

# Chapter 2: Three Approaches to Qualitative Data Analysis

## Introduction

In this chapter, you will learn about the fundamental approach to qualitative data analysis: **thematic analysis**. This is a generic approach to data analysis that enables data sources to be analysed in terms of the principal concepts or themes. These themes are developed by the analyst to enable the data to be reduced to key ideas.

Having examined this basis for qualitative data analysis, we will then consider a particular approach to thematic analysis that has been influential in qualitative data analysis: **grounded theory**. This is both a methodology and an epistemology, and seeks to ensure the validity of qualitative data analysis, by 'grounding' the analysis in the concepts used by the respondents.

Finally, we look at a technical approach to qualitative data analysis that is used in the analysis of 'talk': **discourse or conversational analysis**. This provides a very detailed means to document and analyse qualitative data, and is of use when the way in which something is said is as important as what is said.

In the next two chapters, we use these basics of qualitative data analysis practically, to undertake manual and computer-aided data analysis.

## Thematic Analysis

In their introduction to thematic analysis of qualitative data, Marshall and Rossman (1999: 150) suggest that data analysis is the process of

'...bringing order, structure and interpretation to the mass of collected data. ... It is the search for general statements about relationships among categories of data ... it is the search among data to identify content.

The principal technique that is used by qualitative researchers to analyse data is thematic analysis. This process may be based on prior categories, or on categories that become clear to the researcher only as the analysis proceeds.

For example, in a study of the work that is done by a health visitor, a researcher might (based on previous research) expect that the work will be divided into casework with individual clients, community development work, public health activities

and administration. These four categories would become '**themes**' that would be used by the researcher to analyse interview data once it has been collected from respondents. This would be an example of a **pre-figured** or '**objective**' strategy for thematic analysis.

On the other hand, the researcher might decide that s/he cannot decide in advance what are the themes. For example, a researcher conducted interviews with British Asian women concerning healthy eating. No research had been conducted on this topic previously. Only when the data was read through could the researcher begin to identify what themes were common to the interviewees. She found emerging themes related to cultural preferences, individual preferences, knowledge and affordability. This is an example of an **emergent** or **intuitive** strategy (the intuition refers to the researcher's capacity to discern the important themes in the collected data). We will look more closely at the latter kind of analysis when we consider **grounded theory**.

### **Stages in Thematic Data Analysis**

Marshall and Rossman (1999) suggest that thematic analysis can be divided into six phases:

- organise the data
- generate categories or themes
- code the data
- test emergent understandings of the data
- search for alternative explanations of the data
- write-up the data analysis

We will look at each of these in turn.

### **Organising the data**

The first and crucial step in thematic analysis is that of **familiarisation** with the data. This cannot be skipped without great risks to validity of the analysis, and there is no shortcut to reading the data transcripts, possibly a number of times. This work precedes any efforts to identify themes or test theories, but during the reading process, some broader **understanding** of the data will begin to emerge.

There is of course, the practical issue of organising data: this may take the form of a mass of documents: not only transcripts of interviews but also field notes, documents, photographs, diagrams and scribbled ideas or notes. It is important that you find some kind of way of archiving this material, and this

probably requires indexing every document, with a key index so you can quickly lay your hands on individual documents. Where these can be entered into a computer-aided qualitative data analysis package, this can provide a neat way to archive all your data, but it may still be the case that some kind of physical archive is also required.

However, the main task in this phase of analysis is reading the data. It is during this process that you can start to understand the data.

During reading, you may wish to take notes, perhaps using index cards to remind yourself of something that strikes you. In some software packages, this is known as a **memo**. Usually these notes are not formal themes but more general insights into what is being said, or examples of something unexpected or remarkable.

There are three kinds of reading, according to Mason (2002: 148-150): **literal**, **interpretive** and **reflexive**. Literal reading concerns itself with the structure of transcript, document or other data, simply focusing on how it is constituted (for example, that a document is written for internal consumption within an organisation and takes the form of a confidential report.

Interpretive reading will ‘... involve you in constructing or documenting a version of what you think the data mean or represent, or what you think you can infer from them (ibid: 149). You may focus on the respondent’s own interpretations or impose your own meanings, depending on your epistemological stance and your research question.

Reflexive reading ‘... will locate you as part of the data you have generated , and will seek to explore your role and perspective in the process of generation and interpretation of data’ (ibid). Postmodern approaches to research have introduced reflexive reading as a key aspect, but most qualitative studies incorporate an element of this approach alongside the first two.

### **Generating Categories and Themes**

The process of category generation involves noting **patterns** in the data, perhaps relating to the topics described by interviewees, or in how they describe aspects of what they are describing. For example, as you read an interview, you are reminded by something said by another interviewee, or a recurrent theme in what one interviewee has to say. Where categories for analysis are pre-selected, these should form the basis for the patterning of the data. Where they are emergent, they need to be exhaustive, and it may take a considerable time to find the right categories to sufficiently explicate the data.

Categories should be **internally consistent and externally divergent** (Marshall and Rossman 1999: 154). This means that a category should link together things that are the same as each other, but be distinct from each other. In the example of the study of Asian attitudes to healthy eating above, the four categories fulfil these criteria and as such are useful themes in the data. If two categories overlap, then they are not suitable as distinct themes for analysis. If some data is not covered by the categories developed, the typology may either be incomplete, or the categories developed may be inadequate to cover all the emerging ideas in the data.

Categories may be uni- or multi-dimensional. In a two-dimensional matrix: categories intersect. For example, in a study of health visitors, Twin (1991) suggested that their work could be categorised into a 2 x 2 matrix, where the dimensions were 'individual' versus 'collective' focus and 'development' versus 'information-gathering'. Each box of this matrix contained some aspect of the professional role.

'Themes' may best be understood as 'super-categories': high-level categories that provide an overall structure to the data. Themes provide not only this further level of categorisation, but also a basis for structuring a write-up of data. In my view, a qualitative data analysis should end up with four or five themes: this is a manageable number and makes for a readable data report. Within each theme, there will be one or more categories, and these will obviously be related or they would not fit together as a theme.

This stage is the most critical and the most time-consuming. We will consider the development of categories in more detail in chapter 3.

## **Coding Data**

Coding is a straightforward process once the appropriate patterns of data have been identified, and it is at this stage that software packages come into their own, as they permit storage of large amounts of coded data.

The process of coding requires the application of the set of categories (be they pre-figured or emergent) to the data in a concerted way. Typically, all data should be coded systematically. In manual coding, a code can be written alongside the passage that reflects a theme. In computer-aided analysis, this is done electronically. Software packages have the advantage that the same group of words can be easily coded with more than one categorisation.

For example, interviewee A commented:

‘... I usually buy food from the local shop that sells Asian ingredients, but this more expensive and I may go tot the supermarket for some food ...’

This might need to be coded into two categories: *culture* and *affordability*.

In manual coding, different colour highlighting pens can be used to identify categories, and this permits a quick visual way to keep tabs on each of the different themes in the data.

The purpose of coding is two-fold:

- a) to apply the categories to the data
- b) to enable examples of the data to be used in the write-up of the qualitative data analysis

Computer software has the advantage over manual approaches that once a) has been completed, it is easy to recover data for b).

### **Testing Emergent Understandings**

As categories and themes are developed, some kind of understanding of the data can begin to emerge, including the development of **theoretical constructs**. A qualitative data analyst will start to get a sense of ‘what the data means’ and place the data in a context of theory: applying either established theory or new theory. The extent to which these constructs emerge before or during a qualitative data analysis will depend on whether the study applies a pre-figured or grounded approach.

Marshall and Rossman (1999) suggest that in this phase of qualitative data analysis, a researcher should search the data to challenge the emergent understanding, seek out negative instances that undermine this understanding and start to draw categories of data together to establish the main themes.

### **Search for Alternative Explanations**

During data analysis, a researcher should not commit too quickly to one explanation of the data, but should play ‘devil’s advocate’, seeking alternative understandings of the data, and even trying to undermine the theses that are being used for analysis.

### **Writing the Report**

The issues in writing-up qualitative data analysis will be considered at some length in chapter 5, where we will discuss the application of different epistemological strategies in the

research report, including reflexive reporting, in which the voice of the author is an important aspect of the data as reported.

The main problem that arises in the writing of qualitative data analysis concerns the relative weight of the analyst's interpretation. While it is clear from what has been written so far that qualitative data analysis is all about interpretation, approaches such as grounded theory attempt to minimise the authorial voice, creating an **objective** or **realist** account by focusing on the meanings that the participants use, as opposed to those imposed by the researcher. This approach is criticised by **constructionists**, who argue that all analysis is inevitably a work of interpretation, and that for reasons of honesty, writing-up should make this clear by the application of reflexive techniques and an open-endedness concerning the 'truth' of the report.

These phases of qualitative data analysis are fairly generic, and provide the basis for most approaches. On occasions (for example when documenting a 'case-study'), thematic analysis may be replaced by a more narrative structure, in which categorisation, themes and theory play little part. However, for most analysis, this form of thematic approach will provide a good basis for qualitative data analysis.

Please undertake the following exercise before moving on with this chapter.

### **Reflective Exercise 2.1 Stages of Qualitative Data Analysis**

Read the following article, which reports a qualitative research study.

Hibbert, D et al (2002) 'Consumerism and professional work in the community pharmacy', *Sociology of Health & Illness*, 24 (1) 46-65.

Then answer the following questions

1. What approach to data analysis did the authors take?
2. What are the main themes in the analysis?
3. What theoretical constructs are developed by the authors?
4. Do the authors discuss any issues in writing-up the data analysis?

## Grounded Theory

For many, grounded theory is synonymous with qualitative methodologies in research. However, it is not simply a methodology, it is also an epistemology, concerned with the nature of knowledge and the possibility of knowing the ‘truth’ about the world. As this chapter has already indicated, grounded theory is not the only approach to qualitative data analysis, indeed it is a very specific form of analysis.

Grounded theory was established by Glaser and Strauss (1967) and applied by many qualitative researchers in the 1970s and 1980s. While it was instrumental in countering claims that qualitative research was ‘woolly’ and unsystematic, its realist epistemology is not conducive to more recent developments in qualitative research, most specifically constructionist or postmodern research approaches that doubt the possibility of finding a simple truth from research data. Grounded theory is however still used and whether or not you align yourself with realism, the fundamental approaches in grounded theory are important and can be highly valuable as a means to reduce the ‘bias’ of interpretation when conducting qualitative data analysis.

The main principles of grounded theory are as follows:

- theory should be grounded in the data gathered in a study rather than imposed from a previously existing framework
- theory can be refined by further data collection, so in a grounded theory approach data collection and analysis should be iterative (the ‘constant comparison’ of data and theory approach)
- theoretical sampling (non-random sampling that selects cases to supply a wide range of responses) is used to ensure that data is collected that can assist in the development of the grounded theory
- a researcher using the grounded theory approach needs to be aware of her ‘conceptual baggage’, which could bias the emergent theory
- theory emerges through immersion in the data and the development of a coding framework that is entirely based on the structure of the data.

The techniques of grounded theory are those that have already been described in thematic analysis, but with a much stronger emphasis on the need for categories not to be imposed from a pre-figured frame of reference. Grounded theory researchers are encouraged **not** to conduct a literature review before commencing research, in case this adds to the conceptual

**Realism is the epistemological view that there is an underlying reality that can be uncovered by careful research. This reality may not be obvious to participants in a setting.**

baggage! Immersion in the field is the only way to gain a theoretical framework that is 'true' to the participants in the setting. As noted, collection and analysis of data should proceed hand-in-hand, so even after one interview, analysis might begin: categories and early theoretical constructs then inform the shaping of subsequent data collection.

The writing of grounded theory is a critical element of the process, as it is here that the theoretical framework that supplies the understanding of the data is elucidated. Typically, large examples of the raw data are included, to demonstrate that theory is truly grounded in the data: there is a sense in which the data will 'speak for itself' and the role of the qualitative data writer is simply to organise this in a comprehensible way.

Grounded theory has been influential in the development of software, and many of the packages available have adopted analytical techniques that mirror the kinds of grounded theory analysis that have been developed by proponents of the approach such as Strauss and Corbin (1990). Thus for example, categories can be set into hierarchies, and 'memos' can be used to 'remind' a researcher of insights into the theoretical constructs that the data suggests. Coding in grounded theory should not be constrained by artificial limits on numbers of categories, and again, software applications are well suited to this kind of extended analysis.

The main limitations of grounded theory are:

- that in many cases, qualitative data analysis is not primarily concerned with theory generation, but is an opportunity to apply existing theory to a setting. This is most relevant to case studies, where the intention is not to generalise but to document what goes on in a setting in great detail.
- epistemologically and ontologically, grounded theory is neither possible (because all analysis is also interpretation) nor desirable, because it assumes there is a single truth. Rather, there may be many different interpretative frameworks in use by participants in a field setting, and it is these that should be described, not the 'reality' that an external researcher 'discovers'.

Grounded theory is a technique that has a place in qualitative data analysis, and you need to be familiar with this influential approach. Please undertake the following reading and exercise.



## SAQ 2.1 Grounded theory

Please read the chapter from Seale (1999) in the supplementary reading and answer these questions

1. In what ways is grounded theory similar to positivist quantitative data analysis?
2. What is meant by theoretical saturation?
3. What are the four stages in constant comparative method?
4. What is the difference between **open coding**, **axial coding** and **selective coding**?

## Conversation and Discourse Analyses

Conversation analysis (CA) and discourse analysis (DA), while different in methodology, both emerged from an interest in 'talk': the inherent structure of texts -- be they documentary or interactional (interviews or ethnography). Both use a particular form of notation that details aspects of talk such as pauses, interruptions, emphases etc.

We will not devote too much room to these approaches, as they are less common in qualitative data analysis. DA, however, is significant in some areas of social psychology research and will therefore be considered as an alternative to thematic analysis.

### Conversation Analysis

Silverman (2001) summarises the principles of CA as follows:

1. We can discern in texts organisational structures that are capable of analysis independent of the characteristics (psychological, emotional) of their producers.
2. Texts can be analysed in terms of sequences (for example, turn-taking in a conversation)
3. CA requires precise analysis of detailed transcripts that provide more than just the words spoken. Some CA

analysts have adopted the use of video recording to maximise the data gleaned from talk.

CA has been used to analyse such interactions as the opening sequence (for instance between a GP and a patient), to consider the turn-taking behaviour in conversations, and also to evaluate the effect of context on what is 'permissible' in a conversation (for example, a patient may be expected to respond to a GP's questions but not ask personal questions back).

### **Answers to SAQ 2.1**

1. In quantitative data analysis, the objective is to develop falsifiable statements, and the refinement of theory comes about by the discovery of negative evidence. In grounded theory, the constant comparative method allows the falsification of theory and its replacement with more sophisticated, grounded constructs.
2. Theoretical saturation comes about when no new data that permits new categories to be developed can be collected, however many new sources are studied. At this point, data collection may end.
3. The four stages are:
  - incidents are collected together into a category with certain theoretical properties
  - Categories are integrated and analysed for interactions
  - Theoretical saturation
  - Writing the grounded theory analysis in terms of the categories and their interactions, with plentiful examples from the data.
4. Open coding is the fundamental process of categorisation, while axial coding concerns the interactions between categories. In selective coding, one category subsumes others and forms the basis for the emergent theory. You can see the notation used for CA in the supplementary reading included for this part of the unit (see below).

The methodology of CA rests upon some techniques for talk analysis. Silverman (1999: 177) summarises these as follows:

1. Try to identify sequences of related talk
2. Examine speaker roles or identities from their talk
3. Look for outcomes in the talk (a request for clarification, laughter): and then trace back the roots of that outcome
4. Do not seek understanding in terms of a speaker's intentions, or role as discerned by information external to the talk under analysis
5. Do not try to make sense of a line of transcript independent of surrounding talk

### Discourse Analysis

DA has been used to explore a range of social science topics, but with a focus on the use of language as the means of social interaction. It is not interested in the 'underlying reality' of a situation, but rather with the way that participants in a setting or interaction construct their social worlds and identities reflexively. As such it has something in common with constructionist or postmodern social theory, although its roots (like CA) are in ethnomethodology.

Some proponents of DA have applied this approach to gaining understanding of social institutions, and in the extract you will read, Silverman describes the analysis of 'science' and 'motherhood'. In both case, DA draws on the strategies that are used in talk or other texts to enable participants to achieve their roles, identities or aspirations and sustain these during interactions with others. DA theorists often refer to the creation of **routines** in participants' talk in order to achieve their objectives, or to establish their 'moral' rights to be considered in a certain way by others.

Thus, for example, doctors may use certain forms of talk to routinise their authority in a setting. In a classic example, a doctor even queried whether a mother accurately recalled the number of children she had when she contradicted his claim that she had two offspring!

Methodologically, DA requires detailed analysis of sequences of talk in order to discern the strategies adopted by participants in furtherance of their ends. Transcripts are analysed not as reflections of an underlying reality, but as a construction that is artfully created by a speaker. The role of the analyst is to expose this artful work of construction.

**Ethnomethodology is an approach in sociology that focused on how participants ('members') in a social setting achieved certain outcomes, for example how they 'do' ordering a meal in a restaurant or persuade a colleague of a point of argument.**

To complete this section on CA and DA, please read the extract from Silverman (2001) in the supplementary reading and then do the following exercise.

### **SAQ 2.2 CA and DA**

1. In the example of TV news interviews, in what ways does the context affect the talk?
2. What does Potter mean when he says that DA is **anti-realist**?
3. What is a 'script' in DA?

### **Conclusion**

In this chapter you have been introduced to the principles of thematic analysis, which we will focus upon as we now turn to the practicalities of doing qualitative data analysis. We also looked at a number of different approaches, including the analysis of talk in CA and DA

Before moving on to the practical skills in qualitative data analysis, please complete the following reflective exercise (next page) on the theory that we have considered in this chapter.

### **Further Reading**

Potter, J. (1997) 'Discourse analysis as a way of interpreting naturally-occurring talk', in Silverman, D. (ed.) Qualitative Research: Theory, Method and Practice. London: Sage; pp. 144-60.

Strauss, AL. and Corbin, J. (1990) Basics of Qualitative Research: Grounded Theory Procedures and Techniques. London: Sage.

## Reflective Exercise 2.2 Strengths and Weaknesses of Qualitative Analysis Approaches

You have now been introduced to four approaches to qualitative data analysis. Please reflect on your thoughts concerning these, and what you consider their strengths and weaknesses, or how you might apply them to qualitative data.

| Approach                | Strengths | Weaknesses |
|-------------------------|-----------|------------|
| Thematic Analysis       |           |            |
| Grounded theory         |           |            |
| Conversational Analysis |           |            |
| Discourse Analysis      |           |            |

### Answers to SAQ2.2

1. The interviewers use techniques to suggest their impartiality. They limit their input to a conversation to asking questions while interviewees limit themselves to replying.
2. What is said is not a representation of reality: talk produces versions of the world, a setting or an identity.
3. A script is a way of invoking the supposedly routine character of what is described by a participant. It is a device used to make certain claims to authority, moral right and so on.