

## Chapter 13

# Qualitative Analysis

Qualitative analysis is the analysis of qualitative data such as text data from interview transcripts. Unlike quantitative analysis, which is statistics driven and largely independent of the researcher, qualitative analysis is heavily dependent on the researcher's analytic and integrative skills and personal knowledge of the social context where the data is collected. The emphasis in qualitative analysis is "sense making" or understanding a phenomenon, rather than predicting or explaining. A creative and investigative mindset is needed for qualitative analysis, based on an ethically enlightened and participant-in-context attitude, and a set of analytic strategies. This chapter provides a brief overview of some of these qualitative analysis strategies. Interested readers are referred to more authoritative and detailed references such as Miles and Huberman's (1984)<sup>17</sup> seminal book on this topic.

### Grounded Theory

How can you analyze a vast set qualitative data acquired through participant observation, in-depth interviews, focus groups, narratives of audio/video recordings, or secondary documents? One of these techniques for analyzing text data is **grounded theory** – an inductive technique of interpreting recorded data about a social phenomenon to build theories about that phenomenon. The technique was developed by Glaser and Strauss (1967)<sup>18</sup> in their *method of constant comparative analysis* of grounded theory research, and further refined by Strauss and Corbin (1990)<sup>19</sup> to further illustrate specific coding techniques – a process of classifying and categorizing text data segments into a set of codes (concepts), categories (constructs), and relationships. The interpretations are "grounded in" (or based on) observed empirical data, hence the name. To ensure that the theory is based solely on observed evidence, the grounded theory approach requires that researchers suspend any preexisting theoretical expectations or biases before data analysis, and let the data dictate the formulation of the theory.

Strauss and Corbin (1998) describe three coding techniques for analyzing text data: open, axial, and selective. **Open coding** is a process aimed at identifying **concepts** or key ideas

---

<sup>17</sup> Miles M. B., Huberman A. M. (1984). *Qualitative Data Analysis: A Sourcebook of New Methods*. Newbury Park, CA: Sage Publications.

<sup>18</sup> Glaser, B. and Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Chicago: Aldine.

<sup>19</sup> Strauss, A. and Corbin, J. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, Beverly Hills, CA: Sage Publications.

that are hidden within textual data, which are potentially related to the phenomenon of interest. The researcher examines the raw textual data line by line to identify discrete events, incidents, ideas, actions, perceptions, and interactions of relevance that are coded as concepts (hence called *in vivo codes*). Each concept is linked to specific portions of the text (coding unit) for later validation. Some concepts may be simple, clear, and unambiguous while others may be complex, ambiguous, and viewed differently by different participants. The coding unit may vary with the concepts being extracted. Simple concepts such as “organizational size” may include just a few words of text, while complex ones such as “organizational mission” may span several pages. Concepts can be named using the researcher’s own naming convention or standardized labels taken from the research literature. Once a basic set of concepts are identified, these concepts can then be used to code the remainder of the data, while simultaneously looking for new concepts and refining old concepts. While coding, it is important to identify the recognizable characteristics of each concept, such as its size, color, or level (e.g., high or low), so that similar concepts can be grouped together later. This coding technique is called “open” because the researcher is open to and actively seeking new concepts relevant to the phenomenon of interest.

Next, similar concepts are grouped into higher order **categories**. While concepts may be context-specific, categories tend to be broad and generalizable, and ultimately evolve into *constructs* in a grounded theory. Categories are needed to reduce the amount of concepts the researcher must work with and to build a “big picture” of the issues salient to understanding a social phenomenon. Categorization can be done in phases, by combining concepts into subcategories, and then subcategories into higher order categories. Constructs from the existing literature can be used to name these categories, particularly if the goal of the research is to extend current theories. However, caution must be taken while using existing constructs, as such constructs may bring with them commonly held beliefs and biases. For each category, its characteristics (or properties) and dimensions of each characteristic should be identified. The *dimension* represents a value of a characteristic along a continuum. For example, a “communication media” category may have a characteristic called “speed”, which can be dimensionalized as fast, medium, or slow. Such categorization helps differentiate between different kinds of communication media and enables researchers identify *patterns* in the data, such as which communication media is used for which types of tasks.

The second phase of grounded theory is **axial coding**, where the categories and subcategories are assembled into causal relationships or hypotheses that can tentatively explain the phenomenon of interest. Although distinct from open coding, axial coding can be performed simultaneously with open coding. The relationships between categories may be clearly evident in the data or may be more subtle and implicit. In the latter instance, researchers may use a coding scheme (often called a “coding paradigm”, but different from the paradigms discussed in Chapter 3) to understand which categories represent *conditions* (the circumstances in which the phenomenon is embedded), *actions/interactions* (the responses of individuals to events under these conditions), and *consequences* (the outcomes of actions/interactions). As conditions, actions/interactions, and consequences are identified, theoretical propositions start to emerge, and researchers can start explaining why a phenomenon occurs, under what conditions, and with what consequences.

The third and final phase of grounded theory is **selective coding**, which involves identifying a central category or a core variable and systematically and logically relating this central category to other categories. The central category can evolve from existing categories or can be a higher order category that subsumes previously coded categories. New data is

selectively sampled to validate the central category and its relationships to other categories (i.e., the tentative theory). Selective coding limits the range of analysis, and makes it move fast. At the same time, the coder must watch out for other categories that may emerge from the new data that may be related to the phenomenon of interest (open coding), which may lead to further refinement of the initial theory. Hence, open, axial, and selective coding may proceed simultaneously. Coding of new data and theory refinement continues until **theoretical saturation** is reached, i.e., when additional data does not yield any marginal change in the core categories or the relationships.

The “constant comparison” process implies continuous rearrangement, aggregation, and refinement of categories, relationships, and interpretations based on increasing depth of understanding, and an iterative interplay of four stages of activities: (1) comparing incidents/texts assigned to each category (to validate the category), (2) integrating categories and their properties, (3) delimiting the theory (focusing on the core concepts and ignoring less relevant concepts), and (4) writing theory (using techniques like memoing, storylining, and diagramming that are discussed in the next chapter). Having a central category does not necessarily mean that all other categories can be integrated nicely around it. In order to identify key categories that are conditions, action/interactions, and consequences of the core category, Strauss and Corbin (1990) recommend several integration techniques, such as storylining, memoing, or concept mapping. In **storylining**, categories and relationships are used to explicate and/or refine a story of the observed phenomenon. Memos are theorized write-ups of ideas about substantive concepts and their theoretically coded relationships as they evolve during ground theory analysis, and are important tools to keep track of and refine ideas that develop during the analysis. **Memoing** is the process of using these memos to discover patterns and relationships between categories using two-by-two tables, diagrams, or figures, or other illustrative displays. **Concept mapping** is a graphical representation of concepts and relationships between those concepts (e.g., using boxes and arrows). The major concepts are typically laid out on one or more sheets of paper, blackboards, or using graphical software programs, linked to each other using arrows, and readjusted to best fit the observed data.

After a grounded theory is generated, it must be refined for internal consistency and logic. Researchers must ensure that the central construct has the stated characteristics and dimensions, and if not, the data analysis may be repeated. Researcher must then ensure that the characteristics and dimensions of all categories show variation. For example, if behavior frequency is one such category, then the data must provide evidence of both frequent performers and infrequent performers of the focal behavior. Finally, the theory must be validated by comparing it with raw data. If the theory contradicts with observed evidence, the coding process may be repeated to reconcile such contradictions or unexplained variations.

## Content Analysis

Content analysis is the systematic analysis of the content of a text (e.g., who says what, to whom, why, and to what extent and with what effect) in a quantitative or qualitative manner. Content analysis typically conducted as follows. First, when there are many texts to analyze (e.g., newspaper stories, financial reports, blog postings, online reviews, etc.), the researcher begins by sampling a selected set of texts from the population of texts for analysis. This process is not random, but instead, texts that have more pertinent content should be chosen selectively. Second, the researcher identifies and applies rules to divide each text into segments or “chunks” that can be treated as separate units of analysis. This process is called *unitizing*. For example,

assumptions, effects, enablers, and barriers in texts may constitute such units. Third, the researcher constructs and applies one or more concepts to each unitized text segment in a process called *coding*. For coding purposes, a coding scheme is used based on the themes the researcher is searching for or uncovers as she classifies the text. Finally, the coded data is analyzed, often both quantitatively and qualitatively, to determine which themes occur most frequently, in what contexts, and how they are related to each other.

A simple type of content analysis is **sentiment analysis** – a technique used to capture people’s opinion or attitude toward an object, person, or phenomenon. Reading online messages about a political candidate posted on an online forum and classifying each message as positive, negative, or neutral is an example of such an analysis. In this case, each message represents one unit of analysis. This analysis will help identify whether the sample as a whole is positively or negatively disposed or neutral towards that candidate. Examining the content of online reviews in a similar manner is another example. Though this analysis can be done manually, for very large data sets (millions of text records), natural language processing and text analytics based software programs are available to automate the coding process, and maintain a record of how people sentiments fluctuate with time.

A frequent criticism of content analysis is that it lacks a set of systematic procedures that would allow the analysis to be replicated by other researchers. Schilling (2006)<sup>20</sup> addressed this criticism by organizing different content analytic procedures into a spiral model. This model consists of five levels or phases in interpreting text: (1) convert recorded tapes into raw text data or transcripts for content analysis, (2) convert raw data into condensed protocols, (3) convert condensed protocols into a preliminary category system, (4) use the preliminary category system to generate coded protocols, and (5) analyze coded protocols to generate interpretations about the phenomenon of interest.

Content analysis has several limitations. First, the coding process is restricted to the information available in text form. For instance, if a researcher is interested in studying people’s views on capital punishment, but no such archive of text documents is available, then the analysis cannot be done. Second, sampling must be done carefully to avoid sampling bias. For instance, if your population is the published research literature on a given topic, then you have systematically omitted unpublished research or the most recent work that is yet to be published.

## **Hermeneutic Analysis**

Hermeneutic analysis is a special type of content analysis where the researcher tries to “interpret” the subjective meaning of a given text within its socio-historic context. Unlike grounded theory or content analysis, which ignores the context and meaning of text documents during the coding process, hermeneutic analysis is a truly interpretive technique for analyzing qualitative data. This method assumes that written texts narrate an author’s experience within a socio-historic context, and should be interpreted as such within that context. Therefore, the researcher continually iterates between singular interpretation of the text (the part) and a holistic understanding of the context (the whole) to develop a fuller understanding of the phenomenon in its situated context, which German philosopher Martin Heidegger called the

---

<sup>20</sup> Schilling, J. (2006). “On the Pragmatics of Qualitative Assessment: Designing the Process for Content Analysis,” *European Journal of Psychological Assessment* (22:1), 28-37.

*hermeneutic circle*. The word *hermeneutic* (singular) refers to one particular method or strand of interpretation.

More generally, hermeneutics is the study of interpretation and the theory and practice of interpretation. Derived from religious studies and linguistics, traditional hermeneutics, such as *biblical hermeneutics*, refers to the interpretation of written texts, especially in the areas of literature, religion and law (such as the Bible). In the 20th century, Heidegger suggested that a more direct, non-mediated, and authentic way of understanding social reality is to experience it, rather than simply observe it, and proposed *philosophical hermeneutics*, where the focus shifted from interpretation to existential understanding. Heidegger argued that texts are the means by which readers can not only read about an author's experience, but also relive the author's experiences. Contemporary or modern hermeneutics, developed by Heidegger's students such as Hans-Georg Gadamer, further examined the limits of written texts for communicating social experiences, and went on to propose a framework of the interpretive process, encompassing all forms of communication, including written, verbal, and non-verbal, and exploring issues that restrict the communicative ability of written texts, such as presuppositions, language structures (e.g., grammar, syntax, etc.), and semiotics (the study of written signs such as symbolism, metaphor, analogy, and sarcasm). The term hermeneutics is sometimes used interchangeably and inaccurately with *exegesis*, which refers to the interpretation or critical explanation of written text only and especially religious texts.

## Conclusions

Finally, standard software programs, such as ATLAS.ti.5, NVivo, and QDA Miner, can be used to automate coding processes in qualitative research methods. These programs can quickly and efficiently organize, search, sort, and process large volumes of text data using user-defined rules. To guide such automated analysis, a coding schema should be created, specifying the keywords or codes to search for in the text, based on an initial manual examination of sample text data. The schema can be organized in a hierarchical manner to organize codes into higher-order codes or constructs. The coding schema should be validated using a different sample of texts for accuracy and adequacy. However, if the coding schema is biased or incorrect, the resulting analysis of the entire population of text may be flawed and non-interpretible. However, software programs cannot decipher the meaning behind the certain words or phrases or the context within which these words or phrases are used (such as those in sarcasms or metaphors), which may lead to significant misinterpretation in large scale qualitative analysis.