CHAPTER 21

Qualitative data analysis: an overview

Chapter contents

Learning objectives

When you have read this chapter, you should understand:

1 the wide range of methods to analyse qualitative data

2 the principles and process of coding.

21.1 Qualitative data analysis

In the following section we will discuss a wide range of techniques for qualitative data analysis. While in quantitative research the distinction between methods and analysis is rather clear, in qualitative studies that difference is vaguer. Is content analysis (discussed in Chapter 10) a method or an analysis technique? In this book we distinguish between methods and analysis by arguing that methods are used to obtain information, while analysis is used to summarize, (re)structure and relate information. Following this distinction criteria, we consider the approaches discussed in Chapter 10 as methods and distinct from analysis. The coding described in section 10.5 on grounded theory marks the blurred boundary between methods and analysis.

21.2 Coding

Words are typically at the base of qualitative analysis and words are less easy to understand than numbers, because words can have multiple meanings – they are context dependent, i.e. if you read the word 'lion' which association is evoked? A dangerous animal, king of the savannah, endangered animal or do you think of the coat of arms of the England football team? Despite those difficulties while working with words, what you will get is a 'thick' description[1] and a much deeper understanding than numbers can offer.

Transcribing

Before you can turn to the actual coding, researchers often need to transcribe an oral account, for example of a recorded interview, into a written account. **Transcribing** requires a couple of decisions to be made, as the written mode does not naturally capture what and how something has been said. The first decision when you start transcribing is whether you want to transcribe the entire interview or transcribe the main parts and just summarize the apparently irrelevant parts. You should note that transcription of a one-hour interview takes at least five hours if you are an experienced typist. Thus, especially if you have a vast amount of interview hours, selective transcription might be the only option to manage the task in a reasonable time. However, if you decide for selective transcription you need to be clear on the criteria upon which you decide which interview parts are transcribed word by word and which are summarized. The second decision is whether the transcription is verbatim, that is, word by word including all kind of repetitions, or in a more formal writing style. Exhibit 21.1 provides an example of the two to illustrate the differences. Finally, you need to decide whether you only transcribe the words spoken or whether you include pauses and emotional expressions. This choice puts the researcher in a dilemma, as not considering the non-verbatim parts might change the content of the transcript, but including them introduces subjective judgements. Consider the following sentence spoken by a TV reporter: 'In the last minute Ajax Amsterdam equalized the score board against Celtic Glasgow and moves to the quarter finals.' Now imagine this sentence is spoken by a Scottish reporter, by a Dutch reporter or by a Spanish reporter. Verbatim all three would say the same, but would they all sound the same and mean the same?

***Exhibit 21.1** Verbatim and formal transcription.*

Verbatim	Formal
The next phase of it was the development of the carbon fibre. This period was between 1985 and 2000. The process was supported by EU funding programs. These programs had literally the goal to encourage technologies co-operation between companies and universities. We co-operated with a lot of companies, clients, suppliers but even with chemical companies, even competitors. And this was really the first time that we not only talked about technology issues but talked about developing the market.	The next phase was the carbon fibre development between 1985 and 2000. The development was supported by EU funding programs encouraging co-operation between business firms and universities. We co-operated with clients, suppliers and even competitors, not just about technological issues but about developing the market.

Although transcribing seems to be a simple process, we identify a couple of issues that will affect the reliability and validity of transcription. The more effort you put in transcription the better the quality will be. Thus, if you transcribe not only the words but also emotional expressions, if two persons independently transcribe the same material, and if the recordings are made with good technical equipment to achieve high acoustic quality, then you will achieve high quality. But the important question to be answered is always: What is a useful transcription for my research purpose? Thus, efforts need to be in balance with the needed output.

Coding systems

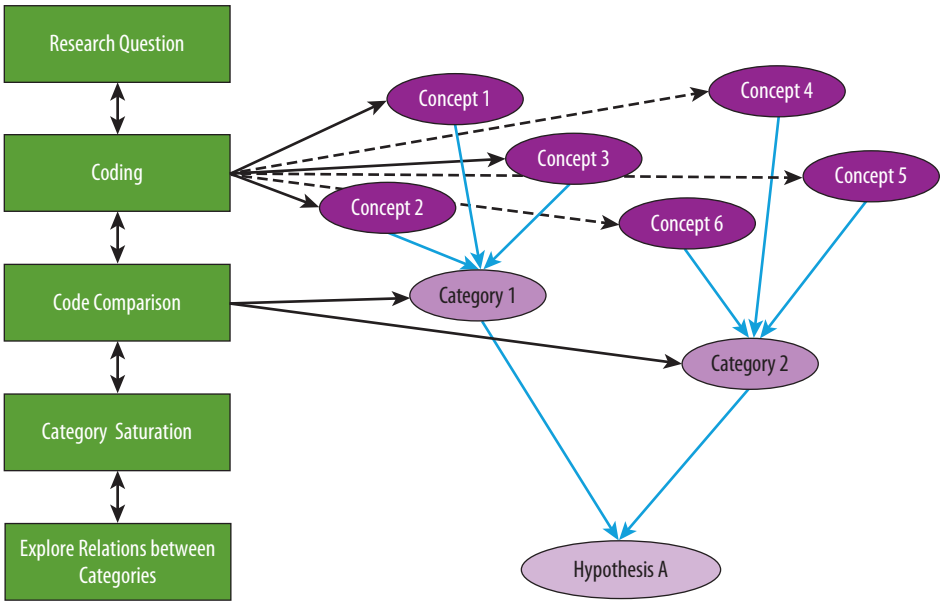
Coding is an interpretive technique with the objective to reduce and structure the information available from sources, such as extensive field notes, transcripts of interviews or any other kind of relevant written document. Coding describes the labelling of information pieces in our material. The labels should reflect in a single word or phrase what that piece of information is about.

Analogous to the general differences in qualitative research designs regarding the degree of structuring, the same differences can be made for the coding procedure. Thus, one of the first questions one needs to ask is whether a list of possible codes before the inspection of the material is produced or should such a list with codes emerge from the material. Whether one should use **prescriptive coding** or **emerging coding** depends on what is already known about the phenomena, the supposed analysis and also pragmatic criteria, such as available time. Especially if the amount of material is large, emerging coding is very time consuming, as it requires an iterated coding process. Every time a new code emerges, material that already had been scanned needs to be re-scanned for the new code. Thus, at least all material has to be scanned twice: the first time to reveal the codes and the second time to mirror the whole material against the completed list of codes. The main advantage of emerging coding is obviously that a more open and unstructured process is more likely to produce new insights and knowledge and is better able to capture idiosyncratic aspects that might be missed if prescriptive coding is applied.

Coding process

Exhibit 21.2 shows the **coding process**. It starts with the research question, which determines what we are looking for in the material. The more we already know about a phenomenon, the more specific the research question will be and the less we know, the more vaguely it will be formulated. The codes we develop, either prescriptive or

Exhibit 21.2 The coding process.



emerging, label pieces of information in our field material and should reflect the content of the field material. Those labels can be seen as concepts as they are closely linked to the perceived reality laid down in the material. In this first phase of the actual coding process, researchers should use too many rather than too few codes. Reducing yourself to a few codes goes along with a loss of information and in the end you will fail to fully utilize the richness of your material. Thus, in case you are in any doubt whether to use a new code or an already-used one, use a new code.

The following step is a crucial one, as you are now asked to compare the coded information fragments. First, you need to check whether fragments that have received the same code express the same phenomena and meaning or whether splitting the fragments under different codes would be a better reflection of the material. The objective of this first step in revising your codes is mainly to ensure consistency in your coding – that it enlarges the validity of your coding. The main objective of the code comparison is, however, to reduce the amount of information. Why do we need to reduce the amount of information? Scientific research always needs a theoretical component, either in the form of theory development or in the form of theory testing. Theories should be parsimonious, that is, they should contain no more elements than necessary to explain a phenomena well. Consequently one attempts to generalize a theory by summarizing idiosyncratic concepts into more general categories. Those categories should align with the constructs that the theory is dealing with. This process of category formation is iterative and will involve going backwards to the concepts used in the coding procedure and try out how well different concepts fit together. The natural end of this category formation process is category saturation, that is, no further concept can be included without a major change in the meaning of the categories.

The final step in the coding process marks the line between pure coding and further interpretation and theory development. Once all categories are formed you need to explore how those categories are related to each other. This step requires the field material to be scanned for possible relations between categories. From a scientific perspective, this final step is most productive and rewarding: it is the contribution of a study in terms of theory development by generating new propositions or hypotheses; and, in terms of theory testing, it is when the qualitative analysis documents the relationship between categories that were previously hypothesized.

A central question that remains is: How to code? Are there some principles a researcher can use to attach a code – a word or short phrase – to a text fragment or an information chunk? Codes can be developed by holding a text fragment against the light of the following questions:

- What does the text fragment describe – an act or activity, an object, a person, a meaning, a relationship or a setting?
- To what question would the text fragment be the answer?
- Does the text fragment reflect a person's own experience or perception or does it inform us indirectly about a third person's experience or perception?
- Which more-general categories or topics would describe the text fragment?
- To which level does the text fragment refer – microscopic, individual, relational, group, community or system level?

Coding is a dynamic and iterative process. Even in the case of prescriptive coding, one will observe that certain codes become popular, while others are (almost) not used. These differences in code usage start a process of reformulating codes. Codes that are used very often need to be checked for consistent usage and one needs to raise the question whether sub-codes would not provide a better reflection of the material. Likewise, hardly used codes raise the question whether they should be merged with other codes and what one would lose through merging codes. These considerations regarding the splitting and merging of codes mainly address the issue of how many codes are useful for the research's purpose, but a much more important issue is not addressed.

The ultimate goal of any scientific study is that we want to understand the world better either by developing new theories or by testing them. Theories provide explanations by relating specific concepts. For example the theoretical considerations on the ambidexterity of firms suggest that firms finding a balance in exploring new opportunities and exploiting existing ones will outperform other firms. Thus, the theory suggests that the two concepts of exploration and exploitation need to be balanced and this balance explains firms' performance. For a scientific advancement, it is therefore not sufficient to just assign codes; one also needs to place the codes within a certain structure. In our example, the structure between the codes' exploration and exploitation would need to indicate

balance. This structure can reflect hierarchical relations between codes, a sequential process along which codes or code clusters of common themes are placed. Below you find a list of recommendations to avoid the coding trap.[2]

- Use software packages, such as NVivo (see section 21.4) for coding, because this ensures that once you have generated a code, other occurrences in the source text are not overlooked.
- Watch the balance between labelling fragments and identification of themes. (Ask 'Why am I interested in this?' rather than 'Where does this go?')
- In the beginning, be generous with coding categories.
- Monitor coding categories and their content periodically.
- Keep generating ideas and questions.
- Stop when you are tired.

Coding concerns

A common concern regarding coding is the researcher's subjectivity in assigning codes. What can be done to reduce this concern? First, during the coding procedure the researcher needs to keep a logbook to note down not just the codes created but also the rules used to decide how to assign a specific code to a text fragment. Although these rules will bear an element of subjectivity, at least it becomes transparent to other people which criteria and rules were used, and then it is up to them to assess to what extent the researcher's subjectivity biased the coding.

Next to a thoroughly kept coding logbook, double coding is another alternative. Double coding requires that the study material is coded independently by other researchers as well. Even just one additional person coding the study material well will increase the codes validity and reliability. Following the proverb 'two pairs of eyes are better than one', agreement between the coders suggests reliability, while disagreement between coders should stimulate discussion. Resulting from this are more appropriate codes or even code structures, which improve the code's validity. You should note, however, that the quality of coding depends crucially on the quality of the coders. Although background knowledge may bias coders, such knowledge is of the utmost importance to see structures – to still see the wood despite the trees.

Another frequent concern of coding methods is that it seeks to transform qualitative data into 'quasi-quantitative' data, thereby draining the data of its variety, richness and individual character. The worry is that the big advantage of qualitative research, namely the possibility to account for context, is lost in the coding procedure. This emphasizes again that coding is not just assigning labels to text fragments, but that it requires also the identification of code structures that are potentially context driven. Moreover, researchers need to expose the rules and definitions underlying the codes, as they incorporate more of the study material richness than the simple code list.

21.3 Analysis tools

Scientific analysis – either quantitative or qualitative – is much about data reduction. In Chapters 16 and 17 we discussed descriptive statistics which we use to summarize data. The comparison of the mean prices for a 60 m^2 flat in the Paris Latin Quarter in 2003 and 2013 tells us more than a long list of selling prices in those two years for the Latin Quarter. Of course, summaries always lead to a loss of information; therefore, we need to remain careful how far we summarize, but without summarizing we will lose the ability to give meaning to our data. Quantitative analysis uses numbers to summarize the information; in qualitative analysis, displays can serve this function and facilitate interpreting the data.

Data analysis in qualitative research

In the following, we briefly discuss a couple of qualitative analysis tools that support you in structuring your codes to identify groups of codes, relation between codes and code sequences. Thus, what we discuss in this section is how to continue once you have coded your source material, i.e. how one can identify structures and clusters between the codes.

Typology or taxonomy

Both refer to the creation of a classification system that groups mutually exclusive categories. In organizational science, a well-known **taxonomy** is the organizational configuration framework by Mintzberg,[3] who distinguished five – later he added a sixth – types of organizational designs:

1 simple structure characteristic of entrepreneurial organization
2 machine bureaucracy
3 professional bureaucracy
4 diversified form
5 adhocracy or innovative organization
6 missionary organization.

Developing such **typologies** requires that one identifies dimensions on which the different types differ. In Mintzberg's typology, important differences between the six types are based on the primary coordination mechanism used and the primary part of the organization.

Induction

In Chapter 5, we already discussed **induction** and **deduction** as two thought processes. Qualitative method and analyses have an emphasis on induction. Induction moves from the specific to the general and is at the core of theory generation and development. The inductive process follows six steps:[4]

1 A phenomenon is defined in a tentative manner.
2 A hypothesis is developed explaining the phenomenon.
3 A single instance is considered to determine whether the hypothesis holds.
4 If the hypothesis does not hold either the phenomenon or the hypothesis is revised to include the instance.
5 Additional instances are investigated to check whether the hypothesis holds for them as well. If the hypothesis holds repeatedly, the certainty that the hypothesis provides a useful explanation increases.
6 If we find instances in which the hypotheses does not hold, the hypotheses needs to be reformulated to cover those exceptions.

These six steps describe an iterative process. Assume your research interest is on massive lay-offs. (1) In the first step, you would need to define the phenomenon more precisely and, being a business student, you are specifically interested what happens within an organization after a massive lay-off occurred. What does it mean for the workers who 'survive' the lay-off? (2) One could develop conflicting hypotheses about this. On one hand, the survivors could be happier and more committed because they survived the lay-off, but on the other hand, the reverse could occur because they feel sorry for their former colleagues who lost their jobs or because they are still afraid they may lose their job in the future. (3) Now you hold interviews with employees of an organization that just had such a massive lay-off. A qualitative analysis of the interviews suggests the remaining workers are less satisfied and committed, suggesting that the negative effect of massive lay-offs reaches the remaining employees. However (4), some interviews suggest that some employees feel more committed now. From the material you have you would now try to understand why some employees are more committed while others are less committed. One explanation you find might be that those who feel more committed were much better informed by the management about the reasons for the lay-off and the future perspective of the firm. This would lead to a revision of the hypothesis suggesting that the effect of massive lay-offs is moderated by the communication intensity. In step 5, you actually move again to step 3 by finding a second firm that had a massive lay-off and then you investigate whether the revised hypothesis holds.

Domain analysis

A domain can be defined as a symbolic category that includes other categories. Thus, several categories that share certain relationships form a domain. **Domain analysis** emphasizes the importance of relationships to form a domain. In Exhibit 21.3, different types of relationships are explained; screening a source text with these different types of relation in mind will help you to identify relevant domains.[5]

Exhibit 21.3 ***Relationship types in domain analysis.***

Relationship Type	Description	Example (X printed in *italics)*
Strict inclusion	X is a kind of Y	*Maastricht University* is an institution of higher education.
Spatial inclusion	X is a place in Y, X is part of Y	The *School of Business and Economics* is part of Maastricht University.
Cause effect	X is a cause of Y	The *foundation of Maastricht University* increased the demand for highly qualified personnel in the region.
Rationale	X is a reason for doing Y	*Stimulating economic growth* in the south of the Netherlands was one reason to found Maastricht University in 1976.
Location for action	X is a place for doing Y	*Maastricht University* is widely known for the excellent cardiovascular research conducted there.
Function	X is used for Y	At Maastricht University we use *problem-based learning* to teach our students.
Means ends	X is a way to do Y	*'Sharing Success', the explicit vision of the School of Business* creates a common culture among staff members and students.
Sequence	X is a step or stage in Y	*Studying at Maastricht University* is a major step in a successful career.
Attribution	X is an attribute or characteristic of Y	*Internationally diverse classrooms* distinguish Maastricht University from other Dutch universities.

To start a domain analysis, one first looks for semantic relationships, just as strict inclusion or means ends. Then you take your data material, read through it, note down all word pairs that built a specific relationship and, finally, you challenge the relationships found. Domain analysis has been extremely useful in investigations of cultural context and social settings, as it allows the researcher to see how consistent different categories and concepts fit with a culture or social setting understood as a domain.

Domain analysis can be used to identify structures to phenomena that appear complex and fragmented at first sight. Scholars often use domain analysis to understand developments within a specific scientific field. For example Jones, Coviello and Tang[6] used domain analysis to structure the field of international entrepreneurship research over a period of 20 years and reveal a sub-structure that distinguishes between studies making international comparisons, studies that investigate international entrepreneurship and studies that do both. Durugbo[7] uses domain analysis to understand collaboration between organizations involved in the design of microsystem technology. His source texts are based on semi-structured interviews and he uses their transcriptions to identify three different functional purposes of design teams, which are related to four strategies enhancing collaboration.

Hermeneutical analysis

Earlier, we stated the importance of context in qualitative research, which is often neglected in quantitative approaches. The opportunity to consider the context of a phenomenon is a major strength of qualitative analysis and there are a couple of approaches that suggest how one can consider context.

Hermeneutical analysis provides important insights into how one can interpret phenomena. Hermeneutics is the theory of interpretation and addresses, in a systematic manner, the challenges one faces in interpretations. One of the problems we face is that the meaning of words differs depending on the sender. The dilemma researchers face is that they are supposed to interpret a specific text, but they are bound by their own experience and background, which often is very different from the sender's background. Think again about the interviews with the employees that survived the massive lay-off. The interviewees might include unskilled workers, foremen, as well as middle managers. As a university graduate, it will be difficult for you to understand the fine nuances that people with a very different background have conveyed. Thus, interpretation of text requires a good understanding of the sender.

Next to that, the historical, social and cultural context of a text needs to be acknowledged. Did the massive lay-off you investigated occur in times of a general economic crisis with high unemployment, or did it occur in otherwise rather prosperous times? What type of company did you investigate? Was it the last lay-off in a series of previous lay-offs or was it the first one in the firm's history? Were lay-offs typical for the sector?

Finally, text needs to be understood from a holistic perspective. Before, we mentioned that coding procedures bear the risk that one focuses too much on text fragments and loses the view of the whole. Hermeneutical analysis calls for a constant shifting between the specific and the whole by reflecting on the importance of the specific to the whole and trying to put each specific in its place in the whole. To summarize, hermeneutical analysis offers us the opportunity to combine different types of context with the interpretation of the phenomena.

Phenomenology and heuristic research

Phenomenological research puts the experience central in the analysis. The starting point is descriptions of experiences that are obtained through unstructured interviews and dialogues. The researcher uses this original material to conduct a reflective structural analysis with the objective to understand what the experience means for those who had the experience. Finally, these interpretations of experiences are used to derive general meanings.

Heuristic research does not start with an experience but rather with a personal question that nevertheless needs to have a wider social importance. The answer to this question is searched by self-enquiry and dialogue. Researcher and participant try to develop jointly a story answering the question and reflecting their personal experiences and thoughts. Taking such stories from multiple participants, the researcher attempts to develop a general meaning by synthesis.

Heuristic research and **phenomenology** share several qualities:

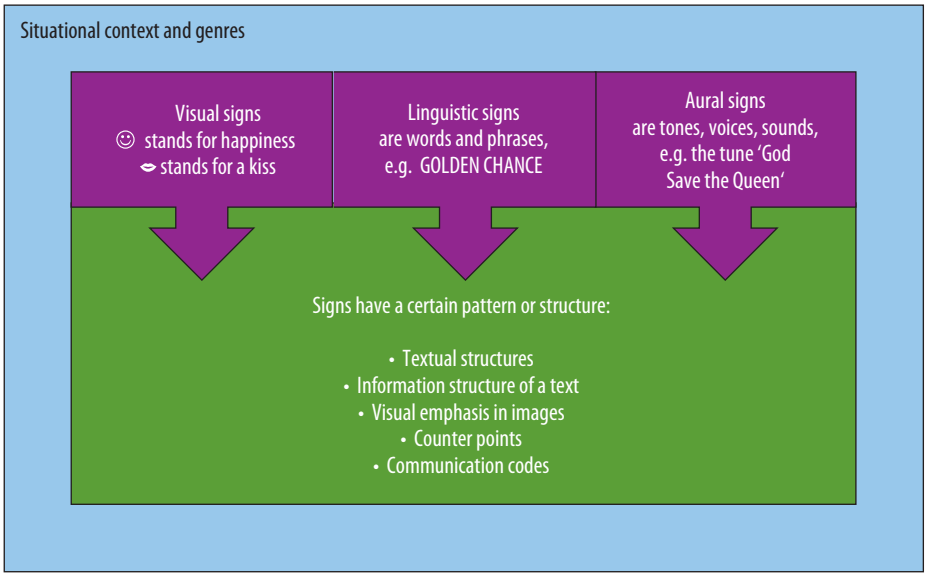
- They have strong preference for qualitative designs, as quantitative approaches are not appropriate to study experiences.
- They have a holistic nature as both emphasize the whole picture and pay less attention to single objects or fragments.
- They are interested in meaning and the essence of an experience and not in measurements and explanations, as they view experience and behaviour as inseparable.
- Their data collection is primary and mostly based on unstructured interviews.

Semiotics

Umberto Eco introduced semiotics to a wider audience by writing a book entitled *The Theory of Semiotics* and by his historic novel *The Name of the Rose*. He said: 'Semiotics is concerned with everything that can be taken as sign.'[8] Like the previous analytical approach, semiotics is used to identify and design the meaning of communication and also of texts by taking a sign-based perspective. When we communicate with people either orally or in writing, they have an at least implicit agreement what a word, image, gesture and so on mean; as soon as we agree on the meaning, it has become a sign.

To understand **semiotic analysis**, the toolbox in Exhibit 21.4 shows the different elements used. It starts with the identification of signs and, as shown, signs can have various forms, but a widely accepted meaning is attached to each sign. Visual signs can be simple like the two examples in Exhibit 21.4: the ☺ is associated with happiness and the 😚 is associated with a kiss, but they can be more complex. Think about famous photographs. The portrayal of Albert Einstein stretching out his tongue stands for the narrow connection between craziness and genius. Press photographs taken in war zones, such as the crying Vietnamese girl running down a field escaping the bombing, evoke in us thoughts about the cruelty of war and the suffering innocents have to endure. Especially in the analysis of text, linguistic signs are of importance. Words are, of course, always a kind of agreement on how to label a certain object or activity, but often words go beyond such labelling and have further meanings. For example, the word 'golden' is not just a description of a specific shiny colour. It has other meanings and we associate luxury or value with it. Finally, we have aural signs – tones and sounds. Hearing the tune of 'God Save the Queen', people will automatically associate it with England, the Royal Family and Monarchy.

Exhibit 21.4 ***The semiotic toolbox.***



As signs are based on an agreed meaning, the situational context in which they are embedded becomes important. Although some signs, such as the ☺ have probably a globally accepted meaning, other signs will be heavily influenced by culture and social settings. Consider how many countries have signed UN declarations calling for respect of human rights; but how different is the meaning of human rights to us, as the agreed balance between individual rights and societal rights differs between countries.

Once we have identified the signs, we need to investigate how systems of signs are structured. In text analysis, we can distinguish between textual structures and information structures. The first refers more to a linguistic analysis and addresses questions whether signs were used rhetorically or what the narrative structure of the text is. Here is the typical narrative structure of any Bond novel by Ian Fleming:[9]

- M moves and gives a task to Bond.
- The bad character appears to Bond.
- Bond gives a first check on the bad character or vice versa.
- A woman appears and Bond seduces her.
- The bad character captures Bond and tortures him.
- Bond beats the bad character.
- Bond has a romantic adventure with the woman, but it will not last.

Information structure refers more to the content of the text – which information is new and which is old, which information is presented and which information is neglected. The analysis of information structure does not need to be limited to text analysis, as images contain information as well. An interesting question is, for example, what is shown and what is not shown on an image. A historic example is that Leo Trotsky was removed from official photographs once he fell into disgrace with Lenin and Stalin. Analysing an image starts with the question 'where does the emphasis lie?' What is in the foreground, does the composition draw your eye to a certain place, are colours used to draw your attention to certain parts of the image?

Another approach is to look for counterpoints, which could include contrasts in images, but also refer to texts in which we can investigate which kind of distinctions and comparisons people make. Like the formation of an identity, giving meaning to something also involves demarking it from what it is not and showing how it is distinctive. Finally, the semiotic analysis looks for communication codes that describe unspoken rules. Those codes are, however, dependent on the situational context and also change over time.

Semiotic analysis is also used in business and management studies and seems to be especially suited to phenomena in which a company's culture plays an important role and where visuals are important, for example in advertising research. Bell[10] studied the closure of the Jaguar plant Browns Lane and how the announced closure formed collective memories. She analysed hundreds of photographs from the firm's archive to understand the organizational memory throughout the firm's history; she showed how important visuals and, of course, the visualization of the firm's name as a wild cat and prominent leaper were in the formation of the collective memory. While Bell focused on visuals, Fiol[11] used semiotics to analyse the letters to the stakeholders of ten chemical companies. Through semiotic analysis, she could identify underlying values that explained the companies' approaches in defining their organizational boundaries and their engagement in joint ventures.

Discourse analysis

Discourses are interrelated sets of texts reflecting an ongoing flow of communication between several individuals. But while in content analysis the emphasis is on the analysis of the text itself, in discourse analysis the emphasis shifts to the relational aspects of text originating from two or more sources. **Discourse analysis** is especially, but not exclusively, useful for research conducted in the tradition of (social) constructivist research philosophies. Rather than assuming a social world and then trying to understand its meaning for the research participants, discourse analysis attempts to uncover how the socially produced ideas and objects are created. Thus while other qualitative approaches interpret a social reality as it exists, discourse analysis explores how it is produced. Exhibit 21.5 shows different types of discourse analysis, which are classified on two dimensions. The first dimension reflects whether the analysis emphasis is on the context – the setting where the discourse takes place – or on the text and the interaction. The second dimension refers to the researcher's perspective, which can be critical, that is, investigating the relation between language and power, or constructionist, investigating the process how reality is created. The objective of critical discourse analysis is to investigate the role of discourses in the emergence and maintenance of asymmetric power relations, while critical linguistic analysis explores how power in a specific context is reflected in the text of the discourse. Interpretive structuralism emphasizes the social context and investigates what discourses are supported. Finally, social linguistic analysis focuses on how the discourse constructs phenomena.

Exhibit 21.5 Types of discourse analysis.

	Context	Text
Critical	Critical discourse analysis	Critical linguistic analysis
Constructivist	Interpretive structuralism	Social linguistic analysis

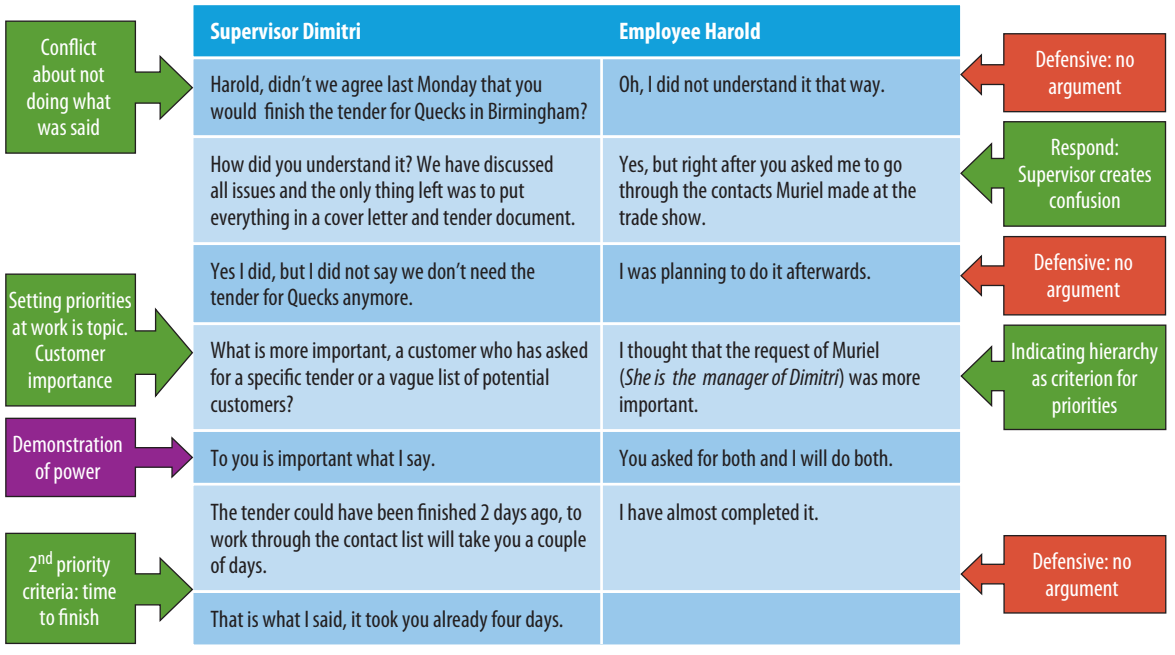
In discourse analysis, we can either collect natural occurring texts, for example, negotiation talks between the management of a company and the workers' council that were obtained by observations, or one can create text through qualitative interviews or focus groups. If your analysis is based on natural occurring texts, you need to select texts and the following questions will guide you through the selection:

- Which texts have been most important in the creation of the phenomena? What texts are often referred to? What are the texts that started the discourse? What are the texts that changed the discourse?
- Who have been the most powerful actors in the discourse and which texts did texts produce?
- Are the texts available and feasible for analysis?

The analysis of the discourse often starts with coding the discourse and exploring what categories the selected texts generated. During the actual analysis of the discourse, the researcher looks for rhetorical schemes (do actors rely on 'facts' in their argumentation, on logic and deduction or do they refer to emotions and feelings), how different actors interpret the categories, whether actors claim to have a stake in the phenomena, whether actors start to align or whether they polarize. Answers to those questions will sketch a picture on how the discourse and the context co-created a specific phenomenon.

Exhibit 21.6 pictures a short sequence of a discourse between a supervisor and an employee about a task conflict. Here we looked whether what the two said was supported by arguments or whether authoritative power was used. The analysis also reveals that the employee often responds with neither an argument nor an acceptance of the power.

Exhibit 21.6 ***Example of a discourse analysis.***



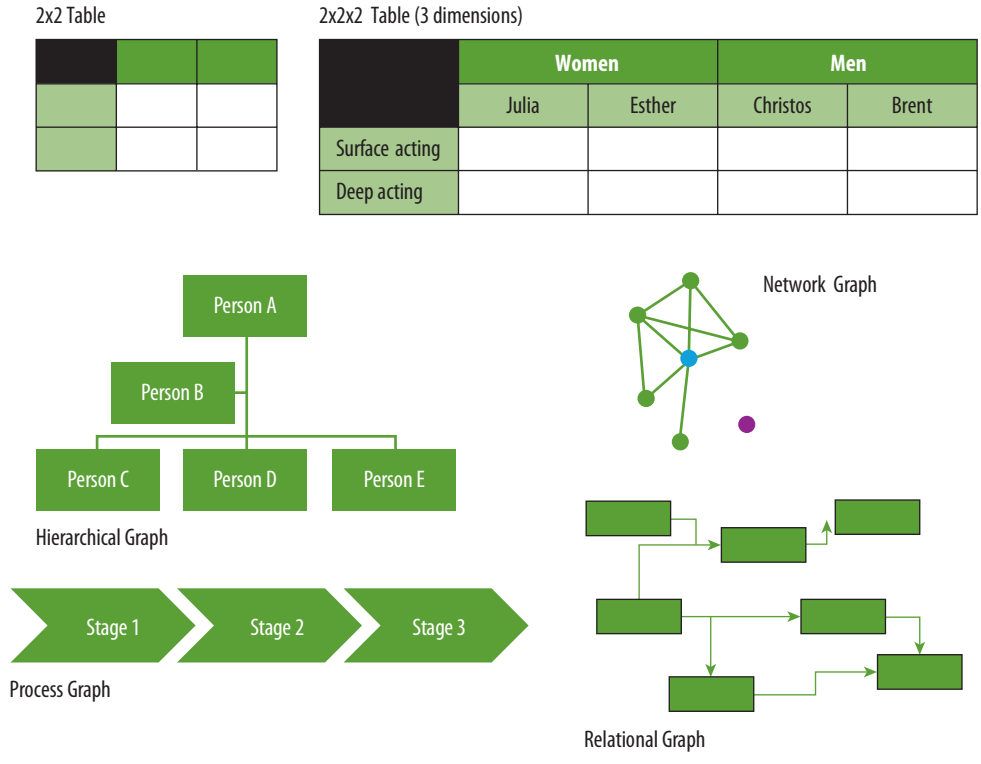
Matrix and logical analysis

Matrix and logical analysis employs graphical features to structure the information obtained and to depict relationships and processes. A major advantage of using graphical features is that it drastically condenses the information and at the same time visualizes the main patterns within the information collected. The drawback is, however, that in the attempts to condense and structure, some information might get lost while other pieces of information might be overemphasized. Thus, here the researcher needs again to ensure that the right balance is kept.

A matrix is usually displayed as a table, often with two dimensions, but three dimensions are also possible (see Exhibit 21.7). A matrix is usually the foundation of all graphical displays. In Microsoft Excel you can design a wide variety of graphical displays (called charts) but all are based on a spreadsheet that is a matrix. The important question to answer is how do you label the two dimensions of your matrices, that is, what do you put in the columns and in the rows? In the following, we provide examples of matrices that are often used in qualitative research:[12]

- The *checklist matrix* has properties or requirements in the row and column headings of observed cases, persons, situations or events and represents the actual unit of analysis you are interested in. The cells are then filled with a simple check mark if a certain property can be attached to the observation. The checklist then allows you to systemically analyse whether or to what extent what you have observed fits the range of criteria (properties or requirements).
- The *time order matrix* crosses time (points) with units of analyses. The cells would be filled with information chunks taken from the field notes or interviews that would refer to a specific unit of analysis, for example the person in the respective time.
- The *concept matrix* is used to align theoretical concepts with the data material – the unit of analysis intersects with the concepts that can be taken directly from the theory or that have emerged during the coding process. The cells of the table are filled with information fragments indicating that the information regarding, for example, the specific person could be described by the respective concept.
- The *effects matrix* intersects concepts that are theoretically related, in terms of theory and proposition development. The row headings could be filled with the explanatory or independent variables, while the column headings represent the dependent variables The cells would then be filled with information fragments from the qualitative data that indicate a relationship between the two concepts.

Exhibit 21.7 ***Forms of graphical representation.***



The most direct graphical representation of a matrix is a table, but other displays are possible as well. In Exhibit 21.7, we show other typical graphical representations used in qualitative data analysis. On the right-hand side is a three-dimensional table depicted in two dimensions. In the first row's column heads, you would place two categories of the first dimension, for example, women and men for the dimension 'gender'. The second row could contain your observations; for example, the names of the persons from the field notes – Julia, Esther, Christos and Brent. The row headings in the first column would mark the occurrence of certain expressions, for example, 'surface acting' or 'deep acting' as forms of emotional self-regulation. Julia, Esther, Christos and Brent are employees in a call centre and the company's policy is to be always polite and friendly towards in-calling customers. We had interviews with the four and we asked them to describe situations in which very angry customers were calling in, how they reacted to them and how they felt about it after they finished the phone call. From the interview information, we would take out information chunks and place them in the cells. Thus, if Esther reported that she could understand the anger of the customer although it was not the fault of the company, you would place the description of the situation and how Esther felt after the call in the appropriate cell. In the case where Brent still smiled and was friendly to an angry customer, but at the same time was thinking what an idiot the customer was, you place the information in the cell Men | Brent | Surface acting. In the next step, you would try to see whether there are certain patterns in the table.

Hierarchical and networks graphs are especially useful to depict the context of a situation and to describe the roles persons hold. But you should be aware that you do not just produce a hierarchical graph like a formal organization chart. In your analysis you are free to add further information to the charts; for example, information describing the lines between the different persons. It might be that the relation between person A and C is amicable, while the one between A and E is characterized by many fights. Network graphs are well suited to displaying social structure. Nodes represent persons or events and lines represent ties or relationships. Look at the network graph above from two perspectives. In the first perspective, you assume that the nodes represent colleagues and the lines represent whether they meet each other in private. In that perspective, we would find that Ms Blue is very popular and maintains contact with almost all her colleagues. It could be that the reason for this is that she is the head of the

department, but it could also be that she is networking very well. On the other side, Mr Purple, he is a loner and to analyse this we would need to fill the graph with further information. Is Mr Purple the head of the department, is Mr Purple working at the same location or is he the lone wolf running the Southeast Asia business from Bangkok? What these examples point out is that producing such graphs does not immediately produce the result of a study. Those graphs are more a tool that facilitates the systematic investigation of the qualitative data one has collected. Finally, the process and the relational graph are suited to structure information in terms of causal relation, that is, what are antecedents and what are consequences. Moreover, one can use them to display sequential processes that evolve over time.

Quantitative quasi-statistics

Quantitative quasi-statistics refers to the fact that one can convert qualitative data into quantitative data, but that the information density of such generated quantitative data is limited and, consequently, also the statistical methods that can be used upon them. In most cases, the statistics are limited to simple frequency counts, sometime the calculation of mean and, infrequently, the estimation of a simple bivariate correlation. Nevertheless, using such simple quantitative tools, such as counting, can help us to develop a better informed picture regarding our qualitative data. In Chapter 10 on content analysis, we discussed how such quantitative quasi-statistics can be produced.

21.4 Computer-assisted qualitative data analysis

While in quantitative data analysis, computer software to conduct analyses, such as SPSS, SAS, stata or R, have a long history, the use of software packages for qualitative analysis is a much more recent development. Currently, the abbreviation **CAQDAS** (computer aided qualitative data analysis software) is commonly used as a label for these software packages. A well-known typology[13] classifies CAQDAS in the following categories:

- code and retrieve
- code-based theory building
- text retrievers
- text-based managers.

The first two categories support researchers in coding qualitative data, that is, in assigning codes to chunks of the information sources. Moreover, these packages usually enable automatic searching of the information sources for specific key words. Text retrievers and text-based managers allow a more content-based analysis. Their search features are more advanced and often they include add-ons that allow the creation of occurrence tables, active word lists and even matrices that count the co-occurrence of words in single sentences. The rapid developments in IT have led, however, to newer packages often building in these different features. Therefore, we discuss the different generic CAQDAS utilities. The following utilities can be distinguished:

- content searching
- linking
- coding
- writing and annotation
- mapping and networking
- integrated quantitative utilities (counting, tabulation etc.).

How CAQDAS can help in the actual analysis of qualitative data can be shown by referring to key elements of the analysis process.[14]

Qualitative data analysis is usually based on multiple information sources, such as transcripts from different interviews, internal memos, etc. CAQDAS allows you to structure the research more efficiently, as the software combines all the information sources into one project database. Comparison of different sources are enabled to detect differences and commonalities. Many packages also allow you to create additional files, such as literature lists, abstracts or outcome files, within the same database, which facilitates the cross-referencing between the collected data and outcome files. Thus, CAQDAS is valuable in organizing qualitative data in meaningful ways.

Data exploration is another key element of qualitative data analysis and CAQDAS supports this process by enabling annotations and linking. Thus, while reading through the data material you can make annotations to note down your thoughts. Linking allows the searcher to mark linkages in the data material. On a paper hard copy, you would draw a line or an arrow to mark such a linkage, but doing so becomes complex if the data volume increase. Linking information chunks that appear far apart on text pages leads to confusing lines on paper, but is manageable in a software package.

Coding and retrieving is at the foundation of qualitative data analysis and, therefore, a key element in CAQDAS. The packages support researchers in all coding procedures, inductive or deductive coding. It supports researchers by allowing them to easily keep track of the codes introduced, re-coding does not require all the material to be re-read anymore but is more or less just the pushing of a few buttons. Most importantly, once the coding is completed retrieval functions make it easy to get an overview of which information chunks are linked to which particular code. This feature is already helpful during the coding. Assume you doubt whether you should assign a specific code to a piece of information you just read. With the retrieval function, you can quickly get an overview of the information pieces to which you assigned the code previously and that will enable you to better judge whether the new information piece should bear that code as well or not. The software packages do not provide the codes; researchers still need to generate them, building on their own insights, knowledge, experience and intuition. CAQDAS is a support tool facilitating the coding process but the software does not take over.

Searching is another element in the qualitative data analysis that has become much more easy through the use of CAQDAS. Assume you have transcripts of 20 long interviews with managers and employees of a service firm. Reading through the interviews you obtain the feeling that the relation with one specific customer has a large impact on how the company perceives the market. Without the aid of a computer, you would need to read through all the documents looking for the name of the customer and the longer the documents are, the higher the chance that you will miss an occurrence. Computers will not miss and highlight each occurrence of the customer's name in your data. Search tools are also useful in the coding procedure, as you can easily identify similar information pieces throughout your information database.

The main strength of CAQDAS is that it facilitates the reduction and structuring of qualitative data. More and more packages also offer advanced *output* tools: the generation of standard reports that can be imported to other programs or printed is a standard feature; for example, you can generate a report that contains all information chunks of one or more codes. In more advanced packages, it is also possible to create tables with either numeric or text information in the cells. We earlier mentioned co-occurrence matrices as another advanced table. Mapping is a further output tool in which one can create graphs that link information chunks and import those graphs to standard text processing or presentation software. A major methodological advantage of CAQDAS is that it increases the rigour of the analysis as it helps to ensure that the same analysis principles are applied to the qualitative data collected; with limited effort, anybody could replicate the analysis once the principles of the analysis have been chosen. Another advantage is that the automation of the processes allows the handling of much larger amounts of data, which increases reliability and validity as well. Finally, CAQDAS makes qualitative data analysis more efficient.

There are, of course, a few drawbacks as well. The major drawback is that CAQDAS might seduce researchers to narrow their research approach in a way that fits the usage of the software. Some researchers are afraid that the inflexibility of the software will hamper their interpretative freedom. Moreover, CAQDAS output depends on the input – garbage in, garbage out. Thus, if your qualitative data has deficits or the queries you start are not well thought through, nothing meaningful will be output. This truth, however, applies to any kind of research. In the end there is also an economic disadvantage, as it takes considerable effort to learn these software packages and also to put the raw data into a format that is suitable for them. An old Dutch saying, '*de kost gaat uit voor de baat*' points to the fact that you need to incur costs before you can reap the benefits.

NVivo

What SPSS is to quantitative analysis, NVivo is to qualitative analysis. An overview of other software packages is depicted in Exhibit 10.3 in Chapter 10. NVivo contains all the utilities mentioned above. The most recent version (10) also allows you to capture information from the web and especially social media sites, such as YouTube and Twitter, as you can directly import information from those sites.

Summary

The clear distinction in this book between methods discussed in the hard copy and analysis discussed in the additional chapters available at the website is somewhat blurred in the case of the qualitative data analysis. One reason for this is that in qualitative studies information collection and its analysis happen partly simultaneously, but the other reason is that certain analysis techniques we discuss are closely interwoven with certain types of data collection; for example, content analysis and coding. Coding is an interpretative process in which we reduce and structure the information available by placing labels, codes, against chunks of information. Those codes can be prescriptive – we define them before the coding starts – or they can emerge during the coding. Coding is an iterative process that starts with generating as many codes as possible, but further on in this process we compare the codes generated to form a smaller and smaller number of categories. The number of analysis tools available in qualitative data analysis is large and which one to use depends on the purpose of the investigation. In this chapter we discussed typology, induction, domain analysis, hermeneutical analysis, phenomenology, heuristic research, semiotics, matrix and logical analyses as well as quantitative quasi-statistics. Due to developments in information technology the use of the computer is no longer limited to quantitative analysis and a reasonable amount of standard software packages specifically developed to support qualitative analysis exists. The main functions of such packages include data exploration, coding and retrieving as well as searching.

Discussion questions

Terms in review

1 What is the purpose of qualitative data analysis?

2 What is the difference between prescriptive and emerging coding?

3 What are the most common concerns regarding coding?

4 Describe the relationship types in domain analysis.

5 Describe the different display formats in matrix analysis.

6 What are the main functions and utilities provided by CAQDAS?

Making research decisions

7 When would you use verbatim transcribing and when formal?

8 How can we balance keeping context and specific information on the one hand, while on the other hand the development of parsimonious theories requires that we need to work with a limited set of categories?

9 What distinguishes domain analysis, hermeneutical analysis, heuristic research, semiotics and discourse analysis?

10 To represent a cause and effect relationship, which graphical representation would you choose?

11 Which graphical representations are best suited to make comparisons?

12 Which data analysis tasks should be left, can be left and should not be left to CAQDAS?

Classroom discussion

13 Provide groups of 2 or 3 students with text – for example, an interview with a politician or CEO from a newspaper – and ask them to code the text. Discuss how the different groups have coded the text.

14 Discuss in class the advantages and disadvantages of CAQDAS, especially the issue that the software package might pre-induce a certain structure onto the qualitative data and analyses.

Recommended further reading

Miles, Matthew B., A. Michael Huberman and Johnny Saldaña. *Qualitative Data Analysis: A Source Book* (3rd edn). Thousand Oaks: Sage, 2013. The widely cited classic on qualitative data analysis.

Get started with understanding statistical techniques!

When you have read this chapter, log on to the Online Learning Centre website at *www.mcgraw-hill.co.uk/textbooks/blumberg* to explore chapter-by-chapter test questions, additional case studies, a glossary and more online study tools for *Business Research Methods*.

Notes

1 C. Geertz, *The Interpretations of Culture*. New York: Basic Books, 1973.

2 L. Richards, *Handling Qualitative Data: A Practical Guide*. London: Sage, 2005.

3 H. Mintzberg, 'Structure in 5's: A synthesis of the research on organization design', *Management Science* 26(3), 1980, pp. 322–41.

4 F. Znaniecki, *The Method of Sociology*. New York: Farrar & Rinehart, 1934.

5 J.P. Spradley, *The Ethnographic Interview*. New York: Holt, Rinehart and Winston, 1979.

6 M.V. Jones, N. Coviello and Y.K. Tang, 'International entrepreneurship research (1989–2009): A domain ontology and thematic analysis', *Journal of Business Venturing* 26, 2011, pp. 632–59.

7 C. Durugbo, 'Work domain analysis for enhancing collaborations: A study of the management of microsystems design', *Ergonomics* 55, 2012, pp. 603–20.

8 U. Eco, *The Theory of Semiotics*. London: Macmillan, 1976 (cited on page 7).

9 U. Eco, 'Narrative structure in Fleming' in E. del Buono and Umberto Eco (Eds.) *The Bond Affair*, London: Macdonald, 1966.

10 E. Bell, 'Ways of seeing organizational death: A critical semiotic analysis of organizational emorialisation', *Visual Studies* 27, 2012, pp. 4–17.

11 C.M. Fiol, 'A semiotic analysis of corporate language: Organizational boundaries and joint venturing', *Administrative Science Quarterly* 34, 1989, pp. 277–303.

12 M.B. Miles and A.M. Huberman, *Qualitative Data Analysis: A Sourcebook of New Methods*. Newbury Park: Sage, 1984.

13 E. Weitzmann and M. Miles, *A Software Source Book: Computer Programs for Qualitative Data Analysis*. Thousand Oaks: Sage, 1995.

14 A. Lewins and C. Silver, *Using Software for Qualitative Data Analysis: A Step by Step Guide*. London: Sage, 2007.