

Chapter 7

Qualitative data analysis

7.1 Introduction

In chapter 2 we very briefly touched upon qualitative research. We presented it – simplifying matters somewhat – as juxtaposed with quantitative research: where quantitative studies describe a limited number of issues, in isolation, for a large number of people, qualitative research describes many issues, in their interrelatedness, in a specific context, for a not necessarily representative set of people. This means that, where quantitative research often isolates variables and studies the impact of these on other variables, in qualitative research pre-defined variables or factors are less often studied. In fact, a qualitative approach is generally chosen because the factors of interest need to be studied within a specific context, or in their complex or layered interrelation with other variables. Also, many qualitative researchers argue that human behaviour cannot be understood ‘from the outside’, but that we need to know the *meaning* that events and behaviour have in order to be able to understand why people behave as they do. To uncover meaning, persons need to be queried repeatedly or deeply, or observed within their natural habitat, necessitating naturalistic and participatory approaches to data collection. Context and meaning are key concepts in any qualitative study.

This has important implications for the type of data that are gathered and the manner in which these can be analysed. As qualitative researchers argue that behaviour is understood best within its context, it is not strange that qualitative researchers rarely use standardized, ‘pre-fab’ instruments, because, given that each context is unique, such standardized instruments are inherently useless as they might easily skip essential elements of the context. Therefore, more often than not, qualitative researchers actually start their analysis from the physical world and human behaviour as observed in it. The data themselves can be the starting point of analysis, and meaning and structure may be derived from the data and not imposed upon it as a *Fremdkörper*. Some grounded theorists even go as far as to do their literature review only once all the data have been collected.

This – for a quantitative scholar ‘topsy-turvy’ – approach is given the epistemological outlook of qualitative research inherently logical. Phenomena can only be under-

stood after the researcher has grasped the meaning of behaviour, of symbols, happenings, occurrences. Hence the strong emphasis of qualitative research on embeddedness, on immersing oneself as a researcher in the world of those who are to be studied – in corporate culture, in the lives of recipients of welfare – and the extensive use of open or in-depth interviews, in order to *understand* in the research why decisions are taken by municipal staff as they are, or whether welfare recipients perceive administrative procedures as fair. For that matter, as we stated in chapter 2, qualitative research is also referred to as *interpretative* empirical, with quantitative research conversely labelled as *analytical* empirical; parallel labels are *inductive* for qualitative research and *deductive* for quantitative.

It must be noted, as we also said in chapter 2, that this contrast between qualitative and quantitative research is slightly overdone and overly simplistic. The two are not completely different, irreconcilable approaches to empirical legal research. In practice, many studies combine qualitative and quantitative elements. Much qualitative research is not entirely ‘grounded’, but works from theoretical assumptions, or employs existing legal categories or theoretical constructs. Nor is quantitative research entirely inductive, as often exploratory, more qualitative phases are built into the research design. This means that there is no watershed between qualitative and quantitative research, and that qualitative and quantitative analysis methods are not in entirely different leagues. Qualitative researchers also aim to describe prevalent or dominant patterns, or are seeking relations between variables. As we will see below, some analysis methods for qualitative data may be in fact entirely quantitative.

7.1.1 The iterative nature of qualitative research

Regardless of the above, the more ‘prototypical’ qualitative analysis methods can be contrasted with quantitative methods as on average less prescribed, less ‘recipe-wise’, and more iterative. This also holds true for qualitative data collection procedures, which are generally not often very closely prescribed and are instead flexible and adaptive. Especially ethnographic researchers generally start by interacting with the people they aim to study, and may draw up hypotheses as they go along. They may ask as many questions as they need to understand the phenomenon they are studying. As they do not know beforehand what will cross their path, these interviews are far from pre-designed questions administered by web survey: they must take place in preferably face-to-face settings, where researchers are able to note facial expressions, hear inclinations and deflections of voice, see emotion or boredom in people’s faces.

In some qualitative approaches, a participatory stance is taken, where researchers attempt to view the world from their research participants’ perspectives. They cannot, goes the reasoning, really understand their respondents’ predilections, motives, constraints, unless they are able to – in a sense – lead their lives with them. An example of such a study is the famous criminological study on gang life and the urban poor in Chicago by Venkatesh (2008), which has been immensely important for theory formation on the pathways in and out of gang membership. The study by Flood (2005) similarly was one where the researcher attempted to immerse himself in the corporate world he studied, and was in that sense ethnographic: Flood was even hired by the firm that he studied, sat in on meetings and joined office parties, like a regular employee.

We discussed at length in chapter 4 how qualitative data collection generally occurs in open or topical interviews, and how the data that results from such fieldwork generally consist of transcriptions of interviews, field notes and analytic memos. Qualitative data can also consist of documents or video material (or even movies, advertisements, YouTube uploads or Instagram pictures). What all these data have in common is that they are generally (highly) unstructured. In order to make sense of them, to analyse them and arrive at conclusions, we must therefore structure the material, reorganize it so that an overview can be made. Then we will be able to see the woods for the trees, and structure may emerge and the research questions be answered.

In qualitative data analysis, that process is often iterative. By that we mean that analysis does not take place in one pre-designed sweep as it does in quantitative data analysis, but that the researcher in fact iterates over explanations of the data. It is possible, after coding textual material and analysing the codes in a second step, to return to the original codes and modify them, and then re-do the analysis step. The qualitative researcher may in each step visit the raw material, draft an analysis or explanation, revisit the material and check – looking at all the transcripts and notes and memos – whether that analysis is tenable or needs to be adjusted, adjust the hypothesis, then revisit the data to see whether the adjusted hypothesis now holds, etc., until a conclusion is reached that is satisfactory and is in line with everything that is in the data.

The research cycle that was introduced for quantitative research in section 2.2 and that is travelled through only once for a quantitative study may in fact be traversed in as many loops as are needed for answering the research question in qualitative research. This can even go so far that hypothesis formulation, data collection and analysis are interwoven, and are performed iteratively, with as many waves of data collection and analysis as necessary. Researchers then collect data, analyse the data and arrive at a tentative conclusion (for example written down in analytic memos) and then return to the field, collect new data and test their tentative conclusion against new information that was gathered. Once such iterative loops do not lead to adjustment of the conclusions (no new insights are gathered from additional interviewing or observation), it is said that *saturation* has been reached (see for a classic reference Glaser & Strauss, 1967). This process, where one samples respondents up to the point where adding new sample members would not change the conclusions, is referred to as *theoretical sampling*. It is often used when purposive samples are drawn. At the point where the researcher decides that data collection can be stopped and saturation has been reached, it is then also said that high *theoretical validity* has been achieved.

It is generally advised to test this to the limit in the sense that researchers are encouraged, once their preliminary interpretation of the data has been formed, to actively seek out cases for inclusion in the sample that do not seem to match the tentative conclusions. This practice of deviant or negative case analysis could be regarded as an aberration if one were to strive for randomness of selection research subjects. It can, however, also be regarded as a kind of quota or stratified sampling, where one aims to have sufficient spread in the sample over possible subgroups of respondents.

A related issue is the analysis of outliers. These can be regarded as deviant cases: they do not fit the general pattern. In quantitative analysis these are generally seen as a nuisance. One should in quantitative analysis scan for such outliers, and if they appear they are often ignored in analysis. They are regarded as an anomaly, as exotic and not

indicative of a relevant pattern for the majority of the investigated sample. Contrarily, in qualitative analysis it is considered imperative to include these outliers. They are perhaps even regarded as the canary in the coal mine, as specific cases that may indicate where the extracted pattern or devised theory is simply not ready yet, which may lead to a major improvement of the conclusions that have been reached so far. Qualitative researchers attempt to incorporate such cases to the very end, revisiting their theoretical framework and analysis to see whether the outlier can be made to fit, or whether one needs to conclude that it does not fit after all. Flyvbjerg (2006) is an interesting read on the usefulness of the study of such outliers, or more broadly case studies, that may put one on the path to discovering a black swan. All in all, sampling and analysis may be interwoven in qualitative research. We will for that reason sometimes in this chapter on data analysis also discuss data collection issues.

One important implication of this is, as discussed, that in qualitative research, we often do not work with just one research cycle, but in a sense a (sometimes very long) corkscrew of mini-research cycles. A qualitative study is therefore by itself a spiral of small research cycles. By the time the explanation or interpretation of the phenomenon under study converges, and does not change anymore with new respondents or data added, nor even with negative cases added, the research ends and conclusions can be drawn. Qualitative research unfolds all in all quite differently from the quantitative research mould. It is for qualitative research generally much harder, therefore, to predict what a sufficient number of sample members will be, how long the study will take, and therefore to budget costs. It must be said that also this depiction is slightly overdone, as in the real world also qualitative researchers must draw up budgets and plan their studies.

Before we start our discussion of qualitative data analysis, it should be noted that there appear more different takes among qualitative researchers than among quantitative researchers on what suitable or preferred methods are for specific analysis questions. As qualitative research may study idiosyncrasies, many qualitative researchers are also bound to adhere to their own, or special analytic 'culture'. Textbooks on qualitative research are for that reason often an amalgam of various methodological approaches and 'schools', working from different epistemological stances. While there is fairly general agreement among quantitative researchers on how a regression analysis should be carried out (although at times there can be wide-ranging differences too on the 'right' approach for answering a specific quantitative research question!), there does appear to be less agreement among qualitative researchers on what suitable methodological approaches are. Given the scope of this book, it will be impossible to address all these different schools and traditions and nuances. What we will do is discuss a number of analytic approaches and tools that are shared among most qualitative scholars.

7.1.2 Modes of qualitative data collection and analysis

As we discussed in previous chapters, qualitative data collection methods can consist of various types: ethnography, where the researcher in a sense immerses him- or herself in the world of the objects of study and where a combination of approaches is generally used (interviews, observation, dialogue, documents), interviewing, where

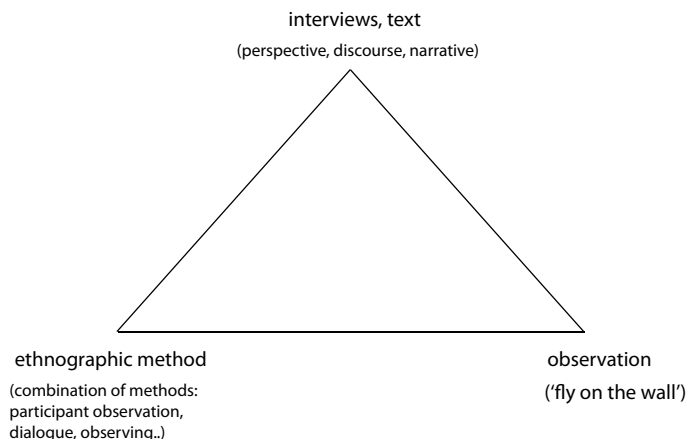
generally open or in-depth modes of interviewing or narratives are used, and observation, where the researcher does not take part in any social interaction with the research subjects (the researcher is in a sense a ‘fly on the wall’, not participating, unobtrusive and only registering behaviour). As we discussed briefly, the use of documents or secondary material, such as published court rulings, where essentially the researcher also ‘observes’, can be considered a separate source. Under such ‘patient’ sources, we also classify blogs, videos, films, advertising material, diaries and the like.

These modes can be thought of as being on a continuum of ‘participation’ or ‘immersion’: the most participatory method is then ethnography, where the researcher may him- or herself even ‘become’ the instrument as we said; interviews assuming a middle position, where the researcher plays a role in what is said as s/he poses the questions, decides on what to probe, registers emotions (or not), so that the researcher influences to a certain extent what behaviour and verbal utterances the respondents show; and observation, under which also falls the study of documents, as the most ‘unparticipatory’, where a researcher need not even have any exchanges with the respondents.

On the other hand, these three modes of data collection can also be ordered differently, on a continuum of ‘naturalness’ with which study participants may be expected to behave. Interviews are then the mode in which respondents’ behaviour would be least natural, ethnography would assume a middle position as the researcher will collect data such that the study participants can go about their ways as normally as possible but the researcher remains a stranger in their midst, and observation (under which again the use of existing materials) would be the data collection mode (especially when unobtrusive methods are employed such as through CCTV cameras, the ‘fly-on-the-wall’ mode) where respondents can behave as naturally as possible. Figure 7.1 depicts this in schematic fashion.

All analysis of qualitative material, whether gathered using ethnographic methods, interviews or observation, can be carried out employing a ‘*formal mode*’ and an ‘*informal mode*’. By the formal mode, we mean methods to structure and analyse the material using tools and methods that are interchangeable between researchers. A researcher can employ for instance first-level coding and second-level coding using the tools that we point out below. However, during data collection, informal methods may also be used. By these we mean for instance that researchers may also pick up implicit information that is unique to one particular researcher, such as a specific response that fieldworkers encountered upon entering gang turf, or a response by an interviewee that may be tied to characteristics of the researcher.

Such information can definitely be used in analysis: in fact, some parts of texts or of interviews may even give you as a researcher a particular feeling, or even a sensory experience such as nausea or a shudder, which may offer important clues for analysis. And such sensations should therefore definitely be employed in analysing the material. During coding (formal mode), researchers should then ask themselves: why does this particular sentence or expression affect me so much? Pondering such particularities that may be tied to yourself as a data collector or researcher is important, and qualitative researchers should indeed in this sense be sufficiently ‘reflective’. The term ‘positionality’ is often used to denote how researchers can and should reflect on their gender, social class or specific privileges that impact data collection and other research methods.

Figure 7.1: Schematic representation of three modes of qualitative data collection

Clearly, such information picked up that is related uniquely to the researcher is less easily exchanged with or found by other researchers. A particular response may be tied to a researcher’s gender, so that a researcher with a different gender would not even pick up the cue, or responses may be due to cultural differences or preconceptions. Researchers may use meta-information that they happened to pick up or gather during the fieldwork or in other phases. Using such informal modes of analysis is part and parcel of qualitative analysis. All in all this means that qualitative data analysis is influenced by personal experiences and individual characteristics of the researcher. This makes it inherently more difficult to achieve reliability. In the following we will mainly focus on the more formal modes of data analysis.

Qualitative data is usually massive and unstructured. Qualitative data analysis therefore generally consists of two steps: one to order the unstructured data, and a second to analyse these ordered data in order to answer the research questions.

7.2 Coding qualitative data

As said, before the analysis of qualitative data can start, the data have to be ordered in a manageable format. Data can be interviews, transcripts, minutes of board meetings, court verdicts, plea bargains, newspaper reports, notes, or even pictures or videos. They have to be transformed into a format that is analysable. If the data are not written text yet, they are generally textified, that is, turned into written text. Interviews are transcribed for that: all that is said is literally typed up, sometimes including coughs and ‘ehhrrmms’ (see section 4.2).

Most researchers these days use software to order and analyse their data. CAQ-DAS (computer-aided qualitative data analysis software) is the generic name for these packages. They all have in common that they make it possible to structure and restructure the data, regroup fragments, attach codings and subcodings to text blocks, and look for patterns in the data. Such software is almost necessary to get an overview of the often massive amounts of material (transcripts of interview can easily add up to hundreds of typed pages). A veteran programme is ATLAS.ti (atlasti.com), which has been around for many years and is available for many platforms (Mac, Windows, and even a smartphone version is available). Friese (2019) offers an accessible and hands-on manual for coding using ATLAS.ti. Other packages are NVivo (qsrinternational.com) and MAXQDA (maxqda.com). In addition to tools for coding, most commercially available packages also aid in aggregation and visualization of the data. Quite a number of free packages are available as well, but these generally offer coding only; an exception is qdap (the acronym for Quantitative Discourse Analysis Package), an R-based package that offers word extracting, statistical analysis and visualization (see cran.r-project.org/web/packages/qdap/index.html).

However, for analysing small bits of data, such a package is not necessary. Especially for the first sweeps in coding, when the aim is to get a feel for the data, the text may simply be highlighted with felt pen, or some first coarse coding done by hand.

7.2.1 Description: first-level coding

Analysis generally starts with first-level coding, also referred to as primary-cycle coding or first-cycle coding. In the analysis of qualitative data, one is generally faced with an immense amount of unstructured, ‘raw’ data. The goal of analysis is to get an overview, to describe what is there, and to find relationships between pertinent constructs, or to detect logical schemes. In order to be able to do so, the data first has to be reduced to manageable shorter units. This is what the first-level phase is about: in the first phase of coding the aim is to develop a list of codes under which the information in the data can be subsumed and therefore ordered. The code list should capture the relevant issues that are there in the data. All the written material should, formulated differently, be classified under a number of codes, after which the codes can become the focus of analysis.

But this may still sound somewhat mysterious, because what is a ‘code’? Saldana (2013, p. 3) defines a code as:

“A code in qualitative inquiry is most often a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data. (...) The portion of data coded during first-cycle coding processes can range in magnitude from a single word to a full paragraph, an entire page of text or a stream of moving images.”

First-level coding is therefore a first, descriptive step. In first-level coding one reads through all the data and attaches codes to relevant parts of text. Here one essentially

tries to attach labels to chunks of the data, to group sentences or pieces of text under various descriptive headers. First-level coding is in a sense quite ‘flat’ and easy as it is descriptive, one only codes *what* is in the data. The aim is essentially to arrive at a sort of inventory of the information content. Pieces of text (or other coded material) are grouped under the codes that are homogeneous: they all refer to the same phenomenon, meaning, interpretation. Segments under the same label refer to the same issue, segments under different labels refer to different issues. Coding is therefore a process in which one identifies similarities and differences in the information content of the data. Every relevant segment of text should be coded (and may in fact also receive two or three codes).

Coding can be done on the basis of an existing list of codes (‘top-down’), or it can be done in a more ‘grounded’ way. In the first approach one would for instance be testing a theory that uses particular constructs, or searches for specific legal arguments used in rulings: encountering these known categories, they are marked and labelled as such. When, on the other hand, coding is done in a more grounded way, one scrutinizes the data for categories or themes that emerge from the data. Coding using existing categories is also referred to as ‘etic’ or ‘deductive’ coding; coding employing categories as they present themselves from the data is also referred to as ‘emic’, ‘open’ or ‘inductive’ coding’. Deductive coding, with a set list of codes, is arguably easier to replicate by other coders and reliability is therefore sooner achieved (see section 2.6).

It is difficult to pinpoint exactly how one determines codes inductively. The most basic technique is essentially reading and eyeballing the data: reading the material once, twice or even a third time, and noting phrases or words that ‘appear’ to be ‘key’. Another observational technique to extract themes from transcribed text is to focus on repetitions: when you interview people repeatedly, or groups of people, there will be issues that recur. These may be ‘fairness’ or ‘respect’ or ‘class differences’. Apparently, as such themes emerge so often, they are important, they clearly constitute a relevant issue. Metaphors may often signal so-called ‘themes’. Quinn (1996, in Ryan & Bernard, 2003) studied American marriages. She noted that respondents often used metaphors to indicate that marriage is intended to last: interviewees would state that a marriage was like ‘the rock of Gibraltar’ or was ‘nailed in cement’. Apparently, this lastingness was a dominant issue. The term ‘theme’ is often used interchangeably with the word ‘code’ in inductive coding.

In spoken and written texts, transitions such as new paragraphs or a pause in speaking often indicate that a new theme is coming. And what may be equally important, is to not only search when identifying themes for what is there, but also for what is *not* there. Silences in interviews, looking away, can be as telling as words. Information that is missing, gaps, things left unexplained may be important clues. Ryan & Bernard (2003) give the example of interviewees who did not respond that they disapproved of government policy, but simply went silent if a question to this effect was asked. Attaching a theme becomes more slippery here, however, because how can one interpret something that is not there? And how can one be sure that the interviewee is silent because s/he disagrees with government policy? Maybe the interviewee is simply fed up, or thought the question trivial.

We illustrate the first-level coding process using data excerpts from a study by Bijleveld et al. (2022). Employing vignettes, the authors asked respondents to what extent

they considered fair (on a rating scale) a sentence of eight years for a perpetrator who had committed war crimes. In the vignettes the authors randomly varied – amongst others – whether the perpetrator had apologized or not, and whether the perpetrator was a commander or a foot soldier. After respondents had indicated whether they believed the sentence was just right, too short or too long, they were asked to elaborate on the reason for their answer. Figure 7.2 gives an illustration of the type of data the authors gathered, and a first round of coding made by the authors.

Because the authors had randomly varied aspects of the vignettes, they first searched for mention of these aspects. They therefore scanned the comments for mention of the apology made (or not made) by the fictitious perpetrator, and for mention of the rank of the perpetrator. In addition, they also employed inductive coding in that they searched for other arguments that respondents thought pertinent. In this study, a mixture of deductive and inductive coding was therefore used. Given that the authors interviewed 570 respondents, the example with 14 motivations can give only a glimpse of the coding process. Even so, what we see is that from the list of given codes, apologies were often mentioned: seven times in all, of which five times that no apology had been given by the perpetrator and twice that an apology had been given. Also twice it is mentioned that the sentence should be long because the perpetrator was a commander in charge. Once it was mentioned only in interaction with the apology: the apology is regarded as insincere because, as a commander, the perpetrator was in charge.

The respondents also gave other arguments, not related to the manipulated factors, for the sentence being too short (mostly) or just right. Seven times it was mentioned that the crimes were too serious for the sentence. Other arguments emerged for favouring a longer sentence: deterrence was mentioned, and the fact that the offender needs more time to rehabilitate, or justice for victims. We also see in Figure 7.2 how the coder was hesitant at times, unsure what code to use. For instance, the second respondent mentioned that the perpetrator should be tried for war crimes in a similar way that Nazi commanders were prosecuted, but the coder did not know yet under which header to capture that part of text.

This brings us to the fact that coding is seldom finished in one round. Researchers generally attach codes to the various segments of text in their data in a first sweep, and then revisit the data and check and change the codes somewhat. Software is very helpful here. Using packages like ATLAS.ti, one can reorganize the data to show all text that was given a certain code, to review. Review of the codes should verify that the text fragments are homogeneous and fitting. It could then emerge that some text fragments actually do not fit under a given label. Or that two subcategories exist within a larger category. During such review rounds, or during later analytic phases, it may also become apparent that two different codes actually describe the same phenomenon: the codes can then be lumped under a new header. For the example we gave above in Figure 7.2, the codes could for instance be regrouped into arguments in favour of a longer sentence, or arguments why the sentence was just right, or why it should be shortened.

Some researchers even go back, after a first analysis has been conducted, to the original text and apply new codes, and then re-do the analysis. Some in fact very quickly start coding once the first data is in. And they may then also go back to their codes and modify them after more data has been collected, and may go back again and

Figure 7.2: Descriptive codes vignette study of international sentencing

<p>I believe it is just because his crimes against civilians were heinous. He also did not exhibit any sort of remorse nor did he apologise, demonstrating that he does not see the error of his actions. As such, 8 years is quite necessary for him. However, it is already 10 years after the fact, and thus I did not give a score of 5 to include the long delay of the sentence, to signify that it is already quite late for this.</p>	<p><i>too serious crimes</i> <i>no apology</i></p> <p><i>fair trial</i></p>
<p>He may have been carrying out orders from even higher superiors (or doing them on his own which is worse) but that does not excuse his behaviour. After the second world war, Nazi commanders were also prosecuted for their crimes and I think in this case it is similar in that David should also be given a higher sentence.</p>	<p><i>duress</i></p> <p><i>serious crimes?</i> <i>consistency?</i></p>
<p>The sentence is way too short for having committed two deaths and tortured civilians.</p>	<p><i>too serious crimes</i></p>
<p>I think 8 years is enough in a sense because it is a reasonable amount of time, but I don't think it serves as justice for the victims. He did apologise to the victims so that's a step in the right direction</p>	<p><i>no justice for victims</i> <i>apology</i></p>
<p>I think that 8 years is very little for him having committed essentially a war crime with the torturing of civilians and the death of two people. Although he apologized we are not even sure whether he meant the apology because why would he have committed those acts as a Commander?</p>	<p><i>too serious crimes</i> <i>apology</i> <i>sincerity unsure because commander</i></p>
<p>8 years is too less for what he did. It doesn't change a person to lock him/her up for only 8 years</p>	<p><i>too serious crimes</i> <i>no effect on offender</i></p>
<p>Prison is necessary only when there is a big threat, which it doesn't seem to be here</p>	<p><i>incapacitation unnecessary</i></p>
<p>I believe that 8 years was unjust. Eight years is not enough for at least two deaths of innocent civilians.</p>	<p><i>too serious crimes</i></p>
<p>He was in charge and in the position to have avoided this.</p>	<p><i>commander position</i></p>
<p>Because he shows no remorse even after 10 years. Should be more, especially as the commander.</p>	<p><i>no apology</i> <i>commander position</i></p>
<p>No, the time period is not reflective of the death and trauma for the survivors. Nor is it adequate rehabilitation time for someone who does not show remorse.</p>	<p><i>too serious crimes</i> <i>no rehabilitation</i> <i>no apology</i></p>
<p>It is too short and thus does not set enough of an example as a deterrent for others who could commit such crimes in the future. He shows lack of remorse.</p>	<p><i>no deterrence</i> <i>no apology</i></p>
<p>He should have received a harsher sentence to reflect the severity of his crimes</p>	<p><i>too serious crimes</i></p>

make new choices again given additional data that they encountered or more recent interview material that has been added. First-level coding is all in all more often than not an iterative process. Spencer, Ritchie, & O'Connor (2010, p. 229) stress that in the final stages of this first-level phase – which is essentially a data management phase where the original data are summarized and synthesized in order to reduce the amount

of material and retain the essence of what was there – three requirements apply. First, key terms, phrases or expressions should be retained as much as possible. Secondly, interpretation should be kept at a minimum. Thirdly, phrases that do not appear to make sense or appear irrelevant should not be dismissed: their relevance may appear at later stages.

So in this coding phase, scholars often attach to the various relevant parts of the text labels that summarize, capture or indicate what is being said or written. The coding process is therefore definitely more than simply organizing the text: coding does always imply some interpretation and therefore analysis. Because labels can also narrow one's vision, some researchers prefer not to attach their own labels, but the very words that the interviewees use to code fragments. In doing so, one postpones the interpretation, reduces the risk of tunnel-vision or imposing pre-set notions. Such coding is called *in vivo* coding. Obviously, by using *in vivo* codes one remains closer to the original text, and retains the actual language used; however, in the choice of fragments, some organizational choices are arguably made which means that some interpretation is effectively there.

Obviously, if one were to strictly apply *in vivo* coding, one could end up with thousands of codes. This can also happen when one applies selected labels, as coding can be done in a very detailed way, giving each part of a sentence its own label. This is referred to as 'fractured' coding, and the process as 'splitting'. The advantage of this kind of fine-grained coding is that all nuances and little fragments are preserved and remain highlighted. Obviously, the downside is that the codings may become too fine-grained to be able to see much: there may be too much detail, one does not see the wood for the trees. Because the aim of the first-level coding is to reduce the amount of information, by combining text fragments into categories, this defies the purpose. In practice, researchers who start out with very detailed coding are for practical reasons often forced to lump categories together.

Some researchers on the other hand start out by using fairly coarse, general labels; this is referred to as 'lumped' coding. Those who start out more coarsely will quickly disaggregate labels into different, more specific labels. But how many labels should one use? There are not more than rules of thumb on how many codes should be developed. The number of codes firstly depends on the richness and breadth of the material. If the respondents all mention the same arguments, a small number of codes will do. If there is a lot of variability in people's answers, one would likely have to employ more codes, and many more so if *in vivo* coding is used.

Using a software program, it is technically possible to develop a list with 500 different codes. However, this list will likely be too large to be able to get an overview of what is there in the data, as so many codes are conceptually unfeasible. It is therefore, to get a grasp of the material, to be able to get an overview, much better not to work with too many codes. In the literature, a maximum of 100–200 codes is regularly mentioned, although in practice much lower numbers are encountered. An extensive and thorough guide to coding is Saldana (2013).

The codes that are given should be described and documented, and explained what kinds of text are given that label. Most researchers therefore draw up – just like during fieldwork – a certain diary-like record of all choices and alterations they made during coding and why, and an *analytic memo*. This is not only for the researchers' own

pleasure but intended to contribute to transparency of the analysis process and thereby reproducibility and replicability.

7.2.2 Analysis: second-level coding, patterns and typologies

After the raw material has been coded, analysis proper starts at the stage of second-level coding. Here researchers will interrogate the data with the research questions in mind. Here meaning has to be extracted, from the manner in which codes co-occur, or by looking at the profiles of respondents who give certain answers. There are various ways in which analysis can be carried out in this second phase, depending on the type of data available and the research questions that need to be answered.

Whichever approach one uses, it is customary to use *exemplars* or *vignettes* to present the findings, that is, parts of interviews or delineated portions of text or other material, collected during fieldwork, that are exemplary for what the researcher believes has been found. In some cases, respondents formulate so succinctly and precisely, or a life-story of a respondent in a sense illustrates the themes so well, that these are reproduced *verbatim*. Vignettes do the same thing, but are artificial exemplars in the sense that the researcher has constructed them. Note that the term ‘vignette’ has a different meaning here than it had in chapter 5.

It is at this second-level analysis stage that theoretical sampling may be employed: as meaning and interpretation emerge from the material, a search for empirical data that do not match the emerging interpretation can aid to strengthen the interpretation (when it turns out such ‘negative cases’ cannot be found) or broaden the scope or generalizability by extending the theoretical findings. Throughout the second-level coding, researchers should also keep a diary-like record of the interpretations that emerge, the choices that were made. Such a diary, which we already briefly mentioned in section 4.2, is referred to as an ‘analytic memo’.

Numerous textbooks have been written on the analysis of qualitative data. While relatively extensive as regards first-level coding, textbooks are less explicit when it comes to the analysis phase. This partly because analysis is generally of itself fairly particular or idiosyncratic. Some textbooks recommend that the reader simply starts at it, and Friese (2019, pp. 5–6) describes holistic analysis as follows:

“Take a printout of one or more transcripts. In summer, look for a delightful place outside; in winter, find a cozy spot in front of the fire. Have something to eat and drink and then spend some time reading your data from beginning to end in sequential order, reminding yourself of the interview situation or other contexts. Do the same with videos or other type of data you may have. There is no need to sit in front of the computer all the time. Reading the data via a different medium, such as on good old-fashioned paper, is likely to give further insights.”

This sounds woolly and far from concrete. While numerous approaches can be distilled from textbooks, we will describe four main approaches to the analysis of qualitative data after they have been coded at first-level. These are: interpretative coding, pattern analysis, axial coding and typologies.

Interpretative coding

A very illustrative example of interpretative coding is the analysis by Lindegaard et al. that we described previously in section 4.5. Lindegaard and her co-authors (2016) coded the behaviour of all people (victims, bystanders and perpetrators) present in a number of robberies from CCTV fragments. They coded numerous fragments of the video images in very general, descriptive and as neutral terms as possible, looking at the movement of people's limbs, the direction they were looking, whether they stood still. Codes given in the first-cycle coding phase were for instance 'hands at waist', 'pushing victim', 'touching arms of bystander'. Alongside to these behavioural categories, they also noted where the robbery took place, how many people were present and other situational characteristics.

In a subsequent coding step, Lindegaard and her co-authors *interpreted* these behavioural codes and attached labels to them that were not descriptive anymore but captured what that behaviour meant in that particular situation. They coded for instance whether the behaviour constituted resistance by a victim, or whether offender behaviour constituted a threat or actual violence. If a supermarket employee raises her hand, this can be because she puts her hands up as a sign of submission, but it could also be an attempt to grab the weapon of the robber (meaning taking control). This means that the authors must decide what the function or intention of the behaviour was.

The authors had set out to know what precursors predicted whether a robbery would turn violent. In a subsequent step, they therefore investigated whether there was a specific pattern or sequence of acts by bystanders, victims and robbers that often led to violence, and whether there were specific patterns or sequences that were associated with a robbery ending non-violently. Their analysis therefore also involved a search for patterns, over the time elapsed – often minutes – that the robberies took.

Pattern analysis

A second approach to analysing the first-level codes is through a search for patterns. It may be that to answer the research question we want to know whether, for instance, certain legal arguments are always used in conjunction: whether respondents who attach prominence to one aspect of sentencing tend to also answer in a particular way with respect to other aspects. This may sound pretty vague, so we illustrate it with an example. In the study by Bijleveld et al. (2022), the authors coded the arguments respondents gave for their assessment of the sentence length.

The authors noted that apologies offered by the perpetrator did not in general lead respondents to opt for a shorter sentence. A total of 54 respondents indicated that the absence of an apology should imply a long(er) sentence: "Lack of remorse deserves longer sentence", "Because [the perpetrator] does not show remorse and the crime is severe, there is a need for an appropriate punishment", "[The sentence is] long, but since not apologised it's suitable", "[The perpetrator's] lack of empathy and apology should also add time". However, an apology by the perpetrator much less often made respondents propose a shorter sentence or deem the sentence too long (13 times): "[The perpetrator] showed remorse, asked for forgiveness and recognized [the suffering of the victims]", "Too short for two deaths, although the apology helps", "[It is] already ten years later, [the perpetrator] offered his excuses and has also built a new life".

The authors also analysed the data quantitatively, which revealed that apologies significantly impacted respondents' preferred sentence lengths. The qualitative analysis now clarified how this impact was in fact non-linear, and adds on to the existing literature on the relationship between apologies and sentencing. That is, while the existing legal and empirical-legal literature on international criminal justice (Hola, Bijleveld, & Smeulders, 2011; Tudor, 2008) points to apologies and remorse as mitigating factors in sentencing, this study showed that for ordinary citizens it is the *absence* of apologies that may be considered an aggravating factor.

Note that pattern analysis may also entail simply tallying the occurrence of arguments. Wijntjens (2020) investigated, in a study of in case law that we return to in section 10.3, whether apologies were regarded by judges as constituting an admission of culpability. Wijntjens scrutinized almost 600 rulings and found that apologies were decisive in the ruling in only a little over 1% of cases. It is here that we see how qualitative analysis can very well incorporate quantitative analysis methods.

Pattern analysis can be greatly supported by software such as ATLAS.ti. The program has features to search for and investigate the co-occurrence of codes, whether certain codes follow each other, whether one code always encompasses another, etc. It is also possible to search for differences, such as whether people with certain educational levels, or different migration backgrounds regard the justice system differently. For more on this, see Friese (2019, chapters 6 and 7).

Axial coding

Axial, or hierarchical coding is the third approach that we describe for analysing the first-level codes. Axial coding is in a way the reverse of the first-level coding process. While in first-level coding one attempts to break up the text into a large number of codes, at this stage one tries to combine these codes into overarching dimensions or axes (hence 'axial' coding). With axial coding, central, overarching codes or constructs are identified that bind or encapsulate the first-level codes. The codes that are arrived at in axial coding are often elevated to the (somewhat confusingly labelled) higher status of 'category'. Axial coding is also sometimes referred to as hierarchical coding.

One way to do this is to reorganize the text fragments (and this is again where the qualitative data analysis software packages come in so handy) so that text that has similar labels is regrouped to appear together. One can then re-read all (parts of) sentences and see whether new, conceptual 'umbrella' concepts describe data under one label.

Figure 7.3 gives a worked example on data as were collected in the study by Bijleveld et al. (2022) that we used previously for illustration purposes. In the figure, the first-level codes and second-level, axial codes have been entered. In this example, three axes have been identified, of which one is eminently clear: 'Utilitarian' indicates whether someone employed arguments that indicated whether a shorter or longer sentence would 'work' for some reason (such as protecting society, rehabilitating the offender). The two other axes are less clear (and would likely be altered or removed in a next analytical run over the data). One of these is labelled 'Normative' and connects to the seriousness of the crime and what – given that seriousness – a suitable sentence would be. A third refers to justice, whether the sentence would be just. Justice can refer to victims, but it also does so twice to the perpetrator: whether it is fair to pun-

ish someone after such a long time, and whether the perpetrator did not commit the crimes under duress, although the respondent rescinds that argument in the same sentence. The codings show once more how second-level coding may require a number of iterations to arrive at a satisfactory analysis.

Typologies

While pattern analysis can be regarded as an analysis at the level of the variables, a similar kind of pattern analysis can also be conducted but now at the level of the respondents or sample units. Then, one is interested to find out whether certain combinations of codes, or profiles, can be distinguished that classify the sample members into groups. Such profiles define what are called 'typologies'.

In constructing typologies one divides the sample into a small number of groups, of which each and every member belongs to one and only one subgroup. Typologies are an organizational tool to divide or sector the real world, and thus to describe and understand it. Ritchie, Spencer, & O'Connor (2010) give an example of parents who had children with learning disabilities who had continued to live in the parental home. The authors of the study had constructed four typologies, using two dimensions or 'axes' along which parents could be ranked. The two axes were first the recognition of the need to find alternative arrangements for the child and second the immediacy of action by the parents. This resulted in the following groups: evaders (parents who felt their children need not leave the home), delayers (parents who knew their child had to leave the home eventually but did not take action yet), debaters (parents described as torn between the need for change and the difficulty of implementing it) and action-takers (parents who saw the need for change and who had already started preparations for it).

Typologies are often constructed with the aid of underlying axes, resulting in four or more typologies. We will return to the construction of typologies in section 10.2. The concept of typologies is fairly strict in the sense that it is required that each and every member must be fitted into one group and one group only (like a taxonomy). It is also fairly strict in the sense that people cannot have partial profiles: they cannot belong to two typologies at the same time. On the other hand, members within the typology groups need not all be identical; some variability is regarded as natural.

7.3 Content analysis

Content analysis is the broad generic name commonly attached to the analysis of large volumes of existing text documents. Content analysis in fact comprises a fairly broad class of analytic techniques that aim to analyse texts to reveal what information they contain. Both quantitative and qualitative approaches are viable to do so. Content analysis can be carried out using first- and second-level coding of the transcribed material, methods that we discussed just now. Often, and especially when the material to analyse is voluminous, special software is used to carry out the analysis of the qualitative material. Automated procedures for content analysis are increasingly used.

Although the category is broader, content analysis has often been used to reveal the meaning or symbolic or covert intent of communications, whether these be written,

Figure 7.3: Descriptive and axial codes in vignette study of international sentencing

<p>I believe it is just because his crimes against civilians were heinous. He also did not exhibit any sort of remorse nor did he apologise, demonstrating that he does not see the error of his actions. As such, 8 years is quite necessary for him. However, it is already 10 years after the fact, and thus I did not give a score of 5 to include the long delay of the sentence, to signify that it is already quite late for this.</p>	<p><i>too serious crimes</i> <i>no apology</i> <i>fair trial</i></p>	<p>NORMATIVE UTILITARIAN? JUSTICE PERPETR.</p>
<p>He may have been carrying out orders from even higher superiors (or doing them on his own which is worse) but that does not excuse his behaviour. After the second world war, Nazi commanders were also prosecuted for their crimes and I think in this case it is similar in that David should also be given a higher sentence.</p>	<p><i>duress</i> <i>serious crimes?</i> <i>consistency?</i></p>	<p>JUSTICE PERPETR.</p>
<p>The sentence is way too short for having committed two deaths and tortured civilians.</p>	<p><i>too serious crimes</i></p>	<p>NORMATIVE</p>
<p>I think 8 years is enough in a sense because it is a reasonable amount of time, but I don't think it serves as justice for the victims. He did apologise to the victims so that's a step in the right direction</p>	<p><i>no justice victims apology</i></p>	<p>JUSTICE VICTIMS</p>
<p>I think that 8 years is very little for him having committed essentially a war crime with the torturing of civilians and the death of two people. Although he apologized we are not even sure whether he meant the apology because why would he have committed those acts as a Commander?</p>	<p><i>too serious crimes</i> <i>apology</i> <i>sincerity unsure</i> <i>because commander</i></p>	<p>NORMATIVE</p>
<p>8 years is too less for what he did. It doesn't change a person to lock him/her up for only 8 years</p>	<p><i>too serious crimes</i> <i>no effect on offender</i></p>	<p>NORMATIVE UTILITARIAN</p>
<p>Prison is necessary only when there is a big threat, which it doesn't seem to be here</p>	<p><i>incapacitation unnecessary</i></p>	<p>UTILITARIAN</p>
<p>I believe that 8 years was unjust. Eight years is not enough for at least two deaths of innocent civilians.</p>	<p><i>too serious crimes</i></p>	<p>NORMATIVE</p>
<p>He was in charge and in the position to have avoided this.</p>	<p><i>commander position</i></p>	<p>NORMATIVE</p>
<p>Because he shows no remorse even after 10 years. Should be more, especially as the commander.</p>	<p><i>no apology</i> <i>commander position</i></p>	<p>NORMATIVE?</p>
<p>No, the time period is not reflective of the death and trauma for the survivors. Nor is it adequate rehabilitation time for someone who does not show remorse.</p>	<p><i>too serious crimes</i> <i>no rehabilitation</i> <i>no apology</i></p>	<p>NORMATIVE UTILITARIAN</p>
<p>It is too short and thus does not set enough of an example as a deterrent for others who could commit such crimes in the future. He shows lack of remorse.</p>	<p><i>no deterrence</i> <i>no apology</i></p>	<p>UTILITARIAN NORMATIVE</p>
<p>He should have received a harsher sentence to reflect the severity of his crimes</p>	<p><i>too serious crimes</i></p>	<p>NORMATIVE</p>

oral, in images or even iconic representations. Once all such communications have been coded, that is, once descriptive, first-level codes have been assigned, one starts to search for meaning in the coded material. Analysing transcriptions of for instance boardroom meetings in law firms, one can note what is said, literally – which is what is summarized in the first-level codings – but one is mostly after more, such as the meaning and impact of communication, or the symbols behind expressions, tones or gestures, the meaning that people attributed to events. In the literature, numerous labels are given to the methods to do so. Essentially what one does in content analysis is uncover themes in the qualitative material (and possibly sub-themes), selecting from all themes the ones that are relevant for answering the research question, connecting these themes in a meaningful way so as to answer the research question, and, lastly, linking the findings back to theory. In that sense, it does not differ in a fundamental way from the coding and analysis of regular interviews or field notes.

In content analysis, quantitative methods are employed relatively frequently. The qualitative material is then quantified, and analysed using regular quantitative methods. In such a quantitative content analysis in its simplest form, one could for instance count the frequency with which certain words appear (such as ‘solution’ or ‘conflict’) in tort cases, or how often victims in criminal cases are referred to as ‘vulnerable’ or similar words. Quantitative analysis can be made more interesting by not only looking at means or percentages (‘univariate analysis’), but by combining occurrences of words. One would then look at relational patterns: do for instance lawyers use different adjectives for male than for female clients? In a sense, one then carries out a bivariate analysis. In multivariate varieties one can look at combinations of words or phrases and contexts. Some such content analysis methods are quantitative – from basic to quite ‘hard-core’ – and others are qualitative and flexible – and hard to describe in a recipe-like manner.

Schmader, Whitehead, & Wysocki (2007) carried out a more quantitative analysis of qualitative material. They analysed letters of recommendation for male and female biochemists. Using special software for text analysis (Linguistic Inquiry Word Count (LIWC) Pennebaker, Francis, & Booth, 2001), they analysed how often particular words appeared in letters of recommendation as gauged against the total number of words in the recommendation letters. These words were: achievement words (e.g. goal), communication words (e.g. talk), positive emotions (e.g. happy), negative emotions (e.g. worthless), tentative words (e.g. perhaps), and certainty words (e.g. always). In addition, they studied the appearance of, amongst others, ‘grindstone traits’ (such as hard-working), ability traits, and standout adjectives (such as ‘great’ or ‘brilliant’).

The study showed that letters for male and female candidates were quite similar in terms of the various word counts, but that recommenders had used significantly more standout adjectives for male candidates as compared to female candidates, even though objective criteria showed no gender differences in qualifications.

See Hsieh & Shannon (2005) for a more in-depth discussion of qualitative content analysis, and see the short, pragmatic and very readable overview of the analysis of themes in qualitative analysis by Ryan & Bernard (2003).

7.3.1 Systematic case law analysis

Systematic case law analysis is a special kind of content analysis, of particular relevance to ELS scholars. On systematic case law analysis Hall & Wright (2008) state:

“Using this method, a scholar collects a set of document opinions on a particular subject, and systematically reads them, recording consistent features of each and drawing inferences about their use and meaning. This method comes naturally to legal scholars because it resembles the classic scholarly exercise of reading a collection of cases, finding common threads that link the opinions, and commenting on their significance. But content analysis is more than a better way to read cases. It brings the rigor of social science to our understanding of case law, creating a distinctively legal form of empiricism.”

In systematic case law analysis, one selects a sample (or an entire population) of document opinions or court rulings, reads and codes the material, and searches to answer the research questions. Very often the first-level codes have been derived in a dogmatic analysis, using standard legal research methods. The codes are then analysed in the second cycle using empirical, qualitative or quantitative methods, to answer the research questions. Systematic case law analysis is therefore also regarded as a hybrid, in between dogmatic and empirical analysis; in section 10.3 we present an example of case law analysis that focused on argumentation schemes employed by judges. However, in its design, systematic case law analysis follows the lines of a regular content analysis, with the raw data simply case law rulings.

Systematic case law analysis is a very promising type of empirical legal research. With the advent of digitalization, a wealth of case law has become easily available. Software packages facilitate the automated analysis of textified material. Some software enables the analysis of parts of texts or entire text documents.

Mascini & Holvast (2023) carried out a systematic case law analysis to investigate the relationship between the level of experience of law clerks and the confidence that is expressed in the writing style of their drafted judgments. Based on US research, concerns had been voiced that the delegation of drafting duties to law clerks “transforms the judge from a draftsman to an editor” (Posner, 1985, p. 104), which would render opinions and judgments less appealing and candid than if the judge had personally written them. Assuming that in many cases final judgments are based on edited versions of drafts made by law clerks, the hypothesis that the authors tested was that law clerks with more experience would write judgments with greater confidence than less experienced clerks.

Mascini and Holvast downloaded all administrative court cases published in the Netherlands in 2020 (N = 4,642) from the website of rechtspraak.nl, which publishes a selection of judgments. As they show, their sample, while large, amounted to not more than 5.9% of all administrative court decisions in that year. Their dependent variable was confidence, that is writing style of the judgments. The latter they operationalized as having three dimensions: judgment length, level of standardization, and number of legal references. They expected that experienced law clerks would draft shorter judgments that were less standardized and had fewer legal references. Given

the number of judgments, these three dimensions were extracted from the court files in automated fashion. Judgment length was measured by simply counting the number of unique characters in each judgment. Standardization was measured with machine learning techniques using Doc2Vec (Le & Mikolov, 2014), software that can calculate the similarity between entire documents. The number of legal references was measured by counting the number of references to case law and legislation – which could be easily retrieved from a linked data portal of the Dutch government – or in case these were references to previous ECLI cases, by using pattern matching. While the data are qualitative, the authors made use of automated extraction methods and software to quantify their dependent measures of interest. From other sources, the authors had obtained data on career length of clerks.

In their analyses, the authors controlled for a number of possible confounders. These were firstly judges' career length (supposing that more experienced judges might depend less on clerk drafts), complexity of the case (which might arguably affect judgment length as well as possibly standardization), and court level (expecting that judgments from first instance courts would be more elaborate than from appellate courts). They ran separate regression analyses, predicting each of the dimensions of writing style separately, in which they each time also included the other writing style dimensions, as these were correlated.

The authors showed that – in line with their hypothesis – law clerks with more experience co-signed judgments that were less standardized, and contained fewer legal references. However, the judgments were not shorter. The authors are not entirely sure to what this can be attributed, as they attempted to control for case complexity. The study shows, according to the authors, that it is important to replicate empirical legal research across various jurisdictions as the findings from one jurisdiction (USA) clearly may not hold elsewhere (the Netherlands). The authors recommend to replicate this research in still different areas of law as well as in jurisdictions that employ a scribe model of clerking, and to investigate the role of other factors such language proficiency to predict writing style characteristics. They also recommend to test the confidence hypothesis on a less heterogeneous set of cases, to reduce the risk of confounding with factors such as case complexity.

We give an extensive example of systematic case law analysis in section 10.3, to which we refer for more illustration.

7.4 Analysing other material

Qualitative research can use a wide variety of materials to analyse. Materials can be primary (when the researcher has gathered them him- or herself) or secondary. In the latter case, the material can be documents (such as discussed under content analysis), but news items can be analysed as well, as can movies, memes, ads or song lyrics. While the analysis of some of these materials may be slightly different, in the end some 'textification' of the materials is generally derived, after which the materials can be analysed by regular methods, using first-level and second-level coding. While there may therefore be slight differences with the steps as we discussed them here, the basic principles are the same.

This is less easily said when the materials have been gathered during participant observation. Here, it is likely that interviews and conversations, observations and text materials, video images and even music comprise the documents on which analysis should take place, but much of the information that will be relevant for information is likely stored in the researcher's head. It is then much harder to describe how analysis should proceed, what the steps are, and how one arrives at conclusions, as much of the material may be tied in with the researcher's experiences and perceptions.

Formulated differently, in such participant observation, the researcher him- or herself has become the data collection instrument, or even the data. While qualitative data analysis is already by nature contextualized, this is even more strongly so in the case of ethnographic research. It is then extremely difficult and even undesired to outline beforehand the steps to be done for the analysis of the data. While the descriptions of first-level and second-level coding were already quite dense, this becomes impossibly so for ethnographic analysis. Apart from generic principles, not many guidelines or specific procedures can be outlined. Especially for such data, it is immensely important to document precisely in analytic memos what steps were taken, what insights presented at what moments etc.

7.5 Reliability and validity in qualitative analysis

Qualitative methods come in many forms and shapes. Some are described with very little guidance: reading the data, over and over again, until you 'see' the themes. Others are more prescribed, such as when deductive codes are used, or when the occurrence of words is tallied in texts.

We stated in section 2.7 that, relatively speaking, quantitative studies tend to be stronger on reliability and qualitative studies may be stronger on validity, but that was perhaps a fairly vague statement, as there is such variation within qualitative research. Some of the studies that we touched upon, such as the studies using CCTV images by Lindegaard and her colleagues (2016), and systematic case law analysis are examples where reliability is in fact easily achievable because the material to be analysed is 'patient' and can be coded, the codebook refined, and re-coded until interrater reliability is reached.

One can assess interrater reliability by hand, inspecting the codes attached to textual units by different coders and decide whether these are identical or substantively different, both the textual units as well as the codes. The coding by the second coder should of course have been done 'blind', independently of the first coder. Software packages for qualitative data analysis also have options for automatically computing interrater reliability.

Such automated assessment of reliability is a meticulous business, however. Within ATLAS.ti one can compute Krippendorff's α (see section 2.6.1), but ATLAS.ti does so using exact text fragments. That is to say, if one compares one piece of text that has been labelled with code X and a second chunk of text that includes a few more words grouped under the same header X, the program will not identify this as a 100% match. The program works with exact 'textual units' and exact labels, and is merciless on textual overlap: it has to be exact. This means that for determining reliability of

interpretative codes one often resorts to doing so oneself, identifying and evaluating each other's attached codes and textual units.

When material is less easily accessed by others than the researcher him- or herself, such as when ethnographic or participatory methods are used, it is important that researchers themselves scrupulously check and re-check the quality of the data and their interpretation. As mentioned previously, an analytic memo is a must, in which the researcher notes all decisions taking during analysis. If it is not possible to compute quantitative measures such as the kappa-coefficient, qualitative researchers should assess the consistency of the fieldwork, whether analysis was conducted systematically and in a reproducible manner. The findings of such a scrutiny cannot be captured in a simple number but should constitute a convincing account.

Miller (2013) showed how much the outcome of qualitative research can depend on the person of the researcher, and one could say, in methodological terms, how difficult it is to achieve reliability. Using two cases (one a fictional description on what happened in a police station when a woman tried to report a crime, the other an actual description of gang violence at an American secondary school by a gang member) she asked 31 qualitative scholars working in criminology to each, independently, retrieve the most important 'themes' (i.e. the topics that are meaningful elements of the text) from the transcript. Miller found that there was not a single theme retrieved from the raw material that was brought up by more than 50% of the scholars. While Miller acknowledges that analysis of qualitative material may differ across scholars and that such divergence is not worrisome per se, she states that when scholars diverge so much on what can be seen as the rough data – the themes that are relevant – reliability may indeed be at stake (see section 2.7).

Validity is equally important and also – like in quantitative research! – not easy to prove with a simple check. Also here, the researcher should question him- or herself, and provide a convincing account of why the conclusions drawn from the research are valid. Lewis & Ritchie (2003) give a few examples: whether the sample was not biased, whether the observation or interview method allowed for a full view on the phenomenon under investigation, whether the categories and axes chosen reflect the meanings of the study participants and the meaning that is there in the material, and whether the interpretation and conclusion are supported by the data. The constant comparative method and deviant case analysis are tools to assert this. For more, see Lewis & Ritchie (2003).

7.6 Wrapping up

As we saw, the analysis of qualitative data can take many shapes. We gave an overview of the gamut: from hands-on coding of field notes, ethnographic observation, the coding of transcribed interviews or of CCTV images, to the analysis of textified material using software packages, to content analysis using automated data extraction and coding, on which quantitative analyses are performed. Qualitative data analysis is all in all a short label for a very broad set of analysis methods.

Qualitative analysis methods for ELS is also a rapidly developing field: the availability of software and digitalization of legal data has given content analysis and sys-

tematic case law analysis an enormous boost. For lack of space we left out exciting new methods such as network analysis, in which networks of verdicts or legal opinions are researched in order to uncover authoritative verdicts or changes in dogmatic reasoning. For more information on these developments, see Van Kuppevelt, van Dijck, & Schaper (2020), for a more technical comparison of such content analysis methods see Dyevre (2021), and for network analysis software, see [maastrichtlawtech.github.io/case-explorer-ui/#/](https://github.io/case-explorer-ui/#/).

7.7 Further reading

We have in this chapter only touched upon qualitative data analysis. There are many more different qualitative approaches, schools and types of analysis that can be employed than we described here. We refer to textbooks on qualitative research for more examples than we gave in our cursory overview of qualitative data analysis. Classics are Stern (1995) and Glaser & Strauss (1967). Chapters 8, 9 and 10 in Ritchie & Lewis (2010) give a readable overview of analysis methods. Webley (2010) discusses qualitative methods specifically for empirical legal research. An often referenced textbook on qualitative research in general is Denzin & Lincoln (2011), and a fairly practical one on ethnography is Hammersley & Atkinson (2007). A concise and accessible textbook is Tracy (2013). We already mentioned Friese (2019) as an accessible and hands-on manual for coding using ATLAS.ti; Saldana (2013) gives an elaborate and accessible overview of the coding process in all its facets.

Chapter questions

1. Argue whether the different modes of qualitative data collection have implications for data analysis (section 7.1.2)
2. Why and when is *in vivo* coding warranted? What are its disadvantages? (section 7.2.1)
3. What requirements apply to the final stages of the first-level coding phase? (section 7.2.1)
4. What are analytic memos intended to contribute to? (section 7.2.2)
5. Argue why content analysis is often regarded as a hybrid of qualitative and quantitative research (section 7.3)
6. Give two aims of the use of deviant case analysis (section 7.1.1 and section 7.5)
7. Describe whether reliability and validity are assessed differently in qualitative and quantitative research (section 7.5)
8. Argue why systematic case law analysis is also regarded as a hybrid between dogmatic and empirical analysis (section 7.3.1)