13

Qualitative Data Analysis

Chapter Summary

Introduction

Qualitative research produces richer, thicker descriptions of social situations, people, and their views, and presents the social world from the point of view of participants. For these reasons, qualitative data analysis is more exciting but also more challenging than quantitative analysis. This chapter discusses two general strategies of qualitative data analysis—**grounded theory** and **analytic induction**—and shows how these methodologies propose to approach qualitative data. It also looks at coding qualitative data and how the use of computer software may facilitate this process. In addition, the chapter considers criticisms of coding and discusses more holistic analysis techniques, described generally as "narrative analysis."

Although qualitative analysis is much less straightforward and offers fewer elaborated rules and strategies than the quantitative one, there are two broader directions it often takes:

1. Using *general strategies* such as analytic induction and grounded theory;
2. Employing basic operations, such as coding and narrative analysis.

General Strategies of Qualitative Data Analysis

Analytic Induction

Analytic induction is an example of qualitative research that uses iterative process. It begins with a very general research question or a definition of the research problem, and then refines the question during additional data collection through ongoing analysis. The following steps are taken in the analytic induction research:

- Begin with a general research question.
- Gather some data.
- Propose a hypothesis.
- Continue data collection—If a case is found that is inconsistent with the initial hypothesis (or explanation), the hypothesis is redefined to exclude the case, or the hypothesis is dropped or fundamentally revised to include the case that did not fit the original hypothesis

The researcher continues to gather data until no contradictory cases are found.

While *analytic induction* proposes a rigorous process of analysis trying to account for all encountered cases, it faces some practical difficulties in implementation. The difficulties with analytic induction are that (1) the hypotheses generated in the process may be too broad to be useful because all cases must be explained, and (2) there are usually no guidelines on how many cases must be reviewed before the validity of the hypothesis is accepted. For this reason, analytic in-

duction is used less in recent sociological research, although it remains a research framework guiding the logic of qualitative analysis.

Grounded Theory

Grounded theory is also the strategy of research based on an inductive, iterative process. This strategy is defined by its creators as a theory derived from the systematically gathered and analyzed research data, a theory that reflects the data as closely as possible (Strauss and Corbin, 1998). Grounded theory is not so much a theory as it is a recurring and iterative process of simultaneous data collection and analysis, where data collection and analysis reinforce each other. Developed from 1960s and until now, grounded theory has several versions, and analytical steps proposed in it differ from one version to another. In addition, its core notions are interpreted a bit differently by several authors. For example, the difference between the notions of *concepts* and *categories* is not always clear.

Despite these differences, the key processes that make up grounded theory are coding, constant comparison of data, and theoretical saturation:

1. *Coding*: the very beginning process in data analysis and development of theory, where the data are broken into parts which are labelled and given names (codes). Coding begins soon after the data collection has started. The given codes are later reconfigured into **concepts** and **categories**, theoretical abstractions that describe a particular type of behaviour, perceptions, or processes.
2. *Constant comparison*: new data are constantly compared to the old ones to create and redefine concepts and categories. Writing *memos*, or broader comparative notes on cases and concepts, is instrumental to development of theory and differentiating between different concepts and categories.
3. *Theoretical saturation*: the point in developing concepts and categories when examining new cases adds no new information to the existing ones; new data no longer contribute to the development of theoretical generalizations.

Coding starts in in the early stage of research project. It is the labelling of activities and issues that are being observed and grouped together. It is the first step in interpreting data and developing a theory. The data are treated as potential indicators of concepts, and are repeatedly compared with each other to see what concepts or categories arise. Strauss and Corbin (1990) distinguished three types of coding:

1. *Open coding*
   - It is the process of breaking down, comparing, and conceptualizing data. Open coding identifies initial concepts that will be grouped into categories later.
   - Open coding generates a lot of concepts and is kept as close to the data as possible. Concepts are often developed in line-by-line coding. While concepts are initial theoretical labels for data, **categories** group specific concepts into broader classes of ideas.
2. *Axial coding*
   - After open coding, the data are reviewed for linkages and re-organized according to those connections. We link codes by making connections between contexts, patterns of interaction, causes, or consequences.
   - As the result of axial coding, the categories can be connected to the contexts in which they were expressed.
3. *Selective coding*
   - Selection of the core category, relating it to the other categories
   - Validating the relationships between categories
   - Identifying gaps in categories that need to be filled in, further elaboration and refining of existing categories
   - Conceptualizing the phenomenon (theory emerges)

Not all grounded theory practitioners use the three types of coding. Some, like Charmaz (2004), have issues with axial coding arguing that it closes the initial open coding too early when categories are developed and that this prevents development of new concepts. Charmaz puts more emphasis on the open or initial coding and then on focused selective coding. Detailed initial coding, according to Charmaz, is necessary to generate as many new ideas as possible, while the selective or focused coding is needed to flag out the codes that appear most frequently or are the most revealing.

Outcomes of Grounded Theory

What does the grounded theory produce? There are several potential outcomes of grounded theory:

- *Concepts*: these are "the building blocks of theory" (Strauss and Corbin, 1998) and are produced in open coding
- *Categories*: categories encompass two or more concepts and are ideas at a higher level of abstraction. For example, a category of "happiness" can encompass concepts of "achievement," "self-realization," "fulfilling personal relationships," "good health," and others.
- *Properties*: attributes of a category
- *Hypotheses*: initial propositions about relationships between categories
- *Theory*: a set of categories interrelated through statements that explain some social phenomenon. Grounded theory differentiates between *substantive theory*—a set of statements about social phenomenon in a specific setting (e.g., racial prejudice in a hospital)—and a *formal theory*, a set of statements at higher level of abstraction that have applicability to several settings and wider area (e.g., racial prejudice in several areas or institutions).

Research based on grounded theory first with a *general research question*, then selects a *theoretical sample* (a sample where cases are chosen on the basis of fitting into a specific theoretical category or notion), *collects data*, and assigns *initial codes*. After these first four steps, the constant iteration between selecting more cases and data analysis begins. New cases are coded, compared among each other and new concepts and categories are created. Simultaneous data collection and analysis continues until a point of **theoretical saturation** occurs, or new data do not generate new concepts or categories. After that, the test of emergent hypotheses leads to formulation of a substantive theory, which can then be generalized to a formal theory when different contexts are researched.

During the process of grounded-theory research, **memos** are a key analytic tool. They are written by the researchers as summary documents and provide reminders about what is meant by different concepts and categories. Memos are also important in conceptual and theoretical comparison between cases and aid in the ongoing reflection throughout the research project. Wring memos can help researchers to refine their ideas and elaborate them.

Criticisms of Grounded Theory

Grounded theory has been criticized on several grounds:

- It may not be as "theory neutral" as claimed. Although analysis in grounded theory approach starts with data, researchers routinely have an exposure to sociological theories even before starting a project, and are influenced by theories during the data collection and analysis. That prior knowledge likely has some impact on the observations they make and interpretations of them. Hence the suspension of the prior theoretical knowledge is quite difficult, if not impossible. What we see in the social world is conditioned by our prior knowledge.
- Grounded theory approach requires a lot of time for both the data collection and transcription. Transcriptions can be quite tedious, and the constant moving between data collection and analysis required by this approach is also a time-consuming process. It takes time to organize and work through all the material collected in the project.

- It is not clear whether the grounded theory approach actually generates *theory* as a way of explaining social phenomena. Although the approach claims to generate formal, more abstract, theories, most often it generates rather specific theories applicable only to specific contexts, rather than to a broad range of contexts and situations.
- The grounded theory approach tends to fragment the data into the less meaningful bits of information that take the observed activities out of context. Critics argue that this results in the loss of the narrative flow in the collected qualitative material.

Despite these criticisms, grounded theory remains one of the dominant approaches in the analysis of qualitative data. Its ideas of coding and memos influence other approaches, and even the computer software for qualitative data analysis is developed with keeping the techniques of grounded theory in mind.

Basic Operations in Qualitative Data Analysis

Steps in Coding and Coding Concerns

Most qualitative data analysis techniques start with coding the text. **Coding** should answer the question about what does the given piece of data represent, what category it describes. It can also mark a particular type of behaviour or participants' reflections on their behaviour, or even suggest a research question to investigate. Coding is a complicated process, and several suggestions can help to make it more manageable:

- *Code and transcribe as soon as possible, as data are being collected.* This improves the understanding of the data and helps with theoretical sampling. Starting coding early also alleviates the feeling of data overload, which might emerge in the face of large quantity of documents or interviews to analyze.
- *Read through the data before considering any interpretation and starting coding.* This first read-through should help to create the very first impressions of the data and produce initial reflections at the end of the read-through on what is important, interesting, or striking.
- *Read through the data again, noting keywords and themes.* This second read-through implies making notes on the margins about significant ideas or themes noticed in the text. As many notes as necessary can be made, and this process actually describes *coding*, or generating concepts.
- *Too many codes should not be generating concerns, particularly at the early stage of analysis.* The first run of coding should be as inclusive and as inventive as possible, even if it generates too many codes. The codes can be tidied and reduced later. It is important to note that one piece of data can often be coded in more than one way, and it is often better to double-code the same passage under different concepts to preserve the complexity of the data and open new lines of inquiry. Some grounded theory practitioners even advocate line-by-line coding to stay connected to the data. The codes are later considered for similarities and grouped into broader concepts.
- *Review the codes to consider associations, redundancy, relationships to existing concepts.* This is the actual process of grouping codes. Some codes may describe the same phenomena, others speak of the concepts mentioned in the literature. The connections between codes should also be noted, such as if respondents commonly describe one thing as the cause of the other.
- *Consider general theoretical ideas related to codes and data.* This step generates initial theoretical ideas, describes connection between concepts, and develops categories and hypotheses.
- *Keep coding in perspective.* The codes have to be reviewed for their broader significance for the research and interpretation. What do the data mean? What is its relevance to the literature? What do they tell about the lives of the studied people?

Turning Data into Fragments

Coding essentially means cutting the data into smaller, more manageable pieces that can be described by concepts. Written data must be coded and then categorized in a way that it is useful for analysis. The data are coded in relation to themes, the research question, content of the research project, and the analytic context.

While grounded theory proposes its own ways of coding the qualitative data, general approach to qualitative coding presupposes three different levels (Coffey and Atkinson, 1996):

1. *Basic coding*: getting the simplest labels for the material, such as negative and positive consequences of the analyzed activity. This type of coding will produce only a superficial analysis, and needs to be followed by other steps.
2. *Deeper awareness of the content* in the text, of how it relates to the focus of the research, of what is included and what is missing.
3. *Exploring broader analytic themes*: this requires the researcher to move away from the content of specific interviews and look for broader analytical themes.

In Coffey and Anderson's approach, qualitative coding should not be thought of as simply a way of fragmentation and retrieval of data. Rather, they defend the holistic analysis of qualitative material, which can explore interconnections between concepts and reveal a broader narrative in the data.

In the recent years, computer programs for qualitative analysis (CAQDAS, or computer-assisted qualitative data analysis software) gained popularity. QSR NVivo software is one of the leaders in the field, proposing many visual, analytical, and quantitative techniques for analyzing qualitative material. There is a debate on whether computer software is suitable for qualitative data analysis, since some researchers argue that the software primarily works to quantify coded text and therefore negates the qualitative, thematic interpretation of meaning, which is a key goal of qualitative research. Another criticism is that qualitative analysis software fragments textual material into very small pieces of data where the natural flow of the story is broken and the holistic integrative aspect of qualitative analysis is lost. An additional reproach to qualitative software is that it is too closely built around the ideas of systematic inquiry characteristic of grounded theory and this diminishes another key strength of qualitative research, its flexibility.

The supporters of using software in qualitative research point to its advantages:

- The software proposes new visual ways of looking at data which stimulate its holistic perception and point to connections between ideas and concepts (e.g., building "trees" of concepts).
- Qualitative software allows researchers to estimate the representativeness of different quotes from qualitative interviews, in response to quantitative critics who argue that such quotes are "anecdotal evidence" with no indication of their overall frequency or significance.
- Software improves the transparency of qualitative analysis, as researchers are more explicit about how they conduct the study and describe how they used CAQDAS.

Problems with Coding

Both the work of Coffey and Anderson and the discussion around using the computer software for qualitative analysis point out to *two main criticisms* of coding:

1. *Risk of losing the context*: by selecting parts of the text, qualitative coding weakens the connection between description and social setting in which the events occur, therefore resulting in a loss of context.
2. *Fragmentation of data*: breaking the data into codes and small chunks of text results in a loss of narrative.

These criticisms lead researchers to increasingly turn to the more integrative types of qualitative analysis, such as *narrative analysis* and *life history* approaches. However, coding remains influential due

to its background in grounded theory and due to a wide use of computer programs in qualitative analysis.

Regardless of how the qualitative data are analyzed, the final results of any research should present a balanced account that incorporates both the data and theoretical considerations. Interpretation and theorizing are as necessary in qualitative research as in any sociological inquiry, and focusing only on a description of data does not do justice to the broader goal of sociology, that of connecting theoretical questions to available evidence.

Narrative Analysis

Narrative analysis is an umbrella term for many approaches used to analyze the stories people tell in order to understand their life and the world as a whole. The term became particularly prominent in the recent years because of developments in life history and biographical approaches. Narrative analysis pays attention to how people link their accounts of past, present, and future; it focuses on the context and events *and* on interpretations people make of them. The focus in narrative analysis shifts from *what* happened to *how people make sense of it*. This brings life into perspective as an experience.

Proponents of narrative analysis stress that most approaches to qualitative analysis ignore the fact that people perceive their lives as continuity and process; they make sense of it through stories, and therefore the narrative analysis helps to understand the perspective of those being studied.

Four models of narrative analysis are suggested by Riessman (2004):

1. *Thematic analysis*: What is said rather than how it is said.
2. *Structural analysis*: The way a story is told, what is emphasized in order to persuade the listener.
3. *Interactional analysis*: Examines the dialogue between the teller and the listener, to see how meaning is co-constructed.
4. *Performance analysis*: Examines the narrative as a performance, the use of words and gestures to get the story across.

Riessman also advocates asking particular questions during qualitative interviews that are more likely to elicit narratives, such as prompts or elaborating questions. Other qualitative researchers note that stories should be examined in terms of functions they serve to the teller (Coffey and Atkinson, 1996). In turn, Miller (2000) argues that what is important in narrative analysis is not reconstruction of the facts, but rather how people reveal their lives in their reflections and interpretations. In short, narrative analysis is an approach that examines the stories people use to explain the events.

Critics of narrative analysis stress its over-reliance on the story and its preference for the accounts of the participants as the only explanations of social phenomena. Bury (2001) makes these points in relation to different studies about "coping with" and "normalizing" illnesses, where the patients are seen as ultimate authority on the issue. The researchers tend to accept at face value the patients' stories, and promote respondents' accounts as the only true assessment of the events. However, respondents' accounts may have other motives beyond simply explaining coping with illness. Uncovering these motives, such as the social situation of the teller, or broader cultural explanations for a particular type of story, can help researchers extend the scope of analysis and produce a better assessment of social problems going beyond interpretations of participants.

Learning Objectives

In this chapter, you should learn to do the following:

- Appreciate and describe the main differences between the qualitative and quantitative analyses, which derive from different nature of the data, different goals of qualitative research, and different logic in the process of qualitative inquiry overall.
- Differentiate between the two general strategies in qualitative analysis: *analytic induction* and *grounded theory*.
- Understand the notions of *theoretical sampling* and *theoretical saturation* related to grounded theory.
- List the criticisms of grounded theory, including difficulties of transcription, data coding, and categorization; difficulty of rejecting the prior theoretical knowledge in data analysis; breaking data into small chunks that can prevent meaningful interpretation; and difficulties in generating theory in proper explanatory sense.
- Be able to apply the main principles of coding in an analysis of qualitative interview: basic coding, deeper awareness of content, and exploring basic analytical themes.
- Know the main criticisms of coding qualitative data including the fragmentation of data and loss of contextual understanding, and be able to describe the narrative analysis as a collection of methods allowing researchers to analyze people's stories and the way they understand the world around them.

Media Resources

Grounded Theory Institute
http://www.groundedtheory.com/

- What is the difference between different methods that claim to be grounded theory?
- What is the impact of learning and using jargon in grounded theory research?
- What is the relationship of data to Grounded Theory?

Gibbs, G.R., Grounded Theory: Core Elements.
http://www.youtube.com/watch?v=4SZDTp3_New
Gibbs, G.R., Grounded Theory: Open Coding Part 2
http://www.youtube.com/watch?v=vi5B7Zo0_OE&feature=related

- How does the use of grounded theory fit into the academic world?
- How well does coding work with grounded theory?
- Will coding interfere with the true notion of grounded theory?
- How many interviews would be acceptable for a grounded theory study?

Gupta, P. Grounded Theory Analysis of *A Beautiful Mind* (Prologue)
http://www.youtube.com/watch?v=13fAjb-y54g&feature=related

- What is the value of coding in this analysis?
- Could this text be analyzed using a different research method?
- Would the results be similar?

Songer, T. Narrative Analysis in Injury Research. University of Pittsburgh.
http://www.pitt.edu/~epi2670/narrative/narrative.pdf

- How does narrative data analysis work in injury research?
- Can the data be coded for computerized analysis?
- Who sets the categories for narrative analysis and by what criteria?

QSR International. What Is NVivo?
http://www.qsrinternational.com/what-is-qualitative-research.aspx

- When would a computer program assist with qualitative research?
- What are the limits of assistance that a computer program can provide to a researcher?
- Is the assistance of a computer program equally useful to academic research and market research?