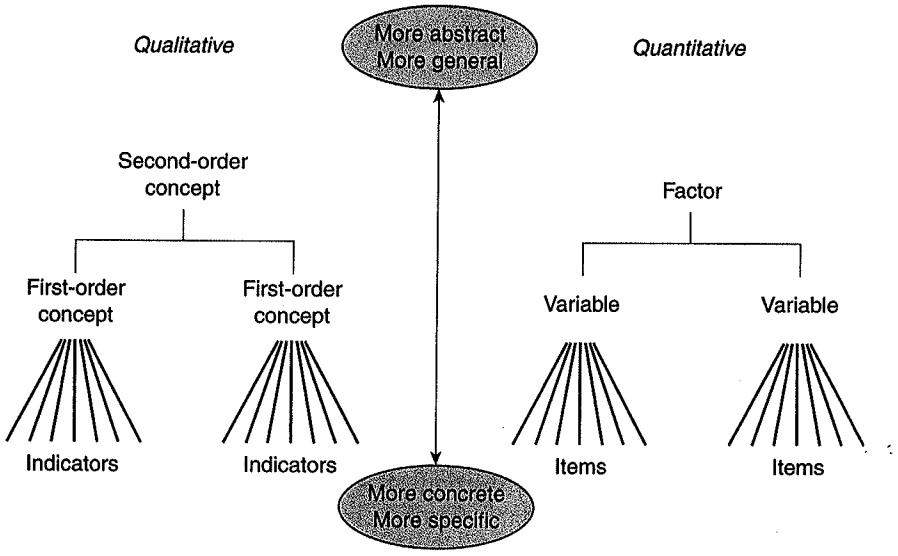
9.4 ABSTRACTING AND COMPARING

The sort of qualitative analysis so far described requires many different intellectual tools and activities, but two stand out as fundamental – abstracting and comparing.

Abstracting

The essential point here is that some concepts are at a higher level of abstraction than others. The terms concrete to abstract describe this continuum of abstraction,

Figure 9.2 Levels of abstraction in data analysis



as do the terms specific to general. The diagram in Figure 9.2 shows this idea, as does the original diagram in Chapter 2. The similarities between those two diagrams have been noted already.

This diagram shows the levels of abstraction in both the qualitative and quantitative contexts, and the close similarities between the two. At the lowest level of abstraction, the most concrete or specific or descriptive level, we have indicators (qualitative) and items (quantitative). At the next level, the first level of abstraction, we have first-order concepts (qualitative) and variables (quantitative). As will be shown in Chapter 11, latent trait theory in measurement formalizes this idea. At the next level, the second order of abstraction, we have second-order concepts (qualitative) and factors (quantitative). Again, factor analysis and cluster analysis in quantitative work formalize this idea. The process of abstraction does not need to stop there. Still more abstract and general concepts are possible, in both approaches, but two levels of abstraction show the idea, and encompass most of what we do.

Two things stand out from this diagram. First, the conceptual structure, in terms of the continuum from concrete-to-abstract and specific-to-general, is remarkably similar in both approaches. Therefore, the general nature of this sort of analysis, developing higher-order concepts to summarize and integrate more concrete levels of data, is also similar in both approaches. Second, quantitative analysis has formalized the way it moves from one level to the next to a much greater extent than qualitative analysis. Thus quantitative analysis aggregates items into variables, to move to the first level of abstraction, and derives factors from variables, to move to the second level of abstraction. By

the nature of its data, qualitative analysis cannot be formalized to the same extent, but the role of abstraction explains the central importance of tree diagrams in qualitative data analysis (O'Leary, 2004: 258; Richards, 2005: 104–21).

Comparing

Comparison is fundamental to all systematic inquiry, whether the data are qualitative or quantitative. In quantitative research we don't often think explicitly about comparison, since comparison is built into all stages of quantitative inquiry. Thus measurement encapsulates the concept of comparison, quantitative design is developed to permit comparison, and the various data analysis techniques are based on comparison. So we are comparing automatically when we use the techniques of quantitative research.

Comparison is not so automatically integrated into qualitative analysis, and it therefore needs stressing. Comparing is essential in identifying abstract concepts, and to coding. At the first level of coding it is by comparing different indicators in the data that we arrive at the more abstract concepts behind the empirical data. Thus it is comparison which leads to raising the level of abstraction, to the 'one-upping' (Glaser, 1978) so essential to conceptual development. The same is true for coding at a higher level. Comparing concepts and their properties at a first level of abstraction enables us to identify more abstract concepts. The systematic and constant making of comparisons is therefore essential to conceptual development at all levels in the analysis of qualitative data.

Tesch (1990), in her comprehensive survey of methods used in qualitative data analysis, sees comparison as the central intellectual activity in analysis. Glaser and Strauss (1967), co-founders of grounded theory, saw comparison as so important that they described grounded theory analysis as the 'constant comparative method'. Thus comparing is at the heart of grounded theory analysis.

9.5 GROUNDED THEORY ANALYSIS

Grounded theory is both an overall approach to research and a set of procedures for developing theory through the analysis of data. The approach was described in Chapter 7. This section deals with the basic ideas of grounded theory analysis. This analysis aims directly at generating abstract theory to explain what is central in the data. All of its procedures are oriented to this aim, and from the start of its coding it recognizes both the central role of conceptual abstraction and the hierarchical structure of theoretical knowledge.

How does the analyst go about generating theory from data? What follows now is an overview of grounded theory analysis, then a description of open, axial and selective coding.

9.5.1 OVERVIEW

The ultimate idea in discovering a grounded theory is to find a core category, at a high level of abstraction but grounded in the data, which accounts for what is central in the data. Grounded theory analysis does this in three steps, which are conceptually distinct but not necessarily sequential. The first is to find conceptual categories in the data, at a first level of abstraction. The second is to find relationships between these categories. The third is to conceptualize and account for these relationships at a higher level of abstraction. This means there are three general types of codes – substantive codes (produced by open coding), which are the initial conceptual categories in the data; theoretical codes (produced by axial coding), which connect these categories; and the core code (produced by selective coding), which is the higher-order conceptualization of the theoretical coding, around which the theory is built.

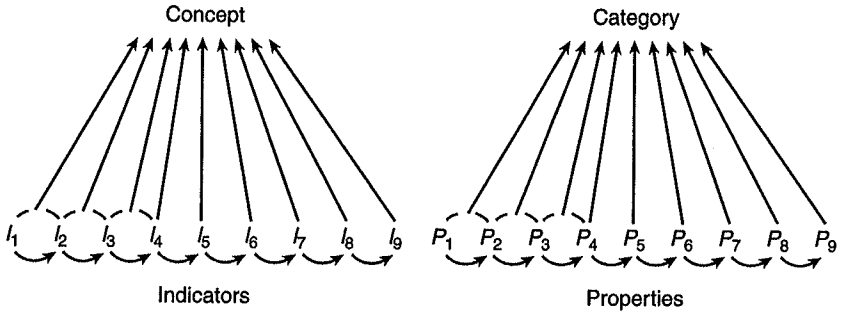
Thus the first objective is to find the substantive codes in the data. These are categories generated from the empirical data, but at a more abstract level than the data themselves. In this first level of analysis, some of these substantive codes will appear as more central in the data than others. The second objective is to bring the main substantive codes together, to interconnect them using theoretical codes. These statements of interconnection are propositions or hypotheses about the data, to be integrated into the grounded theory. The third objective is thus to find a higher order, more abstract construct – the core category – which integrates these hypotheses into a theory, and which describes and explains them.

At the heart of grounded theory analysis is coding – open coding, axial coding and selective coding. These are not necessarily done sequentially – rather, they are likely to be overlapping and done concurrently. But they are conceptually distinct operations. Open coding finds the substantive codes. Axial coding uses theoretical codes to interconnect the main substantive codes. Selective coding isolates and elaborates the higher order core category.

9.5.2 OPEN CODING

Open coding constitutes a first level of conceptual analysis with the data. The analyst begins by 'fracturing' or 'breaking open' the data. This is why the term 'open' in open coding is used. The idea is to open up the theoretical possibilities in the data. The purpose is to use the data to generate abstract conceptual categories – more abstract than the data they describe – for later use in theory building. These are the substantive codes. Open coding necessarily involves a close examination of (some of) the data, identifying conceptual categories implicit or explicit in the data, and the theoretical possibilities the data carry. What makes grounded theory analysis different from other forms of qualitative analysis is its insistence, from the start, on generating abstract conceptual categories to account for the data being studied. Therefore its coding is *not* centrally concerned with simple description, thematic analysis or interpretation of the data, though these may assist the analyst in the

Figure 9.3 Concept–indicator diagram



Source: Glaser, 1978: 62

open coding. It *is* centrally concerned with 'rendering the data theoretically' (Glaser) or 'converting the data analytically' (Strauss). These phrases mean using the data to generate more abstract categories. The focus is on generating grounded abstract concepts, which can become the building blocks for the theory.

The key to understanding both open coding and grounded theory analysis in general is the concept–indicator model. It is shown again in Figure 9.3. It is the same model we saw earlier, both in latent trait theory in measurement, and in the discussion of levels of abstraction in section 9.4. As Glaser (1978: 62) points out, this model directs the coding of empirical indicators in the data.

A concept can have many different possible empirical indicators. When we infer a concept from an indicator in the data, we are abstracting – going upwards from a piece of empirical data to a more abstract concept. Because a concept has many indicators, the indicators are interchangeable with each other for the purposes of inferring the concept. This means that I_1 (indicator 1) is an indicator of the concept, and so is I_2 (and I_3, I_4, etc.). We compare indicator with indicator, assessing their similarities and differences, in order to infer the concept. We also ask, constantly, what more abstract concept this piece of empirical data indicates.

Thus the process of labelling in open coding is guided by two main activities – making comparisons and asking questions. The first means that different pieces of data, as indicators, are constantly compared with each other to help generate abstract categories. For the second, one type of question is constantly asked which is distinctive of grounded theory analysis. It has three forms:

- What is this piece of data an example of? Or,
- What does this piece of data stand for, or represent? Or,
- What category or property of a category does this piece of data indicate?

Open coding, like all coding, is labelling, putting labels on pieces of data. Sometimes these labels will be descriptive, low-inference labels, and sometimes

they will be '*in vivo*' labels, but mostly they will be labels that involve a first level of inference. In grounded theory open coding, the labelling is guided by the model and questions shown above. Codes (labels) at this early stage of the analysis are provisional, and a piece of data may have several labels. Closure on final codes is delayed until substantial coding has been done, and until the analyst has a stable view of what is central in the data. Potentially central categories are also being noted as the open coding proceeds, but closure is delayed here also.

Open coding is about using the data to generate conceptual labels and categories for use in theory building. Its function is to expose theoretical possibilities in the data. It is not about bringing concepts to the data, and no a priori coding scheme is used. Using only concepts and categories generated by the data ensures that they are grounded in the data, and that any concepts to be used in the theory have earned their conceptual status. Thus the analyst starts with no preconceived conceptual categories, but uses only those generated by the data. Open coding is not centrally concerned with summarizing data, or with describing data, or with finding synonyms for words in the data or with interpreting data. It may do these things indirectly, or as part of generating abstract categories, but these things are not the objective – the objective is to conceptualize the data.

Successful open coding generates many provisional labels quickly from even a small amount of data, but this sort of coding does not go on indefinitely. The objective of open coding is not the endless generation of conceptual labels throughout the data. This process of labelling therefore needs to be balanced by two other processes. One is to keep an overview of the data in mind, and to keep looking broadly across the data, rather than only to do the intensive coding. This is what Glaser (1992) calls 'dip and skip', where you intensively code some parts (dip), while at the same time skimming the data using comparisons to look for possible conceptual patterns, and for concepts that tie together different pieces of the data and different incidents in the data (skip). The other is deliberately to step back from the data and to make judgements about what seems to be central and basic in the data, over and above all the coding labels generated. This judgement is made by focusing on such questions as:

- What centrally seems to be going on here?
- What are these data mainly about?
- What is the basic social problem people here are facing, and what is the basic social process they use for dealing with it?

In grounded theory analysis, it is important to find and focus on what is central in the data. The whole process is about successively integrating the data into a smaller set of more abstract concepts and categories. Therefore the focus

is on possible integrating concepts right from the start. Ultimately, a grounded theory will be built around a core category, which accounts for most of the variation in the data, and integrates other parts of the data around it. Therefore coding procedures at all stages are aligned with this purpose of reducing the data through abstracting, of seeking to discover the conceptually central aspects of the data. The outcome of open coding is a set of conceptual categories generated from the data. There will also be some ordering and classification of these categories, and some sense of what is central in the data. There may be some initial views of possible core categories, but whether this has happened or not at this stage, a small number of important categories will have emerged.

EXAMPLE 9.4 OPEN CODING

Strauss and Corbin (1990: 61–74) illustrate open coding using observational data from a restaurant setting.

Strauss (1987: 40–64 and 82–108) demonstrates open coding in a seminar with students, with himself as leader, using observational data from a study of pain management in a cardiac recovery unit.

Corbin (1986: 102–20) describes open coding on observational data from a nurse in a paediatric unit studying children's responses to hospitalization.

Alder (2002) uses open coding in a study examining how caring relationships develop between middle school students and their teachers.

9.5.3 AXIAL (OR THEORETICAL) CODING

Axial coding is the name given to the second operation in grounded theory analysis, where the main categories that have emerged from open coding of the data are interconnected with each other. The word 'axial' is used by Strauss and Corbin, and is intended to denote the idea of putting an axis through the data, where an axis connects the categories identified in open coding. Glaser (1978) uses the more general term 'theoretical coding' to describe this stage. Its meaning is made clear below.

If open coding breaks the data apart, or 'runs the data open' (Glaser, 1978), in order to expose their theoretical possibilities and categories, axial coding puts categories back together again, but in conceptually different ways. Thus axial coding is about interrelating the substantive categories that open coding has developed.

How is this done? To do the interrelating, we will need some concepts that connect things to each other. These connecting concepts are called theoretical codes, which is why Glaser uses the term theoretical coding rather than axial coding. Strauss and Corbin also use the term 'coding paradigm' to describe the set of concepts used for making the connections between things. All of these terms mean the same thing.

We know from quantitative analysis that there are many different ways in which connections between things can occur. For example, causes and consequences is one way, seeing things as different aspects (or dimensions or properties) of a common category is another, seeing things as parts or stages of a process is a third, a stimulus–response association is a fourth, and so on. Some of the ways things can be connected are covered by Miles and Huberman, in their list of tactics noted earlier, and two comprehensive treatments of this topic are by Glaser and Rosenberg. Glaser (1978: 72–82) discusses 18 ways these connections between things can occur. He calls them 'coding families'. Rosenberg (1968: 1–21), writing more from the quantitative perspective, classifies relationships between variables (the quantitative equivalent of connections between things) into three main types (symmetrical, reciprocal and asymmetrical) and then gives several subtypes within each of these classifications.

Strauss and Corbin (1990: 99–107) write exclusively about the interactionist coding paradigm. This identifies causal conditions, phenomenon, context, intervening conditions, action/interaction strategies and consequences as a way of interrelating categories in the data – these are theoretical concepts used to interconnect the data. Thus, if the interactionist paradigm is used, the outcome of axial coding is an understanding of the central phenomenon in the data in terms of the conditions that give rise to it, the context in which it is embedded, the action/interaction strategies by which it is handled, managed or carried out, and the consequences of those strategies.

This idea of theoretical codes is important, but not esoteric – it is about the ways in which things are interconnected with each other. We will see in quantitative analysis (Chapter 12) that there are two conceptually distinct stages in studying relationships between variables. One is finding and describing the relationship. The other is interpreting the relationship, or saying how the relationship has come about, or giving meaning to the relationship. It is the same here, in qualitative analysis. The three sources indicated above, Glaser (1978), Rosenberg (1968) and Miles and Huberman (1994), together give a comprehensive description of possible ways things can be related. These descriptions overlap, and all draw upon ideas on this topic from quantitative research.

EXAMPLE 9.5 AXIAL CODING

Strauss and Corbin (1990: 96–115) illustrate axial coding using pain management data.

Strauss (1987: 64–8) demonstrates axial coding around the category 'monitoring', in a study of medical technology in a cardiac care unit.

Swanson (1986: 121–32) gives several examples of developing categories in axial coding, using as data nurses' accounts of their learning experiences.

Alder (2002) uses axial coding in a study examining how caring relationships develop between middle school students and their teachers.

9.5.4 SELECTIVE CODING

Selective coding is the third operation in grounded theory analysis. The term 'selective' is used because, for this stage, the analyst deliberately selects one central aspect of the data as a core category, and concentrates on this. When this selection is made, it delimits the theoretical analysis and development to those parts of the data that relate to this core category, and open coding ceases. The analysis now proceeds around the core category, and the core category becomes the centrepiece of the grounded theory.

In selective coding, therefore, the objective is to integrate and pull together the developing analysis. The theory to be developed must have a central focus, around which it is integrated. This will be the core category of the theory. It must be a central theme in the data, and in order to integrate the other categories in the data, the core category will have to be at a higher level of abstraction. Potential core categories are noted right from the start of the analysis, though final decisions about the core category should not be made too early in the analysis.

Thus selective coding uses the same techniques as the earlier open and axial coding, but at a higher level of abstraction. The focus now is on finding a higher order concept, a central conceptual category at the second level of abstraction. Selective coding deals with what is central in the data analytically, not simply descriptively. All aspects of grounded theory analysis focus on conceptualizing and explaining the data, not on describing the data. For Glaser (1992), in true grounded theory analysis, the core category will emerge from the constant comparisons that have driven the earlier coding. Once the core category is clear, it is elaborated in terms of its properties, and systematically related to other categories in the data. Relationships are then validated against the data. This stage also shows those categories where further data are required, and thus directs further theoretical sampling. In grounded theory language, this stage is called the systematic densification and saturation of the theory.

EXAMPLE 9.6 SELECTIVE CODING

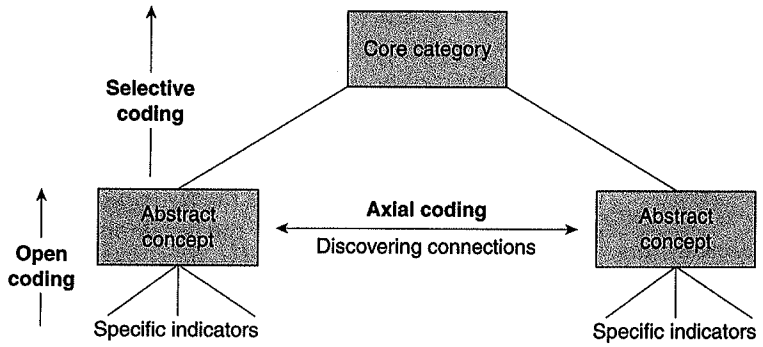
Strauss and Corbin (1990: 116–42) illustrate the steps involved in selective coding using data about how women with chronic illness manage pregnancy.

Strauss (1987: 69–75) illustrates selective coding using nursing work in a cardiac recovery unit.

Corbin (1986: 102–20) illustrates selective coding using memos relating to pregnancy in situations of chronic illness.

The objective throughout is to construct abstract theory about the data, which is grounded in the data. The concepts the theory will use are not brought *to* the data and are not obvious *in* the data. They need to be inferred *from* the

Figure 9.4 Diagrammatic representation of grounded theory analysis



Open coding: discovering abstract concepts in the data; raising the conceptual level of the data.
Axial coding: discovering, in the data, connections between abstract concepts
Selective coding: selecting the core category by concentrating on the basic social process evident in the data; raising the level of abstraction again to the core category; elaborating the core category.

data by induction. This inductive inference is the process of abstraction. By showing a particular piece of data to be an example of a more abstract (first-order) concept, the analyst raises the conceptual level of the data. By showing that first-order concept to be a particular instance, or property, of a more general second-order concept, the conceptual level of the data is raised again. Thus this abstracting is done twice, following the conceptual structure shown in the levels of abstraction diagram (Figure 9.2). By way of summary, a diagrammatic representation of grounded theory analysis is shown in Figure 9.4.

This description has not covered all aspects of grounded theory analysis. Among topics not dealt with here, but covered in the grounded theory literature, are theoretical sensitivity, sampling and saturation, the difference between substantive and formal theories, elaboration and the densification of a grounded theory, basic social problems and processes, and the implications of the grounded theory approach for the study of social processes. As noted in Chapter 7, there has also been considerable recent diversification of grounded theory methods. Further reading on all of these topics can be found in the *Handbook of Grounded Theory* (Bryant and Charmaz, 2007b) and Charmaz (2006).

EXAMPLE 9.7 GROUNDED THEORY

Grounded Theory in Practice (Strauss and Corbin, 1997) is an edited collection of readings of grounded theory studies, from former students of Strauss. The editors provide commentaries for each paper.

Examples of Grounded Theory: A Reader (Glaser, 1993) is an edited collection of 25 grounded theory papers from both quantitative and qualitative projects.

9.6 OTHER ANALYTIC APPROACHES IN QUALITATIVE ANALYSIS

A dilemma for grounded theory, according to Denzin, is how to be subjective, interpretive and scientific at the same time (Lonkila, 1995: 46). This is one reason for recent diversification in grounded theory, and especially for the development of constructivist grounded theory (Charmaz, 2006). The difficulty of doing this may also be one reason for the development of other approaches in qualitative analysis. We will now look briefly at five of these, the first more interpretive, and the other four more focused on language.

9.6.1 NARRATIVES AND MEANING

Data analysis approaches based on segmenting, coding and categorization are valuable in attempts to find and conceptualize regularities in the data. But they by no means exhaust the data, or possibilities for their exploration. Also, they break the data into small pieces, fostering a 'culture of fragmentation' (Atkinson, 1992). In doing this, they can also decontextualize the data. Coffey and Atkinson (1996: 52) write:

> 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.

Miles and Huberman and the grounded theory writers are aware of the problem of fragmentation and decontextualization, and they suggest ways of recombining and of recontextualizing the data. But other approaches (such as the analysis of narratives and stories) deal more holistically with qualitative data right from the start. Much education research data occurs 'naturally' in story form (for example, in participant observation research), and qualitative data can also be solicited and collected in story form, as in oral and life histories, and biographical interviewing. Even where data are not explicitly solicited in story form, they will often come with storied characteristics, as in unstructured interviews, where people may give narrative responses to an interviewer's questions. Thus there is a storied character to much qualitative data, and thinking about stories in the data can enable us to think creatively about collecting and interpreting data (Coffey and Atkinson, 1996).

Narratives and stories are also valuable in studying lives and lived experience, as is often demonstrated in studies concerned with empowerment. Contemporary anthropology and feminism often emphasize the study of lives from the narrator's viewpoint, with data seen as a shared production with the researcher (Manning and Cullum-Swan, 1994). Using stories as a way to capture the lived experience has occurred in many research settings – in medical and illness studies (Brody, 1987; Coles, 1989), in studies of major life events and trauma (Riessman, 1993), in studies in education from both students' point of view (Delamont, 1989, 1990; Measor and Woods, 1984) and teachers' points of view (Goodson, 1992), and in studies of life in organizations (Martin, 1990). Narratives of this sort can give a uniquely rich and subtle understanding of life situations, and the story is often a feasible way of collecting data just because it is such a common device in everyday interaction.[5]

How can qualitative data in narrative and story form be explored and analysed? Elliott (2005) points out that there is no single approach, and that researchers borrow ideas from literary studies and socio-linguistics to assist their analysis. She notes the three-part analytic framework used by Mishler (1995 – meaning, structure and interactional context) and the two-part framework used by Lieblich et al. (1998 – content and form).[6]

The following brief description draws mainly on the writing of Coffey and Atkinson (1996), who use Denzin's framework from interpretive biography for thinking about narratives. They describe formal approaches to narrative analysis, where the focus is on identifying the structural features of narratives, and their arrangement – here, narrative analysis tends towards semiotics (Manning and Cullum-Swan, 1994).[7] They show also how narratives can be studied from the point of view of their function, using function as the unit of analysis. To illustrate the functional properties of narratives, they take the examples of success stories and moral tales, and of narratives as chronicles. The latter lead naturally to oral and life histories, and to biographies, autobiographies and personal experience methods generally (Clandinin and Connelly, 1994). Further avenues for the analysis of narratives are opened up by thinking about whose voices are telling the stories – in any storytelling context, the voices are differentiated and stratified – and by the social and cultural context in which the stories are told. In a general sense, stories are part of the representation of social reality as text, and narratives are therefore social constructions located within power structures and social milieux. In this respect, narrative analysis overlaps with discourse analysis.

In narrative analysis, form and content can be studied together, and a concern with narrative can illuminate how informants use language to convey particular meanings and experiences. How people convey their meanings through language can be looked at from a variety of complementary perspectives. We can

examine, for example, how language is used figuratively. Coffey and Atkinson show how analysis can explore participants' use of imagery, and how such devices as metaphors reveal shared meanings and understandings. The more general exploration of the use of linguistic symbols to convey shared cultural meanings is referred to as 'domain analysis' (Coffey and Atkinson, 1996; Spradley, 1980).

People use metaphors constantly as a way of making sense of experience, and of expressing and conveying its meaning. Qualitative analysts will often do the same thing in making sense of data. Miles and Huberman (1994: 250–52) indicate some of the useful properties of metaphors in qualitative analysis – for example, they are data-reducing devices, pattern-making devices, decentring devices and ways of connecting findings to theory. Metaphors are one important way that people use language figuratively. They are a major type of trope (or literary device), comparing two things using their similarities but ignoring their differences. Other tropes often used are irony (the view from the opposite, sometimes incongruous or paradoxical side), synecdoche (linking instances to a larger concept) and metonymy (representing a whole in terms of one of its parts – Miles and Huberman, 1994: 287). Focusing on these concepts in analysing data for meaning links this sort of qualitative analysis with semiotics, as Coffey and Atkinson point out, and as is shown in section 9.6.4. Example 9.8 shows two books that use narratives in education research.

EXAMPLE 9.8 NARRATIVE ANALYSIS

Connelly and Clandinin (1999) present stories from both teachers and administrators. The authors analyse and reflect upon these stories to make links between knowledge, context and identity, for both teachers and administrators.

The book by Cortazzi (1991) examines important aspects of primary teachers' experience by drawing on the author's analysis of nearly a thousand accounts of classroom events told by 123 teachers. Through their stories, a clear picture is built up of how teachers see teaching.

Gerstl-Pepin (2006) used narrative policy analysis to examine social justice narratives embedded within 'No Child Left Behind' with respect to economic inequities.

Bohanek (2008) examined the ways in which mothers and fathers scaffold conversations about past emotional events with their preadolescent children, using narratives of positive and negative shared family events.

Chapter 8 discussed the analytic status of interview data and the central role of language in qualitative research. This focus, together with the view of language as a form of social action rather than as a neutral medium for 'making pictures of the world out there', provides a convenient way to approach the following types of qualitative analysis.[8]

9.6.2 ETHNOMETHODOLOGY AND CONVERSATION ANALYSIS

Sociological interest in the study of language was stimulated by ethnomethodology, pioneered by Garfinkel (1967). Ethnomethodology sets out to understand 'folk' (ethno) methods (methodology) for organizing the world (Silverman, 1993). The fundamental assumption of ethnomethodology is that people within a culture have procedures for making sense of their daily life. For ethnomethodologists, culture thus consists not of a stable set of things that members are supposed to know, but of processes for figuring out or giving meaning to the actions of members. The primary focus is on how central features of a culture, its shared meanings and social norms, are developed, maintained and changed, rather than on the content of those meanings and norms (Feldman, 1995: 8).

This focus on how the shared common world is created leads the ethnomethodologist to study activities that ordinary people engage in, often without thinking.[9] Most of the time, especially when joint action and interaction is involved, language is central to these everyday activities. With so much of social life mediated through written and especially spoken communication, the study of language is at the heart of ethnomethodology. Thus conversation analysis becomes a central concern, as ethnomethodologists seek to understand people's methods for producing orderly social interaction.

As an indication of its importance, Heath and Luff (1996) refer to a 1990 bibliography of ethnomethodological–conversation analytic studies which contains more than 1400 citations to articles in five different languages. The general purpose of such studies is to understand the social organization of ordinary, naturally occurring human conduct, in which talk is a primary vehicle for the production and intelligibility of human actions. Where talk only is analysed, verbatim transcripts of actual conversations are used. If the data include all interaction including conversation, a video recording is more likely to be used, as in interaction analysis (Heath and Luff, 1996).

Silverman (1993: 125) summarizes Heritage's account of three fundamental assumptions of conversation analysis. They concern the structural organization of talk, the sequential organization of talk and the need for the empirical grounding of the analysis. Following these assumptions, and using specialized transcription conventions, conversation analysis studies the situated production and organization of talk (or action), developing a 'bottom-up' understanding of

how context influences participants' production of the social reality. For example, with respect to sequential organization, a part of that context is the immediately preceding utterance (or action). The next utterance is produced with respect to the immediately preceding utterance, and itself forms part of the context for the subsequent utterance(s). In this turn-taking, the conduct of participants in interaction is doubly contextual, both context-shaped and context-renewing (Heritage, 1984). Some of the tools and techniques used in conversation analysis are described by Coulthard (1985) and McCarthy (1991).

In this way, conversation analysis, like ethnomethodology generally, aims systematically to uncover and analyse the foundations of social life. Silverman (1993: 127–33) lists some of the features discovered so far in this microscopic analysis of ordinary talk. He concludes that conversation analysis, as an empirically based activity grounded in a basic theory of social action, generates significant implications from the analysis of previously unnoticed interactional forms, and goes on to show how conversation analysis can help in analysing and understanding the talk that occurs in organizations and institutions. Similarly, Heath and Luff (1996: 324) conclude that the naturalistic analysis of conversation and interaction has developed a substantial body of findings which delineate the interlocking social organization of a wide range of ordinary social actions and activities.

EXAMPLE 9.9 ETHNOMETHODOLOGY AND CONVERSATION ANALYSIS

Silverman (1993: 125–43) discusses conversational openings, obligations to answer, the structure of turn-taking and institutional talk.

Wooffitt (1996: 287–305) refers to data from various sources in discussing linguistic repertoires, the organization of descriptive sequences and assembling descriptions.

Lynch (2006) argues that ethnomethodology and conversation analysis offer a path not taken in cognitive science – a viable research programme for investigating nominally 'cognitive' themes (memory, learning, perception, etc.) without trading in mentalistic notions of cognition.

Burns and Radford (2008) use conversation analysis to explore parent–child interaction within Nigerian families.

9.6.3 DISCOURSE ANALYSIS

Another view of language looks above its words, sentences and linguistic features and focuses attention on the way language is used, what it is used for and

the social context in which it is used. The term 'discourse' captures this broader focus, and refers to the general framework or perspective within which ideas are formulated (Sapsford and Abbott, 1996). Discourse inextricably permeates social life, since everything people do is framed within some sort of discourse – thus an ideology is framed within a discourse, so are accounts and descriptions (Wooffitt, 1996), and so is science itself (Gilbert and Mulkay, 1984).

Jupp (1996: 300) cites Worrall's use of the term:

> **Discourse embraces all aspects of a communication not only its content, but its author (who says it?), its authority (on what grounds?), its audience (to whom?), its objective (in order to achieve what?). (Worrall, 1990: 8)**

> **Discourse encompasses ideas, statements or knowledge that are dominant at a particular time among particular sets of people ... and which are held in relation to other sets of individuals ... Implicit in the use of such knowledge is the application of power ... discourse involves all forms of communication, including talk and conversation ... In the latter, however, it is not restricted exclusively to verbalized propositions, but can include ways of seeing, categorizing and reacting to the social world in everyday practices. (1996: 300)**

Discourse analysis is not a unified body of theory, method and practice. Rather, it is conducted within various disciplines, with different research traditions, and with no overarching unifying theory common to all types – being heterogeneous, it is difficult to define (Gee et al., 1992). Edley (2001: 189) notes that discourse analysis has become an umbrella term for a wide variety of different analytical principles and practices. In Taylor's view (2001: 5) it is best understood as a field of research rather than a single practice.

Coulthard (1985) gives an overview of its historical development, and shows the various disciplines that have contributed to it, while Potter and Wetherell (1994: 47) list at least four types of work that use the label discourse analysis. The first is influenced by speech act theory, and is directed at accounts of the organization of conversational exchanges.[10] The second, more psychological, focuses on discourse processes, such as the effect of discourse structure on recall and understanding. The third was developed from a sociology of knowledge perspective, studying specifically how scientists construct their talk and texts to present and validate their work and their actions. The fourth derives from European social philosophy and cultural analysis, and attempts to show how institutions, practices and even the individual person can be understood as produced through the workings of a set of discourses. A similar classification is given by Gee et al. (1992)[11] and McCarthy (1991) identifies some differences between British and American discourse analysis. Our interest here is in the third and fourth types described by Potter and Wetherell (1994) – in discourse analysis for qualitative social research.

Despite the diversity and the many disciplinary perspectives, writers point to some fundamental principles and common features of discourse analysis. At the most general level, three principles inform all studies of discourse (Gee et al., 1992: 228): (a) human discourse is rule-governed and internally structured; (b) it is produced by speakers who are ineluctably situated in a socio-historical matrix, whose cultural, political, economic, social and personal realities shape the discourse; (c) discourse itself constitutes or embodies important aspects of that socio-historical matrix. In other words, discourse reflects human experience and, at the same time, constitutes important parts of that experience. Thus, discourse analysis may be concerned with any part of human experience touched on or constituted by discourse.

At a similar general level, Jupp (1996: 305) identifies three features of discourse analysis as used by Foucault: (a) discourse is social, which indicates that words and their meanings depend on where they are used, by whom and to whom; consequently, their meaning can vary according to social and institutional settings and there is, therefore, no such thing as a universal discourse; (b) there can be different discourses, which may be in conflict with one another; (c) as well as being in conflict, discourses may be viewed as being arranged in a hierarchy: the notions of conflict and of hierarchy link closely with the exercise of power. The concept of power is vital to discourse analysis by way of the theoretical connection between the production of discourses and the exercise of power. The two are very closely interwoven and, in some theoretical formulations, are viewed as one and the same.

More specifically, Potter and Wetherell (1994: 48) point to three features that make the sort of discourse analysis they describe especially pertinent for qualitative education research.

- First, it is concerned with talk and texts as social practices; and as such it pays close attention to features that would traditionally be classed as linguistic *content* – meanings and topics – as well as attending to features of linguistic *form* such as grammar and cohesion. Indeed, once we adopt a discourse analytic approach, the distinction between content and form becomes problematic; content is seen to develop out of formal features of discourse and vice versa. Put more generally, the discourse analyst is after the answers to social or sociological questions rather than to linguistic ones.

- Second, discourse analysis has a triple concern with action, construction and variability (Potter and Wetherell, 1987). People perform actions of different kinds through their talk and their writing, and they accomplish the nature of these actions partly through constructing their discourse out of a range of styles, linguistic resources and rhetorical devices.

- A third feature of discourse analysis is its concern with the rhetorical or argumentative organization of talk and texts. Rhetorical analysis has been

particularly helpful in highlighting the way discursive versions are designed to counter real or potential alternatives (Billig, 1991). Put another way, it takes the focus of analysis away from questions of how a version relates to some putative reality and asks instead how this version is designed successfully to compete with an alternative.

Gee et al. (1992) discuss two main stances within discourse analysis research in education – one emphasizes the study of discourse structure for its own sake, using analytic tools from linguistics (discourse as structure); the other studies discourse as it relates to other social, cognitive, political or cultural processes and outcomes (discourse as evidence). Potter and Wetherell (1994) distinguish two different complementary emphases in their style of discourse analysis. One studies the resources used to construct discourse and enable the performance of particular actions, and maps out the broad systems or 'interpretive repertoires' which sustain different social practices. The other studies the detailed procedures through which versions are constructed and made to appear factual. These different stances towards discourse analysis are often combined in research, but they produce different types of research questions, as shown in Example 9.10.

EXAMPLE 9.10 RESEARCH QUESTION IN DISCOURSE ANALYSIS

Discourse as structure: Gee et al. (1992: 229a) list eight types of question that involve the study of discourse structure for its own sake.

Discourse as evidence: the same authors list seven types of questions researchers have used in the study of discourse in relation to social and cognitive processes (1992: 230).

A discourse-analytic research agenda from a critical perspective: Jupp (1996: 306) lists 12 questions that might guide a critical analysis of documents using discourse analysis.

Silverman (1993) shows how a discourse-analytic perspective can change research questions dramatically.

Potter and Wetherell (1994: 55–63) point out that it is difficult to describe and codify explicit procedures that are used in discourse analysis, but they list five considerations that do recur, and they illustrate how each can operate in analysis. They are: using variation as a lever, reading the detail, looking for rhetorical organization, looking for accountability and cross-referencing

discourse studies. Gee et al. (1992) indicate some of the ways the analysis might proceed in the discourse-as-structure and discourse-as-evidence stances noted above, listing some of the tools linguists use when analysing discourse structure, and showing the categories they find useful in studying the social location of texts.[12] Tonkiss (1998: 250–60) discusses 'Doing Discourse Analysis' under the three broad headings of: selecting and approaching data; sorting, coding and analysing data; and presenting the analysis. Under the second heading – sorting, coding and analysing data – she adds two considerations to the list of five given by Potter and Wetherell above: using key words and themes, and attending to silences.

Discourse analysis is an important development in qualitative research, starting as it does from the assumption that discourse at all levels, including people's accounts, is an important resource:

> In our view, people's texts are not trivial outcomes of communicative needs. Rather, they function at many levels and are the product of a person's entire set of political and psychological conditions and entities. Humans are constant creators of complex and multifaceted meanings. (Gee et al., 1992: 233)

Discourse analysis is sensitive to how spoken and written language are used, and how accounts and descriptions are constructed, and to the complex processes for producing social meanings (Tonkiss, 1998). At the microscopic level, it shares much in common with conversation analysis, and some writers (Coulthard, 1985; McCarthy, 1991) see conversation analysis as a particular type of discourse analysis. In a more macroscopic perspective, discourse analysis emphasizes the interrelationships between accounts and hierarchies, power and ideology. Two important directions for this latter type of discourse analysis are critical discourse analysis (Blommaert and Bulcaen, 2000) and Foucauldian discourse analysis (Gubrium and Holstein, 2000). Critical discourse analysis aims to show 'non-obvious ways in which language is involved in social relations of power and domination, and in ideology' (Fairclough, 2001: 229). Foucault examines how historically and culturally located systems of power/knowledge construct subjects and their world. For Foucault, power operates in and through discourse as the other face of knowledge – thus the term power/knowledge. Discourse not only puts words to work, it gives them their meanings, constructs and perceptions and formulates understanding and ongoing courses of interaction (Gubrium and Holstein, 2000: 493–5). At this level, discourse analysis is similar to deconstruction, in dismantling constructed accounts to show connections with power and ideology. It has grown into a wide-ranging and heterogeneous discipline, which finds its unity in the description of language above the level of the sentence, and an interest in systems of meaning and in the context and cultural influences that affect language in use.

EXAMPLE 9.11 DISCOURSE ANALYSIS

Gee et al. (1992: 253–81) describe three examples of discourse analysis: sharing time in a first-grade classroom, reading with story books at home, and verbal analogy items in standardized test-taking.

Potter and Wetherell (1994) use five extracts from their case study of the construction of a TV current affairs programme about cancer charities to illustrate discourse analysis.

Jupp (1996) gives four case studies of discourse analysis, using different kinds of documents.

Coulthard (1985) has many examples of micro-level discourse analysis in the language teaching context.

9.6.4 SEMIOTICS

Language can be seen as a symbolic sign system, where a sign is something that stands for something else. In language, obviously, the signs are words. Semiotics,[13] or the science of signs, lays out assumptions, concepts and methods for the analysis of sign systems. There are many sign systems (for example, mathematics, music, etiquette, symbolic rites, street signs) to which semiotics can be applied, and Eco (1976) points out that semiotics is concerned with everything that can be taken as a sign. At the same time, semiotics is based squarely on language, in line with the view that human linguistic communication can be seen as a display of signs, or a 'text to be read' (Manning and Cullum-Swan, 1994).

The Swiss linguist Saussure and the American philosopher Pierce were the founders of semiotics. Pierce's basic point is that anything can be a sign. For Saussure, being a sign entails being part of a code, and he generated a method which showed that structures and words are inseparable (Silverman, 1993). Semiotics has thus been associated with the structural tradition in literary criticism, but the apparatus of semiotics provides also a way of thinking about any sign-mediated social activity.

An essential idea in semiotics is that surface manifestations derive their meanings from underlying structures (Feldman, 1995). This makes semiotics especially useful in the analysis of language and of texts. Semioticians identify mechanisms by which meaning is produced (the most common ones are metaphor, metonymy and opposition), and have devised techniques using these mechanisms for interpreting qualitative data. Feldman (1995: 22–39) illustrates three of those techniques (semiotic clustering, semiotic chains and semiotic squares) in her analysis of data from her university housing office study.

She provides an example of how the use of these techniques helped her to see relationships in the data of which she was not otherwise aware, thereby illuminating her data in a powerful way. As a rather different example, Manning and Cullum-Swan (1994) present a semiotic reading of the menus at McDonald's.

Semiotics can also be used for the analysis of texts, and we have noted already Silverman's (1993) use of semiotics to analyse narrative structures. With its focus on linguistic structures and categories, it can be used to develop a theory of texts and their constituent elements. This takes text analysis well past the earlier quantitative content analysis (Berelson, 1952) in an effort to get to the deeper meaning. Such meaning is to be found not only in words and phrases, but in the system of rules that structures the text as a whole. It is therefore this underlying structure and the rules it embodies that can tell the researcher what its cultural and social message is. While this semiotic emphasis is valuable, MacDonald and Tipton (1996) remind us that there are limits to the understanding we can develop using only the texts. A text also needs to be studied in its social context.

EXAMPLE 9.12 SEMIOTIC ANALYSIS

McRobbie (1978, 1991) demonstrates semiotic analysis using a magazine aimed at teenage girls.

Feldman (1995: 21–41) discusses semiotic cluster analysis using the example of 'buildings'.

Manning and Cullum-Swan (1994) present a semiotic analysis of a McDonald's menu.

Mavers (2007) conceptualizes writing as a process of design to study how meaning is made by a 6-year-old child who uses semiotic resourcefulness in email exchanges with her uncle.

9.6.5 DOCUMENTARY AND TEXTUAL ANALYSIS

We noted in Chapter 8 the availability and richness of documentary data for education and social research. The analysis of such data shares characteristics with the approaches just described, but it has also some distinctive themes.

One theme focuses on the social production of the document, starting with how the document came into being. All documentary sources are the result of human activity, produced on the basis of certain ideas, theories or commonly accepted, taken-for-granted principles, and these are always located within the

constraints of particular social, historical or administrative conditions and structures (Finnegan, 1996; MacDonald and Tipton, 1996). Words and their meanings depend on where they are used, by whom and to whom. Thus, as discourse analysts point out (for example, Jupp, 1996: 305), meaning varies according to social and institutional setting. Therefore documents and texts studied in isolation from their social context are deprived of their real meaning. Thus an understanding of the social production and context of the document affects its interpretation. Similar considerations apply also to the social production of an archive – what is kept, where and for how long, and what is thrown away (MacDonald and Tipton, 1996: 189).

A second, related, theme is the social organization of the document. We saw these questions from Hammersley and Atkinson (1995: 173) in Chapter 8: How are documents written? How are they read? Who writes them? Who reads them? For what purposes? On what occasions? With what outcomes? What is recorded? What is omitted? What does the writer seem to take for granted about the reader(s)? What do readers need to know in order to make sense of them? Silverman (1993: 63) uses these questions to study the social organization of documents, irrespective of their truth or error. Thus he shows how even such apparently 'objective' documents as organizational files are 'artfully constructed with a view to how they may be read'. He cites the work of Cicourel and Kitsuse in education, Garfinkel with coroners and Sudnow in hospital deaths and criminal statistics to show how the sociological analysis of statistics and files raises fundamental questions about the processes that produce them, quite apart from questions of the truth or error of the statistics themselves. In the same light, he also considers public records and visual images.

EXAMPLE 9.13 ANALYSIS OF THE SOCIAL ORGANIZATION OF DOCUMENTS

Silverman (1993: 61–71) applies textual analysis to files, statistical records, records of official proceedings and images, and includes illustrations from the work of others.

Woods (1979) analyses school reports, and shows the concepts and categories teachers use to make normative judgements about pupils.

A third theme concerns the more 'direct' analysis of text for meaning, this time including questions of truth and error. This analysis can focus on the surface or literal meaning, or on the deeper meaning, and the multi-layered nature

of meaning is now much more widely understood and accepted (Finnegan, 1996: 149). The surface meaning has often concerned historians, whereas sociologists have been more interested in ways of uncovering deeper meaning. Methods used range from interpretive understanding following the ideas of Dilthey (MacDonald and Tipton, 1996: 197) to more structural approaches, especially semiotics, as described above.

A fourth theme is the application of different theoretical perspectives to the analysis of texts and documents. As an example, Jupp (1996) describes the critical analysis of documents, seeing documents as media for discourses, and thus drawing on discourse analysis. Deconstruction is an approach that also has applicability in such a context. Thus, as Silverman points out, there are many ways of thinking about textual analysis, and many different theoretical perspectives that can be applied. Silverman is convinced also (1993: 89) that sociologists make too little use of the great potential of texts as rich data, especially in light of their (often) relatively easy accessibility. The relevance of this point for educational research has been raised in earlier chapters.

9.7 COMPUTERS IN THE ANALYSIS OF QUALITATIVE DATA

While the use of a computer is not appropriate for all of the approaches to analysis described, there are a number of programs to assist the qualitative researcher today. Computer Assisted Qualitative Data Analysis is now known among researchers as CAQDAS.

In choosing among the numerous packages available, several websites provide useful information. Thus, for example, Sage, through its Scolari imprint, publishes a number of commercial CAQDAS packages (www.scolari.com). The website Text Analysis Info Page (www.textanalysis.info) has a comprehensive list of different sorts of packages available. The Computer Assistance of Qualitative Data Analysis Scheme (www.soc.surrey.ac.uk/caqdas) maintains a most useful site to investigate.

There are several factors to consider and questions to ask when a researcher is choosing among the packages. For example:

- Compatibility with my analytic approach. Does this package enable me to do the sort of analysis I want to do?
- Ease of use. Do I have a sense that I can master this software and work with it in ways that will promote my creativity?

- Product support and upgrade path. Is this product well supported by a strong company and is it likely to be further developed and enhanced? This means that the product will continue to grow as my understanding and practice of qualitative research develops and as I grow as a researcher.

  Does the product have previous versions?

  Can I download and try out a trial copy? How does this trial cope and feel with some of my data?

  Is the company active in its engagement with research and researchers?

- Is there a supportive learning community?

  Does the product have good quality tutorials?

  Are there opportunities for training, workshops?

  Is the product supported by a website and is there a discussion forum available?

  Do people actively use this product around me in my context?

  Does the research community in my area use this product? Are there frequent mentions of the product in the recent literature?

- Costs of the software – not all the costs are in the purchase price.

  Does it require specialized or higher-grade hardware?

  Does the ongoing licence require further costs?

  Are training and support expensive?

  Who buys this software or does my institution provide a copy?

Whatever the choice of package, it needs to be remembered that CAQDAS is a tool that can help in qualitative data analysis, but it cannot do the analysis. The researcher's input and creativity will always be required. It is also important to consider the analysis tool when planning the research, rather than as an afterthought once data are ready for analysis. The choice of analysis package will be likely to influence the way data are transcribed and input for analysis (and perhaps also collected). Appendix 2 contains a brief description of QSR N6, one of the most prominent packages in use today to assist in the analysis of qualitative data.

CHAPTER SUMMARY

This chapter has covered a lot of material. The first part of this summary lists the main points of the chapter, in the usual way. In the second part, in Box 9.2, it describes a general approach to the preparation and planning of the data analysis section of a qualitative research proposal.

Diversity in qualitative analysis: multiple approaches and methods and no one right way; but also some features common to all methods; importance of the audit trail (how does researcher get from data to conclusions?)

Analytic induction: uses induction to raise the level of abstraction and to trace out relationships between concepts

Miles and Huberman: data reduction, data display, drawing and verifying conclusions

data reduction – reduce data without significant loss of information

data display – use any type of diagram to display data as analysis proceeds

drawing conclusions – 13 tactics suggested (Appendix 1)

verifying conclusions – 13 tactics suggested (Appendix 1)

Coding: assigning labels to pieces of data; different types and levels of coding (first level descriptive, low inference; higher levels analytic, finding patterns and/or conceptualizing and/or interpreting)

Memoing: recording all ideas (substantive, theoretical methodological, etc.) that occur during coding

Abstracting and comparing: two fundamental activities in qualitative data analysis

abstracting – conceptualizing data at higher levels

comparing – similarities and difference between pieces of data or concepts

Grounded theory analysis: open, axial, selective coding; the concept–indicator model

open coding – discovering abstract concepts in the data

axial coding – discovering connections between abstract concepts

selective coding – raising the level of abstraction again to the core category

Narratives and meaning: preserving storied character of data; multiple methods of analysis; form and content

Ethnomethodology: how shared meanings and social norms are developed, maintained; focus on everyday behaviour; central role of language; conversation analysis

Discourse analysis: field of research with different approaches; the nature of discourse; its structure and relationship to hierarchies, power, ideology

Semiotics: the science of signs; language as a sign system; how language produces meaning

Documentary analysis: social production of document; social organization of document; analysis of meaning; applying different theoretical perspectives

Computers: CAQDAS; choosing software for the analysis of qualitative data

BOX 9.2
Writing the data analysis section of a qualitative research proposal

Students writing qualitative dissertation proposals often have difficulty with the section on the analysis of qualitative data. Faced with the many methods available, an effective way to proceed is:

1. Decide whether your project requires a specialized approach to data analysis. This should follow from the way your research and research questions have been framed and developed. For example, a grounded theory study will require grounded theory analysis, a discourse analysis will require some type of discourse analysis, and so on. If it is specialized, the proposal can then go on to describe the type of specialized analysis to be used, with appropriate support from the literature.
2. If a specialized approach is not involved, one of the more general approaches will be useful. (The Miles and Huberman approach is particularly good in this situation.) When identifying and describing the general approach selected, points to include are the basic operations of coding and memoing, and stressing that the data will be analysed, not just summarized and described. There are different directions the analysis itself might take – for example, it might be inductive, concerned with conceptualizing the data, or interpretive, concerned with analysing meaning, or thematic and concerned with identifying patterns in the data. In all cases, ensure there is support from the reference literature.
3. In doing either (1) or (2), show how the proposed analysis fits with the overall logic of the research. This helps to make your proposal convincing, by strengthening its internal consistency and validity. (A common problem is the lack of fit between the data analysis section and other sections of the proposal.)
4. Show also how the analysis will be systematic, well-organized and thorough. This gives the proposal discipline, suggesting an audit trail through the analysis. In this way, you make your proposal more scholarly.