

## Chapter 5

### How to conduct a qualitative analysis

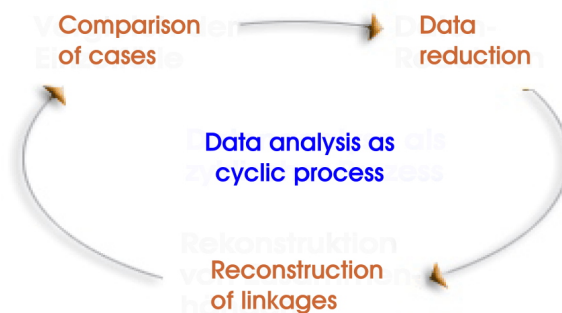
In this chapter we try to outline typical phases in the process of a qualitative text analysis. Particular analytical steps are marked as characteristics of these phases, although concrete processes of qualitative analysis usually follow a cyclic path, i.e., during every phase a researcher may be engaged part of the time in activities characterizing other phases of analysis. In addition, this chapter tries to give an overview on the usage of computer assistance for qualitative analysis as provided by AQUAD. Details are explained in later chapters. However, the following chapters of this manual cannot offer more than a general introduction to methodological approaches to qualitative analyses and possible contributions of computers. We suggest that users who want to learn more about underlying principles read the following books:

- General contributions of computers and software tools to qualitative research are presented in Tesch (1990), Kelle (1995), Fielding and Lee (1998), and Lissmann (2001).
- Miles and Huberman (2nd edition, 1994) give a very detailed introduction to interpretational analysis of qualitative data, based on numerous examples. Concrete examples from a variety of qualitative studies in educational research can be found in Bos and Tarnai (1998) or Schratz (1993) or Schratz (1993), from the field of psychological studies in Kiegelmann (2001, 2002, 2003).
- An introduction to techniques of theory-building in qualitative analysis, based on the approach of "grounded theory", is offered by Strauss and Corbin (1990).

This list is far from being exhaustive, but names some selected books which seem appropriate for beginners in qualitative analysis and which demonstrated their usefulness in Huber's seminars at the University of Tübingen. All of these books also offer excellent approach to more specialized literature. In the following, this chapter borrows from an introduction by Huber (1992) which unfortunately is not available in English.

### 5.1 Which are the steps in qualitative analysis?

As typical phases of qualitative analyses of data we can distinguish the reduction of the original data base, the reconstruction of linkages, and the comparison of findings. Especially in psychological studies researchers often are interested in inferring inductively from *one* subject's data regularities in this person's experiences and behaviors. Whether communalities can be found when taking into account data of several persons is interesting, too, but only during a later stage of investigation.



- The first phase of qualitative analysis is characterized by *reducing* the overwhelming amount of data (texts, sound or video recordings, graphic files) by identifying the content of more or less encompassing data segments. A "code" as abbreviation or name is attached to each segment. In the following, these codes are used as representatives of data segments or "units of meaning" in the data. Fundamentally, this is a process of categorization, where the categories may emerge during data interpretation or may be taken from an already existing category system – depending on the researcher's epistemological orientation.
- During the second phase researchers try to *reconstruct* the data producer's subjective meaning system from the units of meaning in their data. By "data producer" we refer to the researcher's interview partners, to writers of diaries, to observers who took field notes in a setting, the child playing with an object in a video, etc. In order to reconstruct meaning systems we are looking for regular linkages between units of meaning in the data, which are characteristic for a person and/or her situation.
- In the third phase finally researchers try to infer invariants or general communalities by *comparing* individual systems of meaning or "cases" (see Ragin, 1987).

It is important to keep in mind that these phases neither are strictly demarcated nor do they follow each other in a linear sequence, but they overlap and are linked to each other in circular patterns (cf. Shelly & Sibert, 1992). During data reduction we may start to ponder about a person's implicit theory or we may permanently compare the data at hand with other data which we have analyzed earlier. Thus we may perhaps detect in person C's data an aspect of meaning which we overlooked in person A's data. As a consequence, we repeat the process of data reduction for person A. In all of these phases it is necessary to affirm deductively the validity of our generalizations. That is, we try to infer particularities from our general findings and then return to our data and try to find evidence in form of specific information, i.e., statements in the texts or recordings, sequences of action in the video, etc.

## 5.2 How to reduce qualitative data

The principles of reduction are obviously simple, but their application soon proves to demand very much work, to consume very much time, and to be very prone to errors. Voluminous verbal materials have to be reduced to the units or categories of meanings they contain. Tesch (1992) describes these principles as retrieving and marking of text segments, which are relevant for the question under study, by an abbreviation, that is a code for the particular category of meaning. Tesch also compared computer assisted analysis and "traditional" approaches, which use segments of texts or text clippings in the literal sense or which transfer relevant text segments to index cards. The work load is tremendous in both cases. Computer software not only assists in reducing the amount of data, but also in reducing all the mechanical labor otherwise necessary. Instead of handling verbose clippings of text distributed over many piles of index cards, further computer assisted analyses use just the codes of these text segments; that is, after data reduction you work with category names and information where to find the categorized text segments in your texts. If you need to scrutinize the original text segment again during the course of your work, the computer will retrieve it immediately for you.

If the original data, however, are not available in form of text files, but come as hand-written diaries, sound- or video-recordings of interviews, graphic notes or drawings, etc. we had to transcribe these data until recently, that is convert them into text files. Particularly in these cases AQUAD 6 helps to save a lot of time and money, because it allows to reduce directly all these types of data to codes – without a detour via transcriptions. Of course, you may still transcribe critical parts of your data files for (verbal) publication in later reports, articles, etc. There is a small text processor within the "Coding" components of AQUAD, and the "Memo" function (see chap. 8) also offers the possibility of transcribing multimedia data.

In case of any type of original data, the critical question in this phase of analysis is: *How and where do I find units of meaning in my data?* Beginners in qualitative analysis as well as experts who try to get familiar with a new content domain ask this question again and again. Weber (1985) describes six widely used general possibilities to define text segments, namely to choose as unit of analysis single words, meanings of words, sentences, topics, paragraphs, and the complete text (for instance, if the texts are short as in the case of letters to the editor or if you want to produce head lines or abstracts). However, this choice cannot be made mechanically, but it needs itself tentative qualitative decisions. Not so obvious when single words are used as units of analysis, but quite obvious when we use more complex alternatives like word meanings, sentences, typical sequences in a video, etc. we need preceding insights or hypotheses that the unit we have chosen will contribute to answer our research question. Additional qualitative decisions are necessary, for instance which words are used synonymously or which idiomatic expressions have similar meanings for the writers/speakers in case of verbal data.

The strategies of defining units of meaning describe an important difference between quantitative and qualitative approaches to text analysis since the beginnings of a broader reception of text analysis in the social sciences. In the same year, 1952, two trendsetting articles were published by Berelson and by Kracauer. While text analysis serves for Berelson to assess systematically, objectively, and quantitatively the *manifest* contents of communication, qualitative content analysis according to Kracauer tries to reveal the categories of meaning hidden or *latent* in the data. Both authors relate their controversial positions to a debate, which Thomas & Znaniecki (1918) had initiated 35 years earlier with a meanwhile classical analysis of letters of Polish immigrants to the United States. Obviously the question of adequate units of analysis – here words vs. meanings – is confounded with the particular goal of text analysis.

On the level of processing both approaches do not necessarily exclude each other. Especially with assistance of computers we can apply various strategies from the quantitative approach to qualitative text analysis in order to gain support for our interpretational endeavors.

Principally we should keep in mind to develop the analytic units *during* data interpretation. This is true in analyses which try to *understand* the experiences and actions of people from their own verbal descriptions as well as in analyses which try to *explain* specific actions by relating them to the frame of reference of these people's implicit theories. Only this approach helps to open a door to the subjective world views of our interviewees or producers of other types of data. Otherwise, we would be in danger of grasping only some partial aspects of their world views, maybe isolated from their subjective context, which our personal analytic grid is able to comb out. The strategy of developing categories "on the fly" corresponds to the approach of

"grounded theory", an empirically based procedure of generating theories recommended by Glaser & Strauss (1979).

However, this procedure demands enormous work as soon as more than two or three data files are to be analyzed. Because we usually want to compare the results from single cases in advanced stages of analysis, we have to ensure that we defined and coded the units of meaning in all of the cases coherently. Usually this process has to be repeated again and again, and in this process units of meaning and their codes have to be modified. Miles & Huberman (1984; 1994) suggested a compromise, which structures the process of data reduction from the beginning: Before you start analyzing your data files you state a very general *frame of orientation* without any references to particular contents; then you try to find specific units of meaning within this framework. This structuring does not contradict the demand of openness for emerging categories while approaching your data directly, because usually the reduction of potential data by structuring starts even before you enter the phase of coding, for instance when planning and deciding which persons, cases, sites, types of data, etc. should be included in a study.

### 5.3 Scheme of the process of data analysis

(1) In the beginning it is indispensable to become first of all acquainted with the subjects' perspective of the problem under study – above all, if we were not present personally in the phase of data collection, that is, if we did not talk as interviewers with our research partners or did not run the video observation, etc. Or expressed the other way round: We should not try to establish a differentiated system of categories already while reading, watching, listening to the first data set. For sure, this system would be valid only for the first case, and we would have to revise it fundamentally as soon as we get a closer view on the second case.

Supposing you work in your project with a larger number of data sets, we recommend instead that you select a few cases to begin with either by chance or according to the principle of "theoretical sampling" in single case studies (cf. Yin, 1992) from all of your data files. For instance, if we were going to analyze the teaching of mathematics in different schools, we would have to consider possible differences depending on age of the students, school type, and teachers' professional experience. According to these criteria we would select a few data sets, for instance, video recordings of math lessons, and elaborate a provisional frame of orientation. Maybe we video-taped a priori in classrooms, which were known for different approaches to teaching and learning mathematics. Of course, we would then base our first orientation (also) on this criterion. Obviously this first step modifies a recommendation by Miles and Huberman (1994; see above) to decide about a general, not content-specific frame of orientation already before we start to read the data texts or watch/listen to any recordings. Instead, we suggest to draw a sample of data sets/cases to test and differentiate any frame of orientation, which may already be available, at least in outline, just because we have a research question.

(2) If we decide to select a sample of data sets according to the principle of "theoretical sampling", we are forced to give the matter of general characteristics of our cases considerable thought. In other words, we have to ruminate on possible determinants of the "profile" of various cases. In the example of classroom observations we would think about the age of students, school level, professional experience of teachers, number of students, didactical orientations of teachers, etc. It would be a good idea to note these characteristics in a memo (see chap. 8) and to apply them later as "profile codes" (see chap. 6.1) or "singular codes" (which are attached only once) to characterize the data set in general. If we had selected some data sets, however, by chance we could now consider in advance, which characteristics of persons and/or situations involved in our study may become relevant as "profile features" later during more detailed analyses.

(3) We have a look at the selected data sets without getting involved in sophisticated codifications, but try instead to understand the content as a whole, differentiate essential parts, find out about what the cases have in common and what makes them different. Thus we mark out gradually a frame of orientation. A decisive aspect during this step is to note all preliminary interpretations and ideas in memos (see chap. 8), including at the end a short (!) thematic summary, in which we note our "first impression" of the general meaning of each data set, and what we conceive of as its central aspect at this early point in the analytic process.

(4) What follows then is a critical decision about the general strategy of interpretation we want to follow: Should we try – as already initiated in step 3 by writing a thematic summary – to find more general units of meaning, mark them by codes, and differentiate these units stepwise and in repeated sessions? Or should we just in the beginning pay attention to *all* the details, maybe somewhat restricted by our general frame of reference? In this case we would try in further sessions to summarize many of these detailed codes in more abstract categories, thus also making the cases better comparable. The first strategy moves from general to particular considerations ("differentiation"), while the second strategy starts with particular aspects and leads to general insights ("generalization").

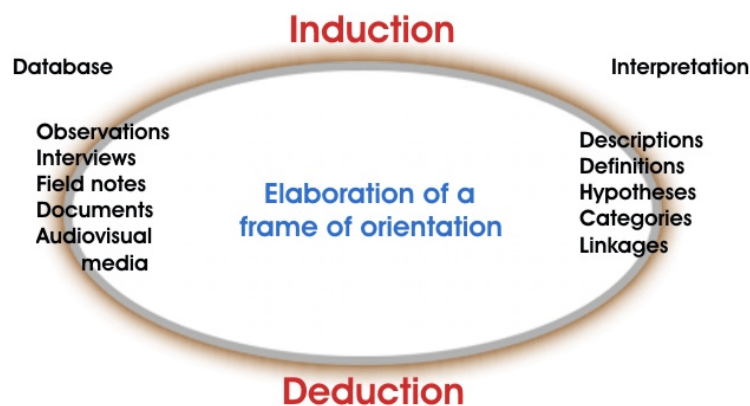
Which strategy we should prefer depends, of course, on the research question, but also on the researcher's experience with his/her topic. Principally, we want to advise beginners *against* the strategy of generalization, which urges them to concentrate on tiny specifics in a text, picture, or a recording and broaden their perspective only later, when they try to develop more abstract, general categories. Thus beginners prevent mostly successfully any deeper, at the beginning uncertain and difficult theory-based analytic considerations, but run the risk of indulging in superficial hustle and bustle. Novices in coding use to be content in this situation, because they feel busy, they produce codes – and notice only much later that their activities were

not necessarily goal-directed. In one of my seminars, for instance, a group of students generated codes for – in their opinion – keywords in a text, basically doubling the data instead of reducing them. In another case, a novice of qualitative content analysis developed about 1500 (no typing error!) codes for his interviews until he realized that he was unable to see the wood behind all of these trees.

Therefore we judge it to be most promising, if above all novices take the principle of "permanent comparisons" very seriously, which is basic within the approach of grounded theory (Glaser & Strauss, 1979). Permanent comparison demands to look within and later across all data sets immediately after introducing a category for congruent and incongruent examples. This orientation prevents to a large extent that the qualitative analyst gets lost in an overwhelming number of interesting details, but tries to elaborate insight into the structure of meaning in his/her data. The following table contrasts both strategies:

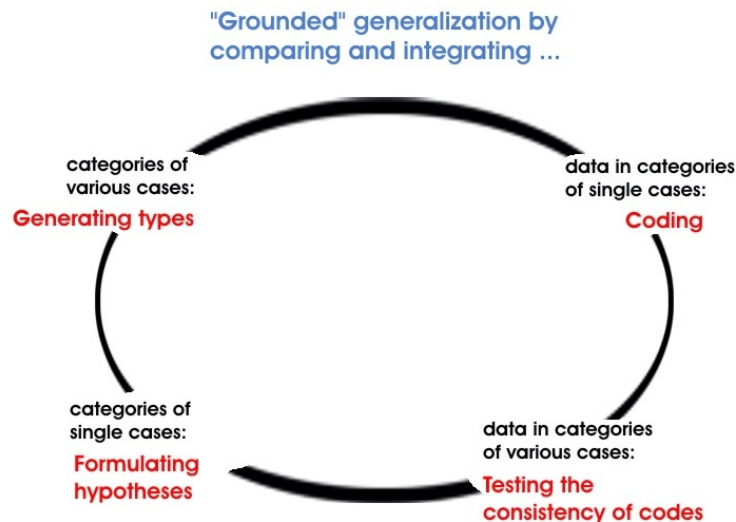
Strategy:	Differentiation	Generalization
Coding starts with	Search for general categories	Search for specific aspects
Further steps	Differentiation with the aim of uncovering specific differences	Generalization with the aim of finding common grounds

Both strategies are not mutually exclusive, but are on the contrary mutually dependent. The interpretation of qualitative data is not a linear, but a cyclic process:



Based on ideas of Dewey, Shelly and Sibert (1992) described the relation of inductive and deductive thinking in qualitative analysis as illustrated in the figure above. From this point of view we should object to the recommendation for novices (see above: start data interpretation from general categories) that there is only a chance to find something new within the data, if we interpret small data segments. In text files this means to code the data more or less line by line. For sure, the higher probability to introduce new ideas by inductive reasoning is counterbalanced by greater danger to miss a general and essential aspect. If you follow our recommendation to approach your data with a strategy of gradual differentiation, the accent lies on the side of "Interpretation" in the figure above. Of course, interpretations and interpretive knowledge may have developed already inductively during concrete experiences in the field or sifting through the data files, for instance in form of some very general categories (example from a study in a kindergarten: "The nursery school teacher directs the kids' attention;" "She supports them in case of difficulties;" "She makes them verbalize their experiences;" etc.) or working hypotheses, which already influenced the formulation of research questions. Necessarily the database will be differentiated as soon as we analyze the data from different

persons who were observed or who talked to us in different situations. That is, if we apply the strategy of generalization we begin to analyze our data with specific meanings, which emerge from within the data. What we need then are signs that show the direction from specific interpretations to general interpretive knowledge. "Permanent comparison," the basic principle of grounded theory was already characterized as guideline for generalizing. These comparisons follow a cyclic pattern, too, as Shelly and Sibert (1992) have underlined. Following their scheme of procedures, we can describe permanent comparisons and their goals during the progress of qualitative analysis as depicted in the figure below:



(5) After the principal strategic decision we start to code a sample of data files (see step 1). Necessarily we will again and again switch from comparing and integrating single data within one file to comparing a category across several files – and back again. That is, we move forward in smaller, less encompassing circles within the general analytic cycle, first of all to test whether we apply our categories consistently.

(6) In this way we develop a set of coding rules, which we finally apply to all data files available in our project. Of course, we must not wonder, if we come upon exceptions and even contradictions, which make us modify the coding rules. In some cases it may even be necessary to leave the cycle of data reduction/-interpretation and enter once more the phase of data collection.

(7) During the whole process there will be many occasions and good reasons to write memos. When you come now to the point, where you try systematically to formulate hypotheses about your subjects' world views or to generate types, you will be grateful to find your initial ideas and assumptions from earlier stages of analysis again. In this stage we concentrate our efforts on elucidating relations or "linkages" between critical categories (see chap. 11). This was described in step 4 as integration of categories in "hypotheses", for example: "If person P observes the events A or B, s/he reacts by doing X." Or: "If teacher T talks about her/his students' lack of motivation, s/he continues by complaining about the influence of TV and the role of parents."

Experiences in this phase, for instance detecting contradictory interpretations, may be a reason for an analytic loop back to the phase of coding and testing consistency (see step 4). Also some additional data collecting, maybe for some selected cases and/or with modified questions or observation methods may turn out to be helpful.

---

(8) Given a sufficient number of cases, we may finally try to group these cases into types, based on comparison and integration of (theoretically selected) categories. As result we achieve a differentiation of cases according to typical configurations of characteristics (see chap. 13).

(9) Finally, we must not fail to point out something that is relevant for the process of qualitative analysis as a whole: Qualitative analysis has something to do with the "quality" of events and states captured in our data. Therefore, our codes should not represent neutrally the fact of specific events or states, but have to inform also about their quality – or expressed bluntly: Were these events good or bad? For instance, we would not be able to formulate hypotheses or generate types in a biographic study on problems of youth, if we marked those interview segments, in which young people talked about their parents and added simply the code "parents." Can you imagine young people who do not talk about their parents in a biographic interview?

The big question is: *How* do they talk about their parents, family climate, social relations, etc.? A characteristic or a variable, which appears without any variation in all cases, does not contribute to "permanent comparisons" and is therefore not suitable for differentiations or explanations that would get us anywhere. Consequently we have to qualify statements like those about parents. In the beginning rough grades like positive/indifferent/negative (or abbreviated as appendix to a code: +/0/-) will be sufficient for a search of differences and common grounds.

From an interpersonal point of view it is a most pleasant trait of many novices in qualitative analysis to refrain from jumping to evaluations and qualifications particularly of personal issues. On the other hand, avoiding what is meant to be rash attributions of quality ("positive", "negative") misses the purpose of qualitative analysis! At least during a test of consistency of coding all "neutral", non-qualifying codes should be checked, whether the attributed data segments need a repetition of earlier coding steps with the goal to make sure that these codes depict differences – if there are any – in the meanings of the data segments.

## 5.4 How to find units of meaning

When looking for units of meaning in the data files, you should be open for emerging structures in your data. However, recipes enforcing structures on the data may often be welcome because they help to shorten a phase of maximal uncertainty particularly during the first steps of data analysis. On the other hand, the price for certainty and savings of time and effort may be too high: concrete guidelines would lead only to those units of meaning which could be foreseen, while surprising, rare, but maybe most interesting aspects are in jeopardy of being omitted from further analysis. The following hints should be understood as heuristics, that is, as general directions which provide assistance in identifying units of meaning and attaching appropriate codes, but not as algorithmic rules which could be followed step by step to a predetermined goal. Three heuristics will be described in the following. AQUAD offers support particularly for the first two of them and their variations, which are also outlined.

- When looking for *categories* we read a text, listen to a recording or watch a video and try to be sensitive to emerging concepts, to statements about situations, events or persons, to opinions, ideas, etc., which can be attributed to a general, super-ordinate category.
- When looking for *sequences* we pay attention to statements of linkages, connections, relations, etc., expressed in the data, and we try to put these subjectively linked statements together in a unit of meaning – which will be more comprehensive than the results of categorical data reduction.
- When looking for *themes* or topics we have to go to the most abstract level; in some cases a complete data set is reduced to its topic, for instance, when we reduce short texts like "letters to the editor" etc.

### 5.4.1 How to find categorical codes

In addition to looking for units of meaning in the data, we have to decide, whether our categories and the corresponding codes should describe, interpret or explain the content of a particular data segment (see Miles & Huberman, 1994, p. 56). This is a decision for which we cannot expect computer assistance. Here, the computer assists "only" in documenting, in sorting, and in revising – if necessary – all the decisions a researcher has made on his/her own responsibility. Then, three possibilities are widely used to find categorical codes:

#### *Applying pre-determined category systems*

If a qualitative study does *not* aim at constructing theories from concepts emerging in the data, there is a very simple means of finding codes: You can use an available category system and reduce your data according to the interpretational schemata contained in the system chosen. Categorical systems may be "available" from earlier studies on the same topic or from publications of other authors.

When using a pre-determined category system we have to decide which segments of a given data set correspond to which of the given category definitions. Within AQUAD we have then to note where each of these segments is located in our data and which codes apply to them. How this is done will be explained in detail in chapter 6. Without fail we will have trouble assigning specific data segments to one of the given categories, and we will not always use the codes consistently. In these cases, the function to *retrieve coded data segments* is very helpful to control our work. After entering a critical code (or a whole list of these codes), the program shows us very quickly all data segments in all data sets analyzed, which up to now were marked with this code.

However, this approach to test the reliability of our coding does lead us only to those data segments which were assigned erroneously to a particular category, but not those segments which belong to the definition of this category but were *not* assigned to this category. In conventional approaches to data analysis we would detect this type of error when we notice inconsistencies within the sets of segments assigned to other categories. With computer assistance it is no problem to control for "missing" data segments by checking all segments which were assigned to related and thus "error prone" categories. In addition, we can make use of

the complementary relations of manifest and latent units of meaning in our data – however, only if these data are texts (!). Three strategies appear to be useful:

- We define *key words* as manifest indicators of a critical meaning and have the computer retrieve all their occurrences in the texts. If we find our key words in a specific text segment, we can decide whether this segment should not better be assigned to the category indicated by this key word. For this strategy most text processors can be used.
- We assemble key words in a *dictionary for analysis*, called a *word catalog* in AQUAD, that is we use a list of key words for the retrieval of critical text segments. Then we find the information needed in a single run of the program.
- AQUAD offers both of these possibilities, automatically combined with a third strategy called *key-words-in-context* (Popko 1980). This function tries to retrieve one or more key words in our texts and prints it within the context of the line where it was found in a text. The researcher is responsible then for further decisions, for instance about the range of a corresponding unit of meaning and about marking it with a particular code.

### *Hypothesis-based categorization*

The research question often supplies the researcher with hypotheses that may be used as a sort of guidance when looking for units of meaning in the data. Based on these hypotheses the researcher tries to define categories and rules of coding for her/his data.

For instance, the approach of Marcelo (see p. 18) to code his interviews with beginning teachers was hypothesis based. A theoretical model of professional socialization served as a framework to design a preliminary category system. With increasing familiarity with the texts and deepened insights into the subjective points of view of the beginning teachers, some of the categories had to be omitted from the system as inadequate, while some new categories emerged from the analytic process. Some "narrow" categories could be combined to more comprehensive units in terms of the teachers' subjective theories (see Huber & Marcelo, 1992). Of course, new or combined categories had to be compared to critical text segments in all interviews again.

Properly, the researcher decides about concrete categories already when designing the collection of data. Thus, for instance, the formulation of guiding questions for an interview would be deductively determined by the researcher's basic hypotheses. On the other hand, the researcher will for sure experience interview situations or find data segments in the interview transcriptions or recordings which cannot be assigned to pre-determined categories. For these data segments categories have to be developed inductively. The new categories may relate the data to the original hypotheses – but they may also be a stepping stone from which modifications of the original frame of reference can be realized. We see that the researcher's degrees of freedom as well as challenges to her/his interpretational aptitudes are increasing when applying this second strategy. Most important, the decreasing structuredness of this approach augments the chances to take into account the subjects' points of view.

When applying this strategy, the researcher has also to meet the demands of the "method of permanent comparison", which is a trade mark of the approach of grounded theory (Glaser & Strauss, 1967, 1979; Strauss & Corbin, 1990). Shelly & Sibert (1992) described this method in detail and related it to models of the researcher's cognitive activities. An activity of major importance in "permanent comparison" is to test every inductive conclusion from particular data to more general principles – here the analytical categories – by means of deductive conclusions, that is deductions of specific units of meaning in the data base.

Together with the demand to apply interpretation in data analysis the importance of computer assistance is increasing. When we try to get an overview on linkages between categories, simple retrieval functions reach their limits. From this point in data analysis on, AQUAD serves you especially well, because it supplies you with routines for deductive conclusions based on the principles of logic programming (Tesch 1990; Shelly & Sibert 1992).

- The functions for *retrieval of coded data segments* and for retrieval of *key words* (including retrieval dictionaries and the KWIC-function) may be used here to check the consistency of coding within and across data sets in the same way as in studies applying ready-made category systems (see above).
- From the hypotheses, which mark out a framework for the development of categories, we get hints to specific *relations of categories*. AQUAD supports hypothesis-based data reduction with functions for testing relations of categories. Among other relations you can find out about the *super-ordination/sub-ordination* of categories, *sequences* of categories or *clusters* of particular categories. These three types of relations probably represent the most frequently tested relational patterns tested in data analyses. The test activities demand that you focus your attention not only on one category and the data segments it represents, but on two or more categories and the defined relations between them -- which may include negations! Of course, non-events or missing relations have to be registered in this process, too.

### ***Categories by theory-building***

The most exacting mode of reducing qualitative data refrains from any prescriptions for data reduction as for instance category systems and from structuring the process of qualitative analysis by hypothetical frameworks. Consequently the researcher has to keep in mind his/her subjective experiences, opinions or prejudices as regards the world views or behavior of his/her subjects, and the researcher has to avoid premature stabilization of emerging principles of data reduction by permanently comparing the momentary categories and relevant statements in all available data files. To be sensible of one's own way of reading a text or interpreting a video is particularly important if data are analyzed which were written or recorded long ago or in a context that differs markedly from the researcher's way of life (Fischer, 1982). Often specific information about the speaker or writer of a text is very helpful. If this information is not available in a text, it may be necessary to look for additional sources. These sources and their information help to approach better the goal of viewing the world of the subjects of a study through their own eyes and to understand it from their own perspectives (or, for more auditive readers: these sources may support listening to the speakers' own voices in their texts).

In this process we try not only to describe subjective world views, but to order them by matching concepts and to reconstruct systematic relations between these concepts, that is, we are occupied with what Glaser & Strauss (1967; 1979) called the *discovery* of a "grounded theory". The term "discovery" accentuates an essential difference from methodological approaches which are applied to confirm given theories. Grounded theories are developed during content analysis, therefore we do not start from an available theory looking for verifying or falsifying data in the data, but we start from a phenomenon the data refer to, which we want to understand and to explain (Strauss & Corbin, 1990).

This approach demands a maximum of interpretational efforts. Exactly this aspect seems to cause problems for beginners of text analysis. Pre-defined category systems come with explicit interpretation rules or they transport these rules implicitly in a set of examples for each category. Researchers using a hypothesis-based approach to categorization can at least rely on a general orientation what to look for in the data. In the process of theory construction, however, researchers have to find out first what a data file is telling them. Beginners often tend to avoid the risk of errors and to interpret as parsimoniously as possible. In extreme cases this tendency results, for instance, in reading a text for the appearances of critical formulations – comparable to the application of key words when using ready-made category systems. If text segments containing such a critical formulation are marked by a specific code, this code does not necessarily represent a "unit of meaning" or signify a subjective viewpoint of the writer/speaker, but more probably the mere fact, that a particular formulation was used in this text segment. Its meaning still may be unclear.

This tendency is especially pronounced, if a researcher does not know about the possibilities of tentative interpretation and easy revisions in computer-assisted qualitative analysis. Supported by AQUAD a researcher is not punished for "playing" with ideas by tremendous labor when it comes to revisions of codes during a later phase of analysis, for instance, when the researcher finds data segments not matching a category introduced earlier. On the contrary, computer assistance encourages creative interpretation, because we have to introduce codes as markers of data segments from the very beginning of an analysis, and changes, summaries, differentiations, etc., of categories can be realized smoothly (Tesch, 1992).

As a *rule of thumb* for theory construction by categorization we can state (cf. chap. 5.3): You should look in the data for units of meaning *as large as possible* in order to find something to interpret at all; at the same time these units should be *as small as necessary* to avoid representing incongruent contents by the same code. This approach could be called a *strategy of differentiation*. AQUAD supports this strategy, because units of meaning can be defined without limitations; above all, units of meaning may overlap, and the same data segment may be assigned to several categories, that is, it may be marked by several codes.

If you are already familiar both with qualitative analysis and with the domain addressed in the data of your study, you may reverse this method and approach your data following a *strategy of generalization*. These prerequisites given, firstly the probability is low that you fall a victim of details in the data and miss to reveal its essential meanings. Secondly you are then also accustomed to software functions for producing meta-codes, which assemble a number of too detailed categories into one more comprehensive category. In a generalizing approach it is particularly important to compare coded data segments permanently in order to find out inductively about all content dimensions expressed in these segments and the adequate super-ordinate categories. AQUAD supports this strategy, too. You can retrieve easily those data segments which are similar to a particularly coded one, which contradict its meaning, which depend on it or are related in other, specified ways to this data segment. Based on findings about such relations you can then try to combine individual codes into a more comprehensive one.

When we reduce data to categorical codes, we try to distinguish data segments in such a way that we can assign them to well defined, mutually exclusive categories. Unequivocal assignment of meanings to categories does not imply, however, that the text segments involved always have to be clearly different from each other. Depending on writing styles or communicative styles of the data producers or depending on the researchers' ways of interpreting their data, segments assigned to differing categories may overlap or the same data segment may even become assigned to more than one category.

#### 5.4.2 How to find sequential codes

The starting point for sequential coding is the detection of specific relations between data segments. If specific linkages of segments emerge from the data, a researcher may want to mark their appearance in a data file and to represent the type of linkage by a particular code – in the same way as the researcher used categorical code. Here, however, the code represents a defined sequence of meanings found in a data file. The unit of meaning is the complete data section, in which the sequence of meanings was found.

Which strategies are at hand that may help us to go beyond assigning categories or subcategories to data segments and to detect sequences or linkages of meaning in a data file? In the following we differentiate between strategies that inquire into simple sequences and complex sequences of meaning.

##### *Looking for simple sequences*

Besides looking for *hierarchical sequences* of super- or subordinate categories as recommended by Strauss & Corbin (1990) for textual data, a number of other simple sequences appear to be interesting from the point of view of grammatical or linguistic properties of the text. When using these strategies, we should be aware, however, that we may be about to abandon looking for emerging categories and begin to enforce categories on the text (see Glaser, 1992). Whether this switch is permissible and which of the possible strategies is adequate cannot be answered absolutely, but depends on the research question. Often researchers are looking for *causal sequences* (using key words like "because", "based upon", etc.), *temporal sequences* ("while", "then", "before", etc.), *concessive sequences* (positive concessions like "not ... but ..." or negative concessions like "indeed ... otherwise ..."), *conditional sequences* ("if ... then..."), *final sequences* ("so that", "lest", "in order to", etc.), *comparative sequences*, *modal sequences*, and *defining sequences*.

Neither this list of types of sequences nor the quoted key words claim to be complete. They only want to stimulate your own trials to retrieve sequences of meaning in the texts according to the researcher's questions. In this context we should again accentuate the function of computers and software as useful *tools* for text interpretation, but not as agents of qualitative analysis. Functions for word retrieval may be especially helpful, but within very narrow limits. Except in cases where a researcher analyses carefully formulated texts,

the yield of dictionaries of particular sequences of meaning is usually limited. In freely spoken interview texts even sentence structures are often hard to define: Where is the principal clause, where does a subordinate clause begin? Often conjunctions critical for defining a specific type of sequence may have been in the mind of a speaker, but they do not appear in the spoken and transcribed text. In addition, many speakers do not use critical conjunctions or constructions according to the grammatical rules. Therefore, looking for key words or applying whole dictionaries of these words never substitutes for empathetic interpretation of texts – but these strategies may be useful heuristics.

### *Looking for complex patterns of sequences*

Structural or content characteristics of data may inspire the search for complex sequences of meaning. For example, when a researcher tries to reconstruct theories of action implicit in a text, audio- or videotape he or she could look for sequences of appraisals of situations, reflections of alternatives for action, expectations of action effects, and evaluations of potential personal consequences like satisfaction or disappointment. This means, the researcher would have to keep in mind a number of categories and to look for them and their proper sequence simultaneously when interpreting a data set. Such an analysis is both demanding and prone to errors. The variety of structures of possible linkages of categories is demanding for software, too. It would not make much sense to implement all sequential combinations of possibly linked categories in form of abstract deductive algorithms into a program, so that the researcher would have to enter during run-time only some concrete codes as variables. For the retrieval of complex patterns a researcher needs access to the source code of her/his software in order to add those rules with minimal programming effort, which s/he expects to apply to the linkage of categories in the data. AQUAD offers exactly this possibility.

### 5.4.3 How to find thematic codes

The most radical approach to data reduction is applied, if we try to code a data file just by one central category, that is, its message or its topic. Since qualitative analysis is a cyclic process, we cannot locate the strategy of thematic coding at a particular place within the interpretational process, let's say at the end as a sort of summary of findings. Thematic coding may play an important role at this stage of content analysis, but this strategy is also useful at other stages:

- When we analyze relatively short and homogenous data sets (for instance, "letters to the editor") or paragraphs in a text, thematic coding may be all we need or want to reach our interpretative goals.
- In the case of large, heterogeneous, and complicated data sets interpretation may start with thematic reduction. Here, this strategy may serve us as an important heuristic. It may help us to find one or some main ideas in the data and not get entangled in a great number of potentially controversial details.
- Usually, thematic coding appears in the final stage of data analysis, often only after several cycles of data reduction, reconstruction and comparison of meaning structures during which the "Leitmotiv" of different data sets was elaborated more and more clearly (see Shelly & Sibert 1992; Strauss & Corbin 1990).

Strauss & Corbin (1990) recommend five steps for thematic reduction of data (in their case: texts), which are compatible with each of the three locations of thematic reduction:

- (1) We start with identifying a main idea or a central category;
- (2) then we look for subordinate categories,
- (3) which often need to be differentiated or to be linked with each other;
- (4) then we examine hypothetical relations among sub-categories as well as between these categories and the main idea by comparing data segments;
- (5) discrepancies, for instance, inconsistent categories stimulate further cycles of analysis.

## 5.5 How to reconstruct systems of meaning

In theory-constructing qualitative analysis we try to generate the speaker's or writer's theory of the issue her/his data file is talking about, that is, we try to reconstruct the speaker's or writer's subjective system of meaning. Solutions can be developed inductively, deductively or by combining inductive and deductive strategies. Which approach is to be preferred depends on the research question.

If we analyze transcriptions of minimally structured interviews or texts, for instance entries in a diary, we usually start by identifying units of meaning and assigning them to specific categories. Then we may look for typical sequences of categories. Finally we will try to subsume such sequences under more abstract categories, that is the message or the topics of the data producer. This analysis starts with inductive processes, but alternates between inductive and deductive conclusions at least from the stage of thematic reduction on (cf. chap.5.3).

If we approach our data under a particular theoretical orientation or from the point of view of particular interests, we will try to retrieve specified relations from the beginning. That is, we start with hypotheses of possible relations and try to confirm them deductively in the data. Inconsistencies or contradictions then will cause processes of inductive analysis, which in turn may modify our hypothetical orientation.

In inductive approaches we try to generalize categories and their systematic linkages from concrete data segments. In deductive approaches we try to find concrete data segments which may confirm general assumptions or hypotheses about how specific categories are linked. The following overview on strategies offered in AQUAD for computer-assisted reconstruction of systematic linkages is structured according to these two approaches.

### 5.5.1 Reconstruction from the context of particular data

We are looking for systematic occurrences of a particular category (or several particular categories) in a data set which is characterized by one or more profile codes, usually socio-demographic codes. For computer-assisted reconstruction we create code matrices (Miles & Huberman 1984) or code tables (Shelly & Sibert 1992). The contents (that is, data segments) in the cells of such a table are determined twice, both by the category serving as column header (profile code) and by the category serving as row header. Thus, the interpretation of findings also is much more guided than in the case of just registering frequent closeness of otherwise "unconditioned" categories in space (line numbers, frame numbers) and time (sequence of data production). On the other hand, the construction of a table for analysis demands much more conceptual investments, that is greater progress in the process of data analysis. The module "Tables" (see chapter 11.2) is available for this type of reconstructions in AQUAD.

### 5.5.2 Reconstruction by testing particular relations

Here we can also distinguish two approaches, which correspond to the just described strategies of coding simple or complex sequences of meaning in the data files.

(1) *Confirming simple code sequences.* Let us assume that an interview with parents talking about their educational practices suggests that a father is trying hard to justify his ways of education. Then we could activate the module "Linkages" (see above and chapter 11.3) and examine our tests, looking for sequences of relevant codes, for instance for final or causal codes, in order to confirm our impression deductively.

(2) *Confirming complex relations of codes.* If a researcher wants to confirm relations of codes which exceed the complexity of those linkage structures already built into AQUAD, s/he can use an "design" module in AQUAD, where the researcher add her or his case-specific linkages of codes. How you put your assumptions about complex linkages of codes into a form which can be confirmed or rejected by AQUAD is described in detail with the help of several examples in chapter 12.

## 5.6 How to compare relations of meaning

Permanent comparisons of interpretations within a data set and across different data sets are at the core of all procedures of qualitative analysis (see Shelly & Sibert 1992). Even when reducing the data within a file to codes it appears to be impossible to attain reliable codings without permanently comparing codes/categories and data segments in the file at hand as well as in the other files of the research project. However, in most projects we want to achieve more than to get access to unique world views expressed in individual files by means of a category system valid for all data sets. At some point in the research process we usually want to establish general assumptions across files/cases. This task confronts the researcher with the proverbial danger not to notice the wood because of all the various trees. Tackling all the different formulations in the data texts and elaborating their precise meanings should not prevent us from looking for common elements. Therefore we have to notice systematic relations across files and to compare them, too. Comparing relations between social phenomena, however, regularly leads to findings of numerous conditions in manifold, sometimes controversial combinations. In addition, in many studies one's own findings have to be compared with findings from other studies, that is to perform a *meta-analysis* of qualitative findings. Even if other researchers have used the same research questions as we did or at least comparable questions, they will have found answers that may be only partially compatible with ours. When looking for configurations of conditions of a particular phenomenon, sciences dealing with the complexities of natural systems usually find highly varying constellations across different studies.

Let us take an example that is often discussed hotly in everyday life: "Is there a relation between cancer of the lungs and smoking?" Whoever wants to affirm this question is regularly confronted with the case of an 80 years old grandfather, who smoked all his life long, or with the opposite case of a victim of cancer that never touched a cigarette. Obviously a lot of conditions are to be considered. Empirical studies provide us with many empirical arguments and you can choose the ones you need in order to immunize your point of view against counter-arguments. Only a meta-analysis of all the relevant studies or at least a representative sample of these studies could give clarity about the relevant constellations of conditions.

By applying Boolean algebra to qualitative data Ragin (1987) has developed an important approach comparing qualitative data. The procedure is based on the Quine-McClusky-algorithm of "logical minimization". According to Ragin (1987, p. 121) this approach fulfils the demands for a qualitative comparative research strategy, because

- a large number of cases can be compared;
- complex causal conjunctures can be addressed;
- "parsimonious" reconstructions or explanations can be produced, if desired;
- individual cases can be investigated both in relevant parts and as wholes (see also above: strategy of thematic reduction);
- competing reconstructions or explanations can be evaluated.

If we compare this strategy to variable-oriented approaches, which assume that variables can be combined additively, we find that Ragin's case-oriented comparisons (1987, p. 51 f.) appear to be able to

- uncover patterns of invariance or constant configurations of conditions by minute comparison of individual cases;
- react more sensitively to meaningful configurations of conditions than to relative frequencies of typical cases -- which implies that even a single contradictory case has to be attended to;
- consider cases as entities, that is to understand the conditions of a case in relation to each other instead of in relation to their distribution pattern in the population;
- provide a basis for examining how the conditions found combine in different ways and in different contexts to produce different results.

In order to apply the rules of Boolean algebra to qualitative comparisons we reduce in every single case all codes radically to "truth values.". We are satisfied with the binary statement "condition true" (i.e., given) respectively "condition false" (i.e., not given). The configuration of conditions of a particular case is represented by a row in the table of truth values. The conditions for a case are combined through Boolean

---

multiplication (logical *and*). The different configurations, i.e., the rows in the resulting table are combined additively (logical *or*). In this way, we first represent the single cases as configurations of characteristics or conditions and then compare the patterns of configurations. It does not matter if we do not use exclusively qualitative data; scores on a quantitative dimension can also be reduced to truth values.

AQUAD offers the possibilities of logical minimization for qualitative analysis in its module "Implicants." Whenever comparative operations are necessary in the process of qualitative analysis, we should consider these possibilities. Because of the cyclic nature of qualitative analysis, logical minimization may be helpful during each of its phases.

As a *heuristic* this strategy is already helpful when we start to generate categories for text interpretation, even if we have only analyzed a few texts, that is even with a small basis for comparative operations. During the final phase of analysis, when we want to create a *summary of results* or when we want to group our findings, to distinguish types of writers or speakers, to determine key texts in our data base, etc., the strategy of logical minimization appears to be indispensable. As Ragin (1987, p. 51) has stated, "the potential volume of the analysis increases geometrically with the addition of a single case, and it increases exponentially with the addition of a single causal condition." AQUAD offers well elaborated procedures for grouping or clustering cases by means of logical minimalization of critical conditions. Finally, we can apply the strategy of logical minimization when we intend to *compare several qualitative studies*. This is a methodological step usually called *meta-analysis*. In the area of qualitative research, the strategy of logical minimization could supply the simplicity, transparency, reliability, and documentation desired for meta-analytical comparisons. You will find more details in chapter 13.