REAL WORLD RESEARCH

A Resource for Users of Social Research Methods in Applied Settings

Third Edition

Colin Robson

WILEY
A John Wiley and Sons, Ltd, Publication

CHAPTER 17

The analysis and interpretation of qualitative data

This chapter:

- stresses the need for a systematic analysis of qualitative data;
- emphasizes the central role of the person doing the analysis, and warns about some deficiencies of the human as analyst;
- discusses the advantages and disadvantages of using specialist computer software;
- explains the Miles and Huberman approach to analysis which concentrates on reducing the bulk of qualitative data to manageable amounts and on displaying them to help draw conclusions;
- suggests thematic coding analysis as a generally useful technique when dealing with qualitative data;
- reviews the widely used grounded theory approach;
- summarizes a range of alternative approaches; and
- finally considers issues involved in integrating qualitative and quantitative data in multi-strategy designs.

Introduction

Qualitative data have been described as an 'attractive nuisance' (Miles, 1979). Their attractiveness is undeniable. Words, which are by far the most common form of qualitative data, are a speciality of humans and their organizations. Narratives, accounts and other collections of words are variously described as 'rich', 'full' and 'real', and

contrasted with the thin abstractions of number. Their collection is often straightforward. They lend verisimilitude to reports.

The 'nuisance' refers to the legal doctrine that if you leave an attractive object, such as an unlocked car, where children can play with it, you may be liable for any injuries they sustain. Naive researchers may be injured by unforeseen problems with qualitative data. This can occur at the collection stage, where overload is a constant danger. But the main difficulty is in their analysis. There is no clear and universally accepted set of conventions for analysis corresponding to those observed with quantitative data. Indeed, many 'qualitative' workers would resist their development, viewing this enterprise as more of an art than a science. But for those who do wish to work within the kind of scientific framework advocated in this book, and who wish to persuade scientific or policy-making audiences, there are ways in which qualitative data can be dealt with systematically. This chapter seeks to provide an introduction to that task.

In the typology of research strategies that has been adopted in this text, the various types of flexible and multiple-strategy designs are the prime generators of large amounts of complex qualitative data.

Qualitative data are often useful in supplementing and illustrating the quantitative data obtained from an experiment or survey. Small amounts of qualitative data used as an adjunct within a largely quantitative fixed design study will not justify detailed and complex analysis. Often the need is simply to help the account 'live' and communicate to the reader through the telling quotation or apt example. However, when methods generating qualitative data form the only, or a substantial, aspect of the study, then serious and detailed attention needs to be given to the principles of their analysis.

Two assumptions

1. *If you have a substantial amount of qualitative data you will use some kind of software package to deal with it.* Standard software, even a simple word-processing package, can do much to reduce the sheer tedium of qualitative data analysis (see Hahn, 2008, on the use of standard Microsoft Office software for a small qualitative project). For anything other than a small amount of data, the amount of drudgery you can avoid, and the ease with which you can relate to the data, make the use of a computer near to essential. There are also specialist qualitative data analysis packages which aid the process even more. See Appendix 8 for further details.
2. *Unless you already have experience yourself, you will be helped or advised by someone who does have experience in this type of analysis.* The dominant model for carrying out qualitative analysis has in the past been that of apprenticeship. Without accepting all the implications of such a model (which tends, for example, to include a principled inarticulacy about process), there is undoubted value in expert advice. The help provided by software is very different from that in quantitative analysis. There the 'expert's' role is largely to point you towards an appropriate test and to ensure that you understand the outcome. In qualitative data analysis, both the experienced person and the computer help you through a not very well-specified process.

Types of qualitative analysis

Box 17.1 provides a typology of possible approaches. Quasi-statistical approaches rely largely on the conversion of qualitative data into a quantitative format and have been covered under the heading of content analysis in Chapter 14 (p. 348). See also Abeyasekera (2005) who provides a range of suggestions.

BOX 17.1

Different approaches to qualitative analysis

1. *Quasi-statistical approaches*
   - Uses word or phrase frequencies and inter-correlations as key methods of determining the relative importance of terms and concepts.
   - Typified by *content analysis*.
2. *Thematic coding approach*
   - A generic approach not necessarily linked to a particular (or any) theoretical perspective.
   - All or parts of the data are *coded* (i.e. identified as representing something of potential interest) and labelled.
   - Codes with the same label are grouped together as a *theme*.
   - Codes and themes occurring in the data can be determined inductively from reviewing the data and/or from relevance to your research questions, previous research or theoretical considerations.
   - The themes then serve as a basis for further data analysis and interpretation.
   - Makes substantial use of summaries of the themes, supplemented by matrices, network maps, flow charts and diagrams.
   - Can be used on a purely descriptive or exploratory basis, or within a variety of theoretical frameworks.
3. *Grounded theory approach*
   - A version of thematic coding where, as a matter of principle, the codes arise from interaction with the data.
   - Codes are based on the researcher's interpretation of the meanings or patterns in the texts.
   - Used to develop a theory 'grounded' in the data.
   - Can be used very prescriptively following rules laid down by founders of the approach, or as a general style of analysis using a specialized terminology for different types of coding.

Note: There are other specialized approaches, including discourse and conversation analysis, and the analysis of narratives (i.e. stories in written, spoken or other forms). See Chapter 14, p. 374.

Thematic coding analysis is discussed in this chapter as a straightforward general approach which can be used in a wide variety of settings. The widespread popularity of grounded theory as a basis for qualitative data analysis demands its coverage in any treatment of the topic. There is a large number of other approaches, many of which call for an extensive understanding of their theoretical foundations. A brief introduction is provided in a later section of the chapter.

Whatever approach is taken, the researcher has the responsibility of describing it in detail. You have to be able to demonstrate the quality of your analysis, including how you got from the data to your interpretation.

The importance of the quality of the analyst

The central requirement in qualitative analysis is clear thinking on the part of the analyst. Fetterman (1998) considers that the analysis is as much a test of the researcher as it is a test of the data: 'First and foremost, analysis is a test of the ability to think – to process information in a meaningful and useful manner' (p. 93). As emphasized at the beginning of Part V, qualitative analysis remains much closer to codified common sense than the complexities of statistical analysis of quantitative data. However, humans as 'natural analysts' have deficiencies and biases corresponding to the problems that they have as observers (see Chapter 13, p. 328). Some of these are listed in Box 17.2. Systematic

BOX 17.2

Deficiencies of the human as analyst

1. *Data overload*. Limitations on the amount of data that can be dealt with (too much to receive, process and remember).
2. *First impressions*. Early input makes a large impression so that subsequent revision is resisted.
3. *Information availability*. Information which is difficult to get hold of gets less attention than that which is easier to obtain.
4. *Positive instances*. There is a tendency to ignore information conflicting with hypotheses already held and to emphasize information that confirms them.
5. *Internal consistency*. There is a tendency to discount the novel and unusual.
6. *Uneven reliability*. The fact that some sources are more reliable than others tends to be ignored.
7. *Missing information*. Something for which information is incomplete tends to be devalued.
8. *Revision of hypotheses*. There is a tendency either to over- or to under-react to new information.
9. *Fictional base*. The tendency to compare with a base or average when no base data is available.
10. *Confidence in judgement*. Excessive confidence is rested in one's judgement once it is made.

11. *Co-occurrence.* Tends to be interpreted as strong evidence for correlation.
12. *Inconsistency.* Repeated evaluations of the same data tend to differ.

(adapted and abridged from Sadler, 1981, pp. 27–30)

documented approaches to analysis help minimize the effects of these human deficiencies. However, there is an inescapable emphasis on *interpretation* in dealing with much qualitative data which precludes reducing the task to a defined formula. Hence, the suggestions made in this chapter are more in the nature of guidelines rather than tight prescriptions.

Common features of qualitative data analysis

While the possible approaches to analysis are very diverse, there are recurring features. Miles and Huberman (1994, p. 9) give a sequential list of what they describe as 'a fairly classic set of analytic moves':

- giving labels ('codes') to chunks (words, phrases, paragraphs, or whatever), labelling them as examples of a particular 'thing' which may be of interest in the initial set of materials obtained from observation, interviews, documentary analysis, etc.;
- adding comments, reflections, etc. (commonly referred to as 'memos');
- going through the materials trying to identify similar phrases, patterns, themes, relationships, sequences, differences between subgroups, etc.;
- using these patterns, themes, etc. to help focus further data collection;
- gradually elaborating a small set of generalizations that cover the consistencies you discern in the data; and
- linking these generalizations to a formalized body of knowledge in the form of constructs or theories.

This general approach forms the basis of *thematic coding analysis* discussed below (p. 474).

Similarity and contiguity relations

Maxwell and Miller (2008) are concerned that an emphasis on coding and categorizing is in danger of losing connections within accounts and other qualitative material. They make a distinction between *similarity relations* and *contiguity relations*. When using coding, similarities and differences are commonly used as the basis for categorization. Relationships based on contiguity involve seeing connections between things, rather than similarities or differences. We look for such relationships within a specific interview transcript or observational field notes, seeking connections between things which are close together in time or space. They can also be sought between categories and codes once they have been established as a next step in analysis.

Similar distinctions have been proposed previously, including Coffey and Atkinson's (1996) 'concepts and coding' as against 'narratives and stories' and Weiss's (1994) 'issue-focused' and 'case-focused' analysis.

Maxwell and Miller review several advantages of combining categorizing and connecting strategies for analysing qualitative data. They suggest that it may be useful to think in terms of

> categorizing and connecting 'moves' in an analysis, rather than in terms of alternative or sequential overall strategies. At each point in the analysis, one can take either a categorizing step looking for similarities and differences, or a connecting step, looking for actual (contiguity based) connections between things. In fact, it is often productive to alternate between categorizing and connecting moves (p. 470).

They provide an exemplar (pp. 471–2) illustrating one way in which the two strategies can be integrated.

In their view, the 'grounded theory' method discussed later in the chapter (p. 489) actually uses this strategy although with a different terminology. In particular, Corbin and Strauss's (2008) 'axial coding' is effectively the same as 'connecting analysis'. The other main approach covered below (p. 474), 'thematic coding' analysis, while essentially based on categorizing, does not preclude following Maxwell and Miller's suggestions.

Using the computer for qualitative data analysis

The single constant factor reported by qualitative researchers is that such studies generate very large amounts of raw data. A small ethnographic style study will generate many pages of field notes including observations, records of informal interviews, conversations and discussions. This is likely to be supplemented by copies of various documents you have had access to, notes on your own thoughts and feelings etc. A multi-method case study will produce a similar range and amount of material. Even a strictly limited grounded theory study relying solely on interviews leaves you with 20 or more tapes to be transcribed and subsequently analysed.

Before getting on with any type of analysis, you need to ensure that you know what data you have available and that they are labelled, stored and, if necessary, edited and generally cleaned up so that they are both retrievable and understandable when you carry out the analysis. A typical first analytic task of labelling or coding the materials (e.g. deciding that a particular part or segment of an interview transcript falls into the category of 'requesting information' or 'expressing doubt' or whatever) involves not only assigning that code but also of having a way of seeing it alongside other data you have coded in the same way.

In the pre-computer era, these tasks were accomplished by means of file folders containing the various sources of data, markers and highlighters, and copious photo-copying. One strategy was to make as many photocopies of a page as there were different

codes on that page, then to file all examples of a code together. It is clear that much of the drudgery of this task can be eliminated by using a word processor. Many data sources will either be directly in the form of computer files or can be converted into them without difficulty. It may be feasible to enter field notes directly into a laptop computer. An interview tape can be entered into the word processor as it is being transcribed. Incidentally, if you have to do this yourself there is much to be said for the use of speech recognition software for this task (listen to each sentence on the tape through headphones, then repeat it out loud to activate speech recognition). It will need to be checked but modern systems can reach high standards of accuracy. Similarly, if you have access to a scanner with optical character recognition software (OCR), it is now straightforward to convert many documents into word processor files. There are some types of data for which this may not be feasible (e.g. handwritten reports).

Word processors are a boon in storing, organizing and keeping track of your data. Obviously you need to observe good housekeeping practices and should take advice on how you can survive possible hard disk crashes, loss, theft, fire, etc. Essentially, this means having multiple copies of everything, regularly kept up to date in more than one location, and in both paper and computer file versions. Word processors can also help with the coding task through 'copy' and 'paste' functions. In this way it is easy to build up files containing all instances of a particular coding whilst retaining the original file with the original data to which codes have been added.

Word processors can also be used to assist in the 'connecting' (as against categorizing) analysis of qualitative data advocated by Maxwell and Miller (2008), discussed earlier in the chapter (p. 469). Marking, extracting and putting together selected data from a longer text can greatly simplify the task of data reduction needed for producing case studies, narratives, etc.

Should you go beyond using standard word processors to one of the many specialist software packages designed to help with qualitative data analysis?

Using specialist qualitative data analysis (QDA) packages

There are many computer packages specifically designed for researchers to use when analysing qualitative data (commonly referred to as CAQDAS – computer-assisted qualitative data analysis). The most widely used has probably been NUD*IST (Non-numerical, Unstructured Data Indexing, Searching and Theorizing), a catchy acronym which encapsulates the central features of many of the packages – indexing, searching and theorizing. NUD*IST has now been superseded by NVivo, developed by the same organization, QSR International (www.qsrinternational.com). It can be used profitably in most situations where you have substantial amounts of qualitative data, and for many different types of study, including grounded theory, conversation and discourse analysis, ethnographic studies, phenomenological studies, action research, case studies and mixed method research. If you have facility in its use it is also a valuable tool when carrying out literature reviews.

While NVivo is the preferred option for qualitative data analysis in many institutions and hence is likely to be readily available and to receive support, there are several other

packages worth considering for particular situations or types of data – see Appendix B for details.

When deciding whether or not to use specialist software, the advantages of time saving and efficiency when analysing large amounts of data (once you have gained familiarity with a package), should be weighed against the time and effort taken to gain that familiarity. Box 17.3 lists some general advantages and disadvantages in their use. García-Horta and Guerra-Ramos (2009) discuss the use of two different packages with interview data, concluding that 'CAQDAS is of great help and can enhance interview data analysis; however, careful and critical assessment of computer packages is encouraged. Their capabilities must not be overestimated, since computers are still unable to perform an independent rational process or substitute the analyst's capacities' (p. 151).

Richards (2002), the prime mover in the development of the NUD*IST and NVivo packages, expresses concerns that the full potential of computer-based analysis is not being realized. More seriously, the packages may actually be having negative effects. Because the coding and sorting tasks can be carried out more effectively and efficiently using a computer package, users tend to focus excessively on this aspect:

BOX 17.3

Advantages and disadvantages of specialist QDA packages

Advantages

- They provide an organized single location storage system for all stored material (also true of word-processing programs).
- They give quick and easy access to coded material (e.g. examples of a particular theme) without using 'cut and paste' techniques.
- They can handle large amounts of data very quickly.
- They force detailed consideration of all text in the database on a line-by-line (or similar) basis.
- They help the development of consistent coding schemes.
- They can analyse differences, similarities and relationships between coded elements.
- Many have a range of ways of displaying results.

Disadvantages

- Proficiency in their use takes time and effort.
- There may be difficulties in changing, or reluctance to change, categories of information once they have been established.
- Particular programs tend to impose specific approaches to data analysis (depends on the program – see Appendix B).
- Tendency to think that simply because you have used specialist software you have carried out a worthwhile analysis. A focus on coding and other technical aspects can give less emphasis to interpretation.

The code-and-retrieve techniques most easily supported by computers and most demanded by users are techniques most researchers had used at some time for sorting out the mess of complex data records. But they were not much discussed in the literature before computing, and not at all clearly associated with the goal of theorizing common to most qualitative methodologies. So computing became associated with techniques that are generic, easily learnt and that emphasize data management and description. Significantly, these are aspects of practical research ignored or even spurned by theoretical writers (p. 266).

While it could be argued that users are simply replicating their previous paper-based, cut and paste, highlighter employing, practices on the computer, it is undoubtedly true that packages are capable of much more than this; for example they can include tools for doing more interpretation once the coding is done. Encouragingly, theory-building software has in recent years been developed to such an extent that it is probably the most widely used type. While, as pointed out by Maxwell and Miller (2008), most of these uses have been based on a prior categorizing analysis (e.g. Richards, 2005), many of the current programs allow the user to create links among and between *any* segments, both within and between contexts, and to display the resulting networks.

Dealing with the quantity of qualitative data

Qualitative data can easily become overwhelming, even in small projects. Hence you need to find ways of keeping it manageable. This process starts before any data are collected when you focus the study and make sampling decisions about people to interview, places to visit, etc. During and after data collection you have to reduce the data mountain through the production of summaries and abstracts, writing memos etc. Miles and Huberman (1994) emphasize that this is a part of analysis and not a separate activity. Decisions about what to select and to summarize, and how this is then to be organized, are analytic choices.

Good housekeeping

Even a small project producing qualitative data can easily leave you overwhelmed with lots of pieces of information of many different types. Possible ways of keeping track include the use of:

- *Session summary sheets.* Shortly after a data collection session (e.g. an interview or observation session) has taken place and the data have been processed, a single sheet should be prepared which summarizes what has been obtained. It is helpful if this sheet is in the form of answers to summarizing and focusing questions. These might include who was involved, what issues were covered, what is the relevance to your research questions (effectively what was the purpose of the session), new questions suggested and implications for subsequent data collection.

- *Document sheets*. A similar sheet prepared for each document collected. This clarifies its context and significance, as well as summarizing the content of lengthy documents. The session summary and document sheets assist in data reduction, an important part of the analysis process.
- *Memoing*. A memo can be anything that occurs to you during the project and its analysis. Memoing is a useful means of capturing ideas, views and intuitions at all stages of the data analysis process.
- *The interim summary*. This is an attempt to summarize what you have found out so far and highlight what still needs to be found out. It is recommended that this is done before you are halfway through the time you have available for data collection. The summary should cover not only what is known but also the confidence you have in that knowledge, so that gaps and deficiencies can be spotted and remedied. *Flexible designs enable you to do this in a way which would not be feasible in a fixed design study but to capitalize on this flexibility you must force yourself to find the time to do this interim summary while you can still take advantage of its findings to direct and focus the later phases of data collection*. The summary can also usefully incorporate a *data accounting sheet* which lists the different research questions and shows, for different informants, materials, settings, etc., whether adequate data concerning each of the questions have been collected.

Thematic coding analysis

Thematic coding analysis is presented here as a generic approach to the analysis of qualitative data. It can be used as a realist method, which reports experiences, meanings and the reality of participants, or as a constructionist method, which examines the ways in which events, realities, meanings and experiences are the effects of a range of discourses operating within society.

Coding has a central role in qualitative analysis. Gibbs (2007), in a very clear and accessible discussion, introduces it as:

> Coding is how you define what the data you are analyzing are about. It involves identifying and recording one or more passages of text or other data items such as the parts of pictures that, in some sense, exemplify the same theoretical or descriptive idea. Usually, several passages are identified and they are then linked with a name for that idea – the code. Thus all the text and so on that is about the same thing or exemplifies the same thing is coded to the same name (p. 38).

Other terms are sometimes used instead of 'code', such as 'incidents', 'segments', 'units', 'data-bits' or 'chunks' (Ryan and Bernard, 2003, p. 87). Because the process involves comparing each new chunk of data with previous codes, so similar chunks will be labelled with the same code, it is sometimes referred to as *constant comparison analysis*.

Coding is followed by grouping the initial codes into a smaller number of *themes*. The term 'theme' is not tightly defined. It captures something of interest or importance in relation to your research question(s). Other terms including 'category', 'label' and 'code'

itself are also used. For example, Miles and Huberman talk about first- and second-level coding. First-level coding is concerned with attaching labels to groups of words. 'Second level coding groups the initial codes into a smaller number of themes.' 'Theme' also tends to be associated with a phenomenological approach to analysis (e.g. Smith, Larkin and Flowers, 2009). However, it is here used more generally, without the implication of necessarily following a particular theoretical approach.

The development of possible themes should be an active concern when you are coding. You need to be continually asking yourself 'what seems to go with what?' and elaborating on and checking these hunches. You will probably start with a very small number of potential themes, modify and add to them during the course of analysis (perhaps organizing them into major themes and sub-themes) and finally be left with a small number once more as various 'runners' are disconfirmed by the data. The work that you do in creating these codes is central to developing an understanding of your data. It lays the foundation for your subsequent analysis and interpretation.

Coding and the development of a thematic framework is central to many qualitative data analyses (although it is not without challenge, e.g. Coffey, Holbrook and Atkinson, 1996). In contrast to many 'named' approaches to qualitative data analysis (such as grounded theory, and discourse or conversation analysis), thematic coding analysis is not necessarily wedded to a particular theoretical framework. It has been used within different theoretical frameworks and can also be used in purely descriptive or exploratory studies. Ritchie, Spencer and O'Connor (2003) and Attride-Stirling (2001) provide introductions.

Guidelines for carrying out a thematic coding analysis

The steps listed below are not unique to thematic coding analysis; many approaches to the analysis of qualitative data have similar steps. Whenever feasible, analysis should be involved at an early stage of carrying out the project. From the start of data collection, you should be looking out for issues of interest in the data including possible patterns or themes. As the data collection proceeds, you move back and forward between the data itself, the extracts from the data which you have coded as possible themes and the analysis you are producing. The various techniques for data reduction, discussed earlier in the chapter (p. 473) will help you to keep on top of this task. Use of memos to jot down ideas including thoughts about likely themes and sub-themes is particularly crucial. Ideas will always come but they can very easily go and be lost, unless you note them down.

Thematic coding analysis can be used inductively where the codes and themes emerge purely from your interaction with the data (as in the grounded theory approach discussed later in the chapter, p. 489). However, there is nothing to stop you starting the analysis with predetermined codes or themes, perhaps arising from your reading of the research literature and/or the research questions you are interested in (as in *template analysis*, King, 2004). At a practical level it can be argued that such preconceptions can bias you toward some aspects of the data, perhaps leading to you ignoring other potentially important themes. An alternative view is that prior engagement with the

BOX 17.4

Phases of thematic coding analysis

1. *Familiarizing yourself with your data.* Transcribing data (if necessary), reading and re-reading the data, noting down initial ideas.
2. *Generating initial codes.* May be done by first devising a framework or template or inductively by interaction with the data. Extracts from the data are given codes in a systematic fashion across the entire data set, with similar extracts being given the same code.
3. *Identifying themes.* Collating codes into potential themes, gathering all data relevant to each potential theme. Checking if the themes work in relation to the coded extracts and the entire data set. Revising the initial codes and/or themes if necessary.
4. *Constructing thematic networks.* Developing a thematic 'map' of the analysis.
5. *Integration and interpretation.* Making comparisons between different aspects of the data using display techniques such as tables and networks. Exploring, describing, summarizing and interpreting the patterns. Demonstrating the quality of the analysis.

literature can enhance your analysis by sensitizing you to features of the data that might otherwise be missed (Tuckett, 2005).

While the phases are, necessarily, presented sequentially, this should not be taken as implying that we are dealing with a linear process where one step is completed before moving to the next one. There is much movement to and fro, where the results of a later phase prompt you to return and rethink what you did at an earlier stage. Remember also that this is a process and it is counter-productive to try and rush it. Just as it is important to start the analysis at an early stage, it is equally important to give yourself time at the end to review the analysis. It is not unknown to get an epiphany at a late stage where you realize that the data are better interpreted in a radically different way!

Box 17.4 lists the steps involved, together with a brief description. In some circumstances, particularly in purely exploratory or descriptive studies, the analysis is terminated after the first three phases have been completed. A substantial proportion of published accounts are limited in this way. Providing that attention is given to demonstrating the quality of your analysis, and hence the trustworthiness of the findings, this may well be appropriate in a real world context. However, it is likely that it provides a very limited understanding of the meaning of your findings and how they might be interpreted, whether in terms of realist mechanisms or whatever. Box 17.5 reviews some of the advantages and disadvantages of thematic coding analysis.

Phase 1: Familiarizing yourself with your data

If you are collecting the data yourself, using a flexible design, this is an ongoing process. After initial data collection, allow time to immerse yourself in the data so you are really

BOX 17.5

Advantages and disadvantages of thematic coding analysis

Advantages

1. Very flexible, can be used with virtually all types of qualitative data.
2. By comparison with other approaches to qualitative data analysis which call for considerable time and effort to understand and require an appreciation of their philosophical and theoretical basis to use legitimately, it is a relatively easy and quick method to learn and use.
3. It is accessible to researchers with little or no experience of qualitative research.
4. The results of the analysis can be communicated without major difficulties to practitioners, policy makers and an educated general public.
5. It is a useful method to employ when working within a participatory research paradigm, where participants are acting as collaborators in the research and in the analysis of findings.
6. It provides a means of summarizing key features of large amounts of qualitative data, using a principled approach acceptable to fellow researchers and journal editors.
7. It is not tied to a particular level of interpretation and can be used in a wide variety of fields and disciplines.

Disadvantages

1. The flexibility of the method means that the potential range of things that can be said about your data is broad, which can be inhibiting to the researcher trying to decide what aspects of their data to focus on.
2. Thematic coding analysis is frequently limited to description or exploration with little attempt made at interpretation.
3. It is not uncommon to find reports where it is claimed that thematic coding analysis has been carried out, and themes are discussed, but there is little or no information about the details of the procedure.
4. Compared to 'branded' forms of analysis such as grounded theory, interpretative phenomenological analysis, discourse analysis or conversational analysis, it is a generic approach which currently has less kudos as an analytic method.

(based, in part, on Braun and Clarke, 2006, pp. 96–7)

familiar with what you have collected. This usually involves repeated reading of the data, doing this in an active way where you are searching for meanings and patterns. If you are presented with data from some other source or for some reason have the full data set before you start the analysis, it is absolutely crucial that you thoroughly immerse yourself in the data as the first step.

When collecting your own data, depending on the time scale and organization of your project, you may have several bouts of data analysis before completing data collection. You then need to immerse yourself in the entire data set to confirm or modify your earlier views about meanings or patterns. While familiarizing yourself with the data you can, and should, be taking notes, writing memos about ideas for formal coding and initial thoughts about themes. In practice, the first three phases shade into each other and are difficult to disentangle.

The familiarization process is time consuming. Don't be tempted to skimp it or try to base it on a selection from the full data set.

Transcription issues

For many projects, much of the data may not be originally in the form of written text (e.g. audio recordings from interviews). It is not essential that you transcribe all (or even any) of such data into text format. It may be feasible to work directly from the original recording. However, having a detailed transcript of some, or all, of the data is often necessary to carry out the analysis. While this is a very time-consuming task (taking several times the time it took to make the recording), it is an excellent way of starting to familiarize yourself with the data. If you are in the fortunate position that someone else will do the transcription for you, it is important that you still have to spend time familiarizing yourself with the data and should also check the transcripts against the original recordings for accuracy. This should always be done with transcripts, of course.

Some forms of qualitative data analysis, such as conversation or discourse analysis call for very detailed transcripts. Typically, thematic analysis does not require the same amount of detail. The necessity for a full 'verbatim' (i.e. word for word) account of all verbal utterances will depend on the nature of your project. It can help if the original recording is to hand when carrying out the analysis as intonation, pauses or other non-verbal features may clarify how a particular utterance should be coded.

Gibbs (2007, pp. 10–21) reviews the issues involved in detail.

Phase 2: Generating initial codes

To do this, you must be thoroughly familiar with the data available (whether this is the full, or an initial, data set), and have a first set of ideas about what is in the data and what you feel is interesting and may be important about them.

Codes refer to 'the most basic segment, or element, of the raw data or information that can be assessed in a meaningful way regarding the phenomenon' (Boyatzis, 1998, p. 63). The process of coding is part of analysis as you are organizing your data into meaningful groups. Once you have coded the data, you can start on deciding what themes you can see in them. By working systematically through the entire data set and giving full attention to each data item, you will then try to identify interesting aspects, which may form the basis of themes within the data set. Include rather than exclude when doing this. Code for as many potential themes as you can come up with. Include some of the context (i.e. surrounding text) if it looks as if it may be relevant when thinking about themes. You can code individual extracts of data in as many different themes as they fit into.

Your themes, which you start to develop in the next phase, are where the interpretative analysis of the data occurs. Coding will in part depend on whether the themes are 'data-driven' or 'theory-driven' (they can be both). In the former, the themes arise from consideration of the data. In the latter, you start by approaching the data with specific questions in mind. Extreme versions such as the 'framework approach' (Ritchie, Spencer and O'Connor, 2003) or 'template analysis' (King, 2004) depend on identifying examples of codes from a pre-existing list. Coding also depends on whether you aim to code the content of the entire data set or are simply coding to identify specific features of the data set. It can be performed manually (lots of printed copies of the data, a big table, highlighters, post-it notes, scissors, etc.) or using a word processor or other standard package, or specialist software (see Appendix B).

What do you code?

This depends on the kind of analysis you are planning to do, which is in turn dependent on your research question(s). As Gibbs (2007, p. 46) points out, fortunately, in much applied real world research, there is a lot of common ground in the kind of phenomena that researchers tend to look for when doing qualitative analysis. Typical things you might consider are listed in Box 17.6. Some codes are essentially descriptive, often in the participant's own words. At some stage in the process, you should be attempting to

BOX 17.6

What can you code?

1. *Specific acts, behaviours* – what people do or say (e.g. getting the opinions of friends).
2. *Events* – these are usually brief, one-off events or things someone has done. It is not uncommon for the respondent to tell them as a story (e.g. moving into a homeless hostel).
3. *Activities* – these are of longer duration than acts and often take place in a particular setting and may have several people involved (e.g. helping partner with dementia get washed and dressed).
4. *Strategies, practices or tactics* – activities aimed towards some goal (e.g. getting divorced for financial reasons).
5. *States* – general conditions experienced by people or found in organizations (e.g. working extra hours to get the job done).
6. *Meanings* – a wide range of phenomena at the core of much qualitative analysis. Meanings and interpretations are important parts of what directs participants' actions.
   a. What concepts do participants use to understand their world? What norms, values, rules and mores guide their actions? (e.g. the idea of 'on-sight climbing' among rock climbers to describe doing a climb without artificial aids, etc. as a superior way of climbing).

b. What meaning or significance does it have for participants, how do they construe events, what are their feelings? (e.g. 'his letter made me feel I was to blame').
c. What symbols do people use to understand their situation? What names do they use for objects, events, persons, roles, settings and equipment? (e.g. teaching referred to as 'work at the chalkface').

7. *Participation* – people's involvement or adaptation to a setting (e.g. 'I find I have to be careful what I say now').
8. *Relationships or interaction* – between people, considered simultaneously (e.g. 'I'm enjoying the family now – the boys like to come home and have friends to stay').
9. *Conditions or constraints* – the precursor to or cause of events or actions, things that restrict behaviour or actions (e.g. firm's loss of markets before lay-offs).
10. *Consequences* – what happens if . . . ('people that haven't got no qualification, but have got a few months' experience are walking into jobs').
11. *Settings* – the entire context of the events under study (e.g. day care centre).
12. *Reflexive* – the researcher's role in the process. How intervention generated the data (e.g. 'It must be hard for you in that situation').

(based on Gibbs, 2007, Table 4.1, pp. 47–8)

move from the descriptive (unless your research questions indicate purely descriptive concerns) to a more theoretically oriented search for categories and themes in the data. This is the central task in the next phase of thematic coding analysis but this should not inhibit you from using codes which themselves move beyond the descriptive.

Marking the coding

Figure 17.1 shows how a short extract might look after coding. In a paper-based approach code names are written in the margin or marked in some way, perhaps using

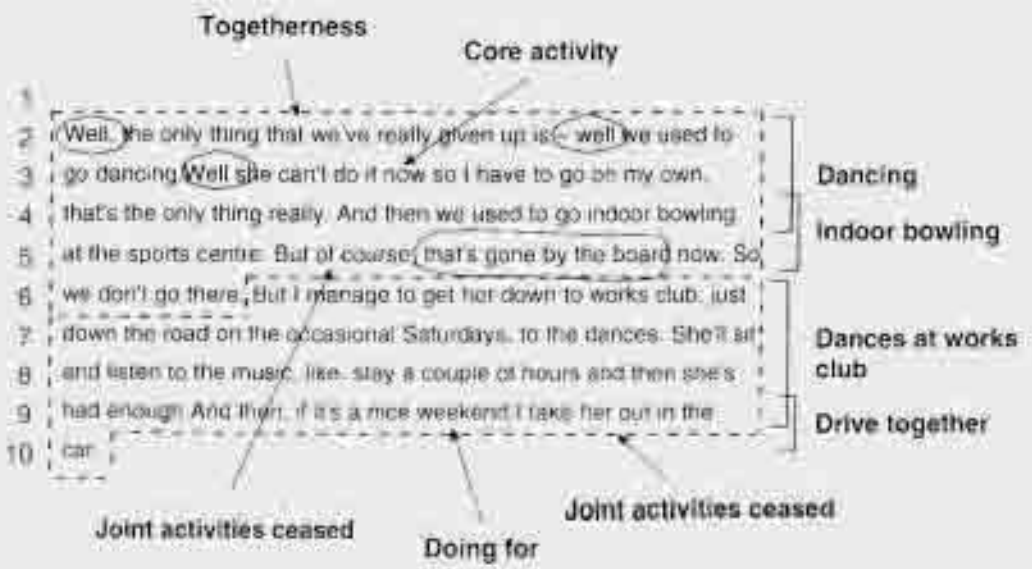

Figure 17.1: Response after coding.

a highlighter pen. Words which stand out in some way (e.g. used for emphasis, unusual terms) could be circled. The figure illustrates several different types of codes, including *descriptive codes* (e.g. dancing, drive together); *categories* (e.g. joint activities ceased); and *analytic codes* (e.g. togetherness, core activity). The extract is unusually heavily coded to illustrate the process. Don't think that you have to code everything. You are just looking for extracts relevant to your analysis, either because they represent codes in the framework you are using or they seem potentially relevant as something likely to be of interest.

While each code is only used once here, in practice, you would be looking for multiple examples from different parts of the text or from other participants. Don't be worried if you have inconsistencies and bits that don't seem to fit in. No data set is without contradictions.

An equivalent process takes place when coding using a software package such as NVivo. Whether paper- or computer-based, each coded extract should have a unique label to assist in retrieving it later. This indicates where it comes from (e.g. which document, person, or whatever), whereabouts in the document (e.g. line number) and other features you might use in the analysis (e.g. gender, age, job status).

Phase 3: Identifying themes

This aspect moves into centre stage when initial coding is completed and you have a long list of the different codes that you have identified. Your task is to sort the different codes into potential themes and to put together all the relevant coded data extracts within the themes you identify. This should have been at the back of your mind when doing the initial coding so it is likely that you will already have some candidates for themes.

Visual representations of various kinds, discussed in connection with the following phase, can also help here. Matrices (tables with rows and columns) and networks (sets of boxes with links between them) may assist in visualizing relationships between codes, between possible themes, and between different levels of themes (e.g. potential main themes and sub-themes within them). Having an initial attempt at the thematic 'map' discussed below may be useful. You may find that some initial codes may end up as themes in their own right, possibly even forming a main overarching theme. Others you will need to combine to form a theme or sub-theme. It is highly likely that several codes do not seem to belong anywhere. Put them, for the time being in a residual category – perhaps a 'miscellaneous' theme. You may find a way of incorporating some of them as the analysis continues.

Box 17.7 suggests some of the things you can look out for when trying to identify themes. The techniques suggested tend to be at a more abstract level relating to the formal linguistic properties of the text than the suggestions for things to code in the previous box. However, there is no hard and fast distinction and it won't matter particularly if you find that you used one of these techniques when coding. Coding is only a tool to get at the themes in the data.

With a collection of possible themes and sub-themes, and the extracts of data coded in relation to them, you can start to refine them. The first step is to read all the collated extracts for each theme and consider whether they appear to form a coherent pattern. If they don't, is it because the theme itself is problematic? Or simply that some data extracts

BOX 17.7

Techniques for identifying themes

1. *Repetitions.* One of the easiest ways to identify themes. Some of the most obvious themes are topics that occur and reoccur. The more the same concept occurs in a text, the more likely it is a theme. How many repetitions are enough to constitute an important theme, however, is an open question which only the investigator can decide.
2. *Indigenous categories.* Terms and concepts that participants themselves use. They may sound unfamiliar to the researcher or be used in unfamiliar ways, providing clues as to their existence. Contrasted with categories or typologies constructed by the analyst.
3. *Metaphors and analogies.* People often represent their thoughts, behaviours and experiences with analogies and metaphors. For example Quinn (1996) found that people talk about their surprise at the breakup of a marriage using extensive metaphors (they thought the couple's marriage was 'like the Rock of Gibraltar' or 'nailed in cement').
4. *Transitions.* Naturally occurring shifts in content may be markers of themes. In written texts, new paragraphs may indicate shifts in topics. In speech, pauses, changes in voice tone or the presence of particular phrases may indicate transitions.
5. *Similarities and differences.* Searching for similarities and differences by making systematic comparisons across units of data. How is a statement, text or whatever similar to or different from those preceding or following? The abstract similarities and differences that this question generates are themes. If a particular theme is present in both expressions, then the next question to ask is 'Is there any difference, in degree or kind, in which the theme is articulated in both of the expressions?' Degrees of strength in themes may lead to the naming of subthemes.
6. *Linguistic connectors.* Look carefully for words and phrases such as 'because', 'since', and 'as a result,' which often indicate causal relations. Words and phrases such as 'if' or 'then', 'rather than', and 'instead of' often signify conditional relations. Time-oriented relationships are expressed with words such as 'before', 'after', 'then' and 'next'. Ryan and Bernard (2003) list several other kinds of relationships that may be useful for identifying themes.
7. *Missing data.* Instead of asking about what is in the data, we can ask about what is missing. For example, in a study of birth planning in China, Greenhalgh (1994) reported that she could not ask direct questions about resistance to government policy but that respondents used silence to protest against aspects of the policy they did not like. Themes that are discovered like this will need to be carefully scrutinized to ensure that you are not just finding what you are looking for. A variant is to scrutinize any expressions that are not already associated with a theme by reading a text over and over.

8. *Theory-related material.* In addition to identifying indigenous themes that characterize the experience of informants, your research questions and issues of theoretical importance can be used to suggest possible themes.

(summarized from Ryan and Bernard, 2003, pp. 89–94)

don't fit in the theme? Sort this out by reworking the theme, finding a new home for the extracts that don't fit or jettisoning them.

Once you are satisfied with this, reread your entire data set. Do the themes seem to capture adequately what you have in the data? You should also look out for possible additional data extracts to code within your themes, which were missed in earlier coding stages. It is also likely that you now think you've got some of the coding wrong, and need to re-code some extracts. Don't worry, this is an iterative process where you go from data to analysis, analysis to data, etc., until you feel reasonably satisfied. Again, don't expect perfection. You could go on forever if you are not careful.

By the end of this phase, you should have at least an intuitive feeling about what the different themes are, how they fit together and the overall story they tell about the data.

Phase 4: Constructing thematic networks and making comparisons

You can now move to formalize the 'fitting together' of the themes into one or more maps or networks. Focus on ways in which themes can be put together, perhaps on the basis of content or on theoretical grounds. It may be that the themes are few enough and about similar enough issues to fit under one network. If they are too numerous, or they seem to be concerned with very different issues or aspects, then put them into two or more groupings. Each will represent a main theme with sub-themes. Attride-Stirling (2001) suggests using three levels of themes: 'global' (super-ordinate themes); 'organizing' (groups of basic themes); and 'basic'. While there are no hard and fast rules about how many themes should make a network, she recommends that from a practical stance, more than 15 may be too many to handle later on in the analysis and less than four may be too few to do justice to the data.

The main or global theme represents your view on what the sub-themes (organizing and basic) are about. It is the core, principal metaphor encapsulating the main point in the data set. If you see more than one such point, then split the themes into two or more networks with different main themes. Figure 17.2 shows an example of a thematic network. When you have produced the thematic network, it is good practice to go through the coded data extracts once more to satisfy yourself that the themes reflect the data and the data support the themes. Modify the themes or network if necessary.

Phase 5: Integration and interpretation

Thematic networks are a tool in analysis, not the analysis itself. Your next task is to explore within and across the themes, to try to understand what the data are telling you.

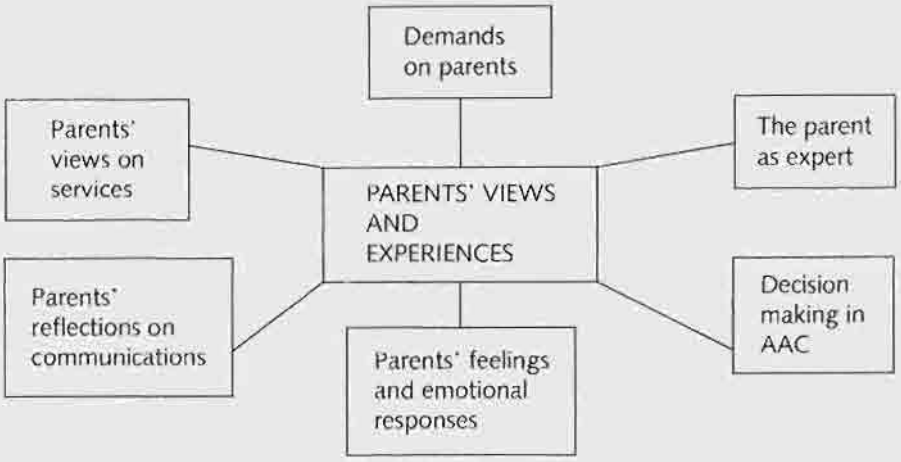

Figure 17.2: Example of a thematic network.
Source: Goldbart, J. and Marshall, J. (2004) Pushes and pulls on the parents of children using AAC. *Augmentative and Alternative Communication*, 20, 204–8. Fig. 3, p. 201.

Miles and Huberman (1994, pp. 245–6) talk about this as 'generating meaning' and list a range of tactics you can use. These include:

1. *Noting patterns, themes and trends.*
2. *Seeing plausibility.* Do the trends, patterns and conclusions make sense?
3. *Clustering.* Grouping events, places, people, processes, etc. together if they appear to have similar patterns or characteristics.
4. *Making metaphors.* Metaphors are rich, data-reducing and pattern-making devices which help to connect data with theory.
5. *Counting.* Helps to see what's there by counting the frequency of occurrence of recurrent events.
6. *Making contrasts and comparisons.* Establishing similarities and differences between and within data sets.
7. *Partitioning variables.* Splitting variables may help in finding more coherent descriptions and explanations.
8. *Subsuming particulars into the general.* Linking specific data to general concepts and categories.
9. *Factoring.* Attempting to discover the factors underlying the process under investigation.
10. *Noting relations between variables.* Using matrix displays and other methods to study interrelationships between different parts of the data.
11. *Finding intervening variables.* Trying to establish the presence and effects of variables intervening between observed variables.
12. *Building a logical chain of evidence.* Trying to understand trends and patterns through developing logical relationships.
13. *Making conceptual/theoretical coherence.* Moving from data to constructs to theories through analysis and categorization.

Some of the tactics have already been used when coding, identifying themes and the relations between them in the thematic network. Many of them are unexceptionable in

that they simply represent a labelling of common practices but several, such as 'factoring' and the use of 'variable-speak', reveal Miles and Huberman's direct translation of concepts from quantitative analysis into qualitative analysis. Researchers used to working with qualitative data may well view this as inappropriate. They may also resist the 'counting' tactic, turning qualitative data into numbers, while happily using pseudo-numbers such as 'few', 'many', 'almost all', etc.

However, there is no requirement to take these tactics en bloc – or to use their terminology.

Using tables for comparative analysis

Several of the tactics call for making comparisons between different aspects of the data. Displaying the data, in the form of tables with rows and columns, provides a simple and useful technique. As discussed in Chapter 16, tables are widely used in quantitative analyses where they are referred to as cross-tabulations (p. 418). Qualitative tables provide a convenient way of displaying text from different parts of the data to make systematic comparisons. There are many different possible types, including:

- Time-ordered tables where the columns are arranged in time sequence. A specific version of this is known as an event listing, where concrete events, sorted into categories, are arranged into time sequence.
- Role-ordered tables where the rows represent data from sets of individuals occupying different roles (e.g. doctor/nurse/administrator/patient).
- Conceptually clustered tables where the columns are arranged to bring together items 'belonging together' (e.g. relating to same theme).
- Effects tables which display data on outcomes.
- Issues tables where the columns concern issues and what happens in connection with them (e.g. who does what).

Using networks

Networks (i.e. a set of boxes or 'nodes' with links between them) are also useful display techniques when you seek to understand patterns and relationships in your data. The thematic network developed in the previous phase is one example. The various types include:

- Context charts showing interrelationships between e.g. roles, groups, organizations which provide the context for the situation studied.
- Event flow networks which show how events (e.g. experiences, incidents) are ordered by time and sequence.
- Activity records displaying a specific recurring activity (e.g. a classroom lesson) as a sequential pattern with specific actions each assigned to a link.
- Flow charts giving sequential decisions made.

- Conceptually ordered tree diagrams[1] showing how phenomena are classified and subcategorized.
- Cognitive maps displaying a person's representation of concepts about a particular domain or area of interest, showing the relationships between them.
- Causal networks consisting of boxes showing the most important variables or factors in a study, with arrows showing the relationships between them.

Your task is to tell the story of your data in a way which convinces the reader of the merit and trustworthiness of your analysis (discussed below in terms of assessing the quality of your analysis). As Braun and Clarke (2006) put it, you have to provide:

> a concise, coherent, logical, non-repetitive and interesting account of the story the data tell – within and across themes. Your write-up must provide sufficient evidence of the themes within the data – i.e. enough data extracts to demonstrate the prevalence of the theme. Choose particularly vivid examples or extracts which capture the essence of the point you are demonstrating, without unnecessary complexity. The extract should be easily identifiable as an example of the issue. However, your write-up needs to do more than just provide data. Extracts need to be embedded within an analytic narrative that compellingly illustrates the story you are telling about your data, and your analytic narrative needs to go *beyond* description of the data, and make an *argument* in relation to your research question (p. 93, emphasis in original).

Demonstrating the quality of the analysis

Box 17.8 lists tactics you might use when assessing the trustworthiness of your analysis.

BOX 17.8

Assessing the quality of qualitative data analysis

Assessing data quality

1. *Checking for representativeness.* There are many pitfalls to the gathering of representative data. The informants, and the events or activities sampled, may be non-representative. Safeguards include the use of random sampling where feasible; triangulation through multiple methods of data collection; constructing data display matrices; and seeking data for empty or weakly

[1] Tree diagrams provide an alternative way of displaying a thematic network where they are sometimes referred to as the 'coding hierarchy'. Note that whereas trees have their roots (i.e. main theme) at the bottom and branches (sub-themes) above, coding hierarchies are, conventionally, the other way round.

sampled cells. Your analysis may be biased, not only because you are drawing inferences from non-representative processes, but also because of your own biases as an information processor (p. 468). Auditing processes by colleagues help guard against this.

2. *Checking for researcher effects.* These come in two versions: the effects you have on the case; and the effects your involvement with the case have on you.
3. *Triangulation.* Not a panacea and it has its own problems (what, for example do you do when two data sources are inconsistent or conflicting – answer, you investigate further, possibly ending up with a more complex set of understandings). However, it is very important: 'triangulation is not so much a tactic as a way of life. If you self-consciously set out to collect and double-check findings, using multiple sources and modes of evidence, the verification process will largely be built into data collection as you go' (Miles and Huberman, 1994, p. 267).
4. *Weighting the evidence.* Some data are stronger than others and you naturally place greater reliance on conclusions based on them. Stronger data are typically those you collect firsthand; which you have observed directly; which come from trusted informants; which are collected when the respondent is alone rather than in a group setting; and which arise from repeated contact.

Testing patterns

5. *Checking the meaning of outliers.* These are the exceptions, the ones that don't fit into the overall pattern of findings or lie at the extremes of a distribution. Outliers can be people, cases, settings, treatments or events. Don't be tempted to hide or forget them. Evaluating an in-service training package we found very high levels of teacher satisfaction with its effectiveness with a very small number of dissidents. Further interviews with the latter established that in all cases an element of the package involving role-playing sessions had been omitted or problematic, which both strengthened the explanation and helped isolate the mechanisms involved.
6. *Using extreme cases.* These are outliers of a particular type, defined in terms of being atypical situations or persons rather than by the data they provide, which may or may not be atypical. An innovation which failed in a school where the circumstances appeared close to ideal appeared linked to the unexpressed resistance of the deputy head teacher responsible for timetabling. Hence suggesting a key factor.
7. *Following up surprises.* Surprises can be salutary. You may well be surprised because something is at variance with your (possibly implicit and not thought through) theory of what is going on. This then provides the opportunity to surface that theory, to possibly revise it and to search for evidence relevant to the revision.
8. *Looking for negative evidence.* This is the tactic of actively seeking disconfirmation of what you think is true. While this is in principle straightforward, you

are likely to have some reluctance to spending a large amount of effort on this activity. Miles and Huberman (1994, p. 271) make the helpful suggestion of giving a colleague your conclusions and free access to your original data with a brief to try to find evidence which would disconfirm your conclusion. If they manage to do this, then your task is to come up with an alternative or elaborated explanation.

Testing explanations

9. *Making if-then tests.* Testing possible relationships, i.e. if one condition obtains or is the case, look to see if a second one is. If it is, we are on the way to understanding what is going on and can make further similar tests. If it isn't true, we have to make other conjectures.
10. *Ruling out spurious relationships.* If you appear to have established a relationship consider whether there may be a third factor or variable which underlies, influences or causes the apparent relationship. In the relationship between guardsmen fainting on parade and the softness of the tar in the asphalt of the parade ground, it appears highly likely that the temperature is providing an intervening causal link rather than noxious fumes from the tar having a direct effect. Note that this is essentially the same tactic discussed above under the heading of 'finding intervening variables' but used for a different purpose. It can also be thought of as finding rival explanations for a relationship.
11. *Replicating a finding.* If a finding can be repeated in a different context or data set, then it is more dependable. Given that once you find a relationship or develop a theory, there is a strong tendency for you to find confirming evidence (and to ignore disconfirming evidence), it is even better if someone else, not privy to your findings, confirms it. Note that this is a particular type of triangulation.
12. *Checking out rival explanations.* It is good practice to try to come up with one or more rival explanations which could account for all or part of the phenomena you are studying. Keeping these 'in play' while you are analysing and gathering further data helps to prevent the premature closure effect discussed above.
13. *Getting feedback from informants.* This process of 'member checking' performs several useful functions. It honours the implicit (or preferably explicit) contract between researcher and informant to provide feedback about the findings. It also provides an invaluable means of corroborating them. While problems of jargon and terminology may need to be attended to, you should be able to present findings in a way that communicates with informants and allows them to evaluate the findings in the light of their superior experience of the setting.

(summarized from Miles and Huberman, 1994, p. 262–77)

The website gives further examples of research using thematic coding analysis.

Data analysis in grounded theory studies

Many analyses of qualitative data are influenced by grounded theory. They typically reference the work of either or both of the originators of the approach, Barney Glaser and Anselm Strauss, which was discussed in Chapter 6, p. 146. However, while some of these analyses follow the detailed prescriptions and terminology discussed in the following sections, others are much more 'in the general style of' and virtually indistinguishable from the thematic coding approach discussed in the previous section.

The aim of grounded theory analysis

The aim is to *generate* a theory to explain what is central in the data. Your task is to find a central core category which is both at a high level of abstraction and grounded in (i.e. derived from) the data you have collected and analysed. This is done in three stages:

1. Find conceptual categories in the data.
2. Find relationships between these categories.
3. Conceptualize and account for these relationships though finding core categories.

It is achieved by carrying out three kinds of coding:

- *Open coding* to find the categories.
- *Axial coding* to interconnect them.
- *Selective coding* to establish the core category or categories (Corbin and Strauss, 2008).

Throughout the analysis theory is built through interaction with the data, making comparisons and asking questions of the data. It is sometimes referred to as the *method of constant comparison* (Pidgeon and Henwood, 1996; pp. 92–4).

Open coding

Here data (e.g. interview transcripts, field notes, documents) are split into discrete parts. The size of the part chosen is whatever seems to be a unit in the data, perhaps a sentence, or an utterance, or a paragraph. The question asked is 'what is this piece of data an example of?' The code applied is a label. It is provisional and may be changed. A piece of data may have several codes (labels), i.e. it may be considered to fall within more than one conceptual category. Labels can be of whatever kind that seems appropriate, including descriptive (e.g. 'accepting advice'), 'in vivo' (i.e. a direct quotation from the data) or more inferential.

These conceptual categories arise from the data. Using pre-determined coding categories and seeking to fit data into such categories is against the spirit of grounded theory. However, this distinction is somewhat metaphysical as the 'conceptual baggage' you bring to your data (whether derived from a pre-existing theory or from analysis of data collected earlier) will inevitably have some influence on what you are likely to 'see' in the data.

Open coding is essentially interpreting rather than summarizing. It is about teasing out the theoretical possibilities in the data. There is much to be said for doing it in a small group. This will enhance the ideas pool about what the data are examples of and it will assist in keeping individuals 'on task'. Is a particular code really grounded in the data? Is the central purpose of open coding being kept in mind?

While carrying out open coding, you should bear in mind any ideas that occur from working with the data about relationships between the categories, and even first thoughts about the core category. This can be encouraged by stepping back from the data from time to time and getting an overall feel for what is going on. As with other approaches to flexible design this initial analysis will be taking place before data collection is complete. There is no requirement to code all the data fully at this stage but you do need to have done a substantial amount of coding and to have a good appreciation of what you have captured overall in the various data sets arising from the interviews, field notes, etc. In an ideal world, you will have got to the stage where the various categories are 'saturated'. That is, you have squeezed as much conceptual juice as you can out of the data so that continuing analysis is giving severely diminished returns in the new categories and insights that it is yielding.

Axial coding

Axial, or *theoretical*, coding is about linking together the categories developed through the process of open coding. Glaser and Strauss, the begetters of grounded theory, now have diverging views about the approach to be taken when trying to establish these relationships. Strauss works within an interactionist paradigm where axial coding is viewed as leading to an understanding of the central phenomenon in the data in terms of its context, the conditions which gave rise to it, the action and interaction strategies by which it is dealt with and their consequences (see Corbin and Strauss, 2008). Glaser (1992) takes a more purist grounded line. He argues that the axial codes, and the form that they take, should emerge from the data rather than being forced into any particular pre-determined format.

Whichever line is taken, axial coding is about in some way putting together again the data which have been effectively split apart into categories by open coding. As Mertens (2005) puts it:

> During this phase, you build a model of the phenomena that includes the conditions under which it occurs (or does not occur), the context in which it occurs, the action and interactional strategies that describe the phenomena, and the consequences of these actions. You continue to ask questions of the data; however, now the questions focus on relationships between the categories (p. 424).

If you are simply concerned with exploring or describing the phenomena being studied, this completes the analysis. However, grounded theory, as the term suggests, seeks to go further. For this, you need to go on to selective coding.

Selective coding

In this third stage, selective coding, you select one aspect as the *core category* and focus on it. The basis for doing this arises from axial coding which provides you with a picture of the relationships between various categories. In doing this, you should begin to get a feeling for what the study is about. In what way can you understand and explain the overall picture? This may well involve limiting the study to the major relationships which fit with this conceptualization.

In grounded theory, there must be a central integrating focus to those aspects which remain in the study. If more than one remain, the notion is that they have to be integrated into a single one at a higher degree of abstraction. This must remain grounded in the data but is abstract and integrated as well as being highly condensed. The core category is the centrepiece of your analysis. It is the central phenomenon around which the categories arising from axial coding are integrated.

Corbin and Strauss (2008) approach this task via the *story line*. This starts as a description of what axial coding has produced. You have to move from this descriptive account to a conceptualization of the story line. In other words you are seeking a core conceptual category which enables you to understand the story line.

Doing a grounded theory style analysis

The preceding section is intended to give the 'flavour' of a grounded theory analysis. Studied in conjunction with one or more of the examples provided on the website, it should be of assistance to anyone wanting to carry out an analysis following a general grounded theory style. Madill, Jordan and Shirley (2000) discuss two simple studies which use grounded theory to analyse interviews with relatives of individuals diagnosed as schizophrenic. They compare analyses carried out using realist, contextualist and radical constructionist epistemologies.

Much published research gives little or no attention to the theoretical background to grounded theory. It is simply regarded as a set of procedures, and as such can be regarded as a particular version of thematic coding analysis covered earlier in the chapter. Substantial further reading is called for if you wish to get on top of its theoretical background preparatory to doing a genuine grounded theory analysis. Punch (2005, pp. 204–12) provides a helpful introduction. He also cites Denzin's warning that, just as grounded theory is being widely adopted in many areas of social research:

> it is being challenged by a new body of work coming from the neighboring fields of anthropology and cultural studies . . . [They] are proposing that postmodern ethnography can no longer follow the guidelines of positivist social science. Gone are words like theory, hypothesis, concept, indicator, coding scheme, sampling, validity, and reliability. In their place comes a new language: readerly texts, modes of discourse,

> cultural poetics, deconstruction, interpretation, domination, the authority of the text, the author's voice, feminism, genre, grammatology, hermeneutics, inscription, master narrative, narrative structures, otherness, postmodernism, redemptive ethnography, semiotics, subversion, textuality, tropes (Denzin, 1988a, p. 432).

This is a reprise of an earlier theme in this book: the demise of positivism and the challenge of interpretive, constructivist and relativist voices (see Chapter 2). The argument put forward there was not to deny the value of this new body of approaches but to insist that there were continuing virtues in maintaining a broad scientific approach for real world research. And that realism provides a viable means of doing this in a post-positivist scientific era. Admittedly some of the language and terminology traditional in grounded theory harks back to an earlier era but there appears to be no basic incompatibility between grounded theory and realism. The approach is one way of finding out underlying structures and mechanisms, and realism has no quarrel with theory being generated from analysis of the data gathered in a study.

The website gives examples of research using a grounded theory analysis.

Alternative approaches to qualitative analysis

As discussed in Chapter 6, there is a wide range of other traditions of flexible research design additional to the three types selected as particularly appropriate for real world research. Of those covered in that chapter, the *phenomenological* tradition has a detailed and fully developed approach to analysis. Creswell (2007, pp. 57–62) provides an introduction; Moustakas (1994) a detailed account. See also Smith, Larkin and Flowers (2009), who discuss in detail a currently popular version known as 'interpretative phenomenological analysis'. There are also specific strategies for analysing text central to *hermeneutic* research. Bentz and Shapiro (1998, pp. 105–20) provide references. *Narrative methods* also have distinctive approaches to analysis. Creswell (2007, pp. 54–7) provides an introduction; Riessman (2008) a detailed account.

There are many other possibilities. The reader still unsure about which route to take may well find inspiration in Coffey and Atkinson (1996) where a single data set is analysed using a range of different strategies including different versions of *narrative analysis*, linguistic or *semiotic analyses*, types of *textual analysis* and interpretive or *hermeneutic* goals.

Integrating qualitative and quantitative data in multi-strategy designs

Multi-strategy (mixed methods) designs, discussed in Chapter 7, are characterized by the collection of substantial amounts of both qualitative and quantitative data. Depending on the particular design selected qualitative data collection may precede

that of quantitative data, or vice versa, or the collection may be in tandem. The techniques and approaches discussed in this chapter and the preceding one can be used for separate analysis of the two types of data. Again depending on the type of design, and in particular on your research questions, such separate analyses and their interpretation may be all that is envisaged. Bryman (2007) makes the point that, even when the initial intentions are not to integrate the findings, there can be value in exploring the connections between the qualitative and quantitative findings. Seeking the opportunity to integrate them in some way takes fuller advantage of the potential benefits of multi-strategy designs, as rehearsed in Chapter 7, Box 7.2 (p. 167).

Onwuegbuzie and Teddlie (2003, p. 375) suggest the following steps in the process of data analysis (not necessarily followed in a linear sequence):

1. *Data reduction*. Involves summarizing quantitative data (e.g. using descriptive statistics) and qualitative data (e.g. using thematic analysis).
2. *Data display*. Using tables, graphs, etc. with quantitative data and matrices, charts, networks, etc. with qualitative data.
3. *Data transformation*. 'Qualitizing' quantitative data and/or 'quantizing' qualitative data.[2]
4. *Data correlation*. Correlating quantitative data with qualitized data.
5. *Data consolidation*. Combining both data types to create new variables or data sets.
6. *Data comparison*. Comparing data from different data sources.
7. *Data integration*. Integrating all data into a coherent whole, or separate quantitative and qualitative coherent wholes.

Slate *et al.* (2009) provide a clear example of the process involved. Caracelli and Greene (1993) discuss several of these strategies in detail and consider their appropriateness for different research designs.

Bryman (2007), in a paper on the barriers to integration, focused on the degree to which researchers link the quantitative and qualitative in the course of analysing and writing up their findings. He cites evidence that reviewers have found this a problem area in published research. His own research (Bryman, 2006a) which carried out a content analysis of over 200 mixed methods research articles, found fewer than one in five 'genuinely integrated' their qualitative and quantitative findings (i.e. analysed, interpreted and wrote up the research in such a way that the two components were mutually illuminating).

Integrated analyses

Rather than viewing the qualitative and quantitative data analyses as initially separate, with or without subsequent attempts to integrate the two sets of findings, an alternative approach is to bring the two aspects together at an early stage. This strategy is commonly used in case study research, as discussed in Chapter 6, p. 135, although for largely

[2] As the term suggests, 'qualitizing' refers to turning quantitative data into a qualitative form (e.g. by converting them into narrative codes that can be analysed qualitatively). 'Quantizing' refers to turning qualitative data into a quantitative form (e.g. by determining the frequency of occurrence of themes in a thematic analysis). While quantizing involves data reduction, qualitizing is an interpretive step which adds information (Maxwell, 2010).

BOX 17.9

Strategies for integrating quantitative and qualitative data through analysis

- Employment of the results from analysis of one form of data in approaching the analysis of another form of data.
- Synthesis of data generated from a variety of sources, for further joint interpretation.
- Comparison of coded or thematic qualitative data across groups defined by categorical or scaled variables (matched, where possible, on an individual basis).
- Pattern analysis using matrices.
- Conversion of qualitative to quantitative coding to allow for descriptive, inferential or exploratory statistical analysis.
- Conversion of quantitative data into narrative form (e.g. for profiling).
- Inherently mixed data analysis, where a single source gives rise to both qualitative and quantitative information (e.g. some forms of social network analysis).
- Iterative analyses involving multiple, sequenced phases where the conduct of each phase arises out of or draws on the analysis of the preceding phase.

(based on Bazeley, 2009, p. 205)

historical reasons it is usually discussed under a separate heading to multi-strategy designs.

Bazeley (2009), in a review of current developments in the integration of data analysis, lists a range of possible strategies summarized in Box 17.9. She also points out the value of computer software in the task of integrating analyses. The use of the widely available spreadsheet software package Excel was discussed in connection with quantitative analysis in Chapter 16, p. 414 – see also Appendix A. It is also a useful tool for tasks involving synthesis of varied forms of data from a range of sources (Niglas, 2007). Several of the Computer Assisted Qualitative Data Analysis (CAQDAS) packages discussed in Appendix B (e.g. NVivo, MAXQDA and QDA Miner) although designed primarily for qualitative analysis, support the combination of quantitative variable data within the qualitative database for matrix-based analyses of coded text and the conversion of qualitative coding to variable data (Bazeley, 2006).

The website gives examples of research integrating the analysis of qualitative and quantitative data.

Further reading

The website gives annotated references to further reading for Chapter 17.