Böhm, A. (2004). Theoretical Coding: Text Analysis in Grounded Theory. In U. Flick, E. Kardorff & I. Steinke (Eds.), *A Companion to Qualitative Research* (pp. 270-275). London: SAGE Publications.

5.13 Theoretical Coding: Text Analysis in Grounded Theory

Andreas Böhm

Barney Glaser and Anselm Strauss (1967) created in their *grounded theory* a comprehensive idea of the epistemological and research process in the social sciences (see **2.1**, **6.6**). It extends from the first ideas of a research question to the production of the report on results (see **5.22**). Data collection, analysis and formulation of theory are closely interrelated. The label grounded theory is often used to refer to both the method and also the research result that is sought through the use of this theory. On the basis of empirical research in a particular object area it makes it possible to formulate a valid theory for this area consisting of interrelated concepts and suitable for the production of a description and an explanation of the social phenomena investigated.

1 PROCEDURE ACCORDING TO GROUNDED THEORY

Grounded theory is a *Kunstlehre* (art), and so its procedure cannot be learned in the form of prescriptions. A clear example of the use of the procedure may be found in the chapter by Hildenbrand about Anselm Strauss (see **2.1**). The following summary of the procedure relies in particular on the presentations of Glaser (1978), Strauss (1987) and Strauss and Corbin (1990). The data material here is text in the broader sense of the term (transcribed interviews, field notes, observation reports, and so on). The data collection is oriented to theoretical sampling (see **4.4**): in the early stages as many different people, situations and documents as possible are selected to obtain data covering the complete spectrum of the research question. Subsequently data are sought that will confirm or modify the (provisional) categories of the theory that have already been developed. 'Sensitizing concepts' as guiding principles are the starting point of the research and have the character of open questions ('what happens and how?'). The researchers' own questions, their prior understanding and, related to this, their own prejudices concerning the research issue can be worked out by means of brainstorming and group discussions. The reading of relevant literature also belongs to this (specialist publications, but also journalistic work, novels and stories). The most important intellectual activity in the analytical process consists of *comparison*. This refers less to the search for identical contents than to the search for similarities and differences (Busse 1994). Coding may be described as the *deciphering* or *interpretation* of data and includes the naming of concepts and also explaining and discussing them in more detail. The explanations are reflected in coding notes. The result of coding is then a list of terms as well as an explanatory text. Three types of coding may be distinguished that may be partially considered as phases in the research process – open, axial and selective coding (see

below). 'Code' is a technical term from the analytical procedure and signifies a named concept. In the data indicators are sought of the phenomenon being studied. The target of the first analyses is the production of codes that relate *directly* to the data. Initially, concepts always have a provisional character, and in the course of the analysis they become more differentiated, numerous and abstract. The differentiated concepts are known as *categories*.

Writing of memos

Theoretical memos are based on the coding notes mentioned above and on broad interrelations that are gradually revealed by the investigator. The writing of theoretical memos requires researchers to distance themselves from the data, and also helps them to go beyond purely descriptive work (motto 'Stop and memo!'). In the course of the analysis memos can become starting points for the formulation of the final manuscript. Exactly as with theoretical memos, there is a constant process of writing and revision (theoretical sorting). Working in a team of colleagues prevents one-sidedness and can speed up the epistemological process, for which reason working in a team of investigators and (research) supervision have proved to be of value.

Open coding

In open coding data are 'broken down' analytically, and in this the principle of grounded theory shows itself: from the data, that is from the text, a succession of concepts is developed that may ultimately be used as building blocks for the model. As a first step it is advisable to analyse single short textual passages (line by line). Subsequently larger paragraphs or even whole texts may be coded. In order to avoid simple paraphrasing, the following 'theory-generating' questions are asked of the text.

- What? What is at issue here? What phenomenon is being addressed?
- Who? What persons or actors are involved? What roles do they play? How do they interact?
- How? What aspects of the phenomenon are addressed (or not addressed)?
- When? How long? Where? How much? How strongly?
- Why? What reasons are given or may be deduced?
- For what reason? With what intention, and for what purpose?
- By what means? What methods, tactics and strategies are used to achieve the goal?

In coding researchers use their background knowledge about the context of the textual passage being investigated and, in general terms, their knowledge about the area of investigation. The result of the work is an interpretative text which adheres to analytical thinking about the phenomenon and which often contains questions about how the phenomenon might be further investigated (see **2.1** for an example). *Theoretical codes* in the sense of terms from scientific theories should initially be avoided. More profitable are *in-vivo codes,* which, as colloquial interpretations of the phenomena, are taken directly from the language of the field of investigation. In-vivo codes are components of 'theories' formulated personally by the producers of the text in question. Traditional categories such as age, gender, level, and so on, should only be used after a thorough scrutiny of their relevance. The text and the researcher's background knowledge make it possible to specify different aspects or properties of the phenomenon being investigated. Mental comparisons (including false and extreme instances) provide some indication of the possible variation in these aspects or in their characteristics. If a particular aspect or property may be plotted on a continuum, then we are dealing with a dimension.

Open coding is an expanding procedure in the sense that considerable quantities of interpretative text can be added to a small segment of an original text. To retain an overview, the investigator should continually write memos, and sort and weigh up the results of the work. In ordering the interim results it will become clear what concepts are important for the researcher's own question and therefore require deeper analysis, and what results should be discarded and not pursued in greater depth.

Axial coding

This step serves to refine and differentiate concepts that are already available and lends them the status of categories. One category is located at the centre and a network of relationships is

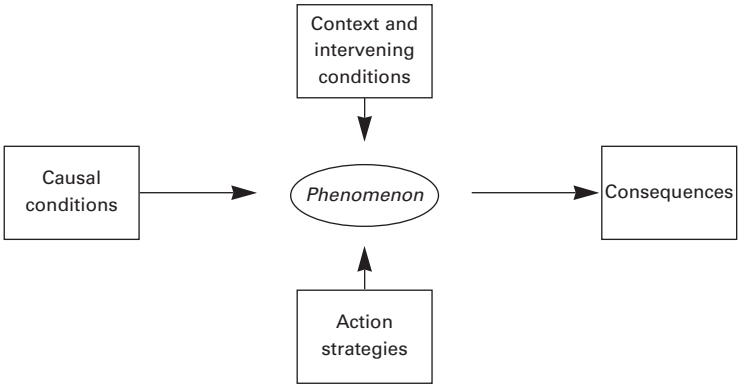

Figure 5.13.1 Coding paradigm for social science research questions

developed around it. Typically, axial coding is used particularly in the middle and later stages of an analysis. In the same way as open coding, axial coding is applied to very short textual segments (in the sense of a detailed analysis), to larger extracts or to the entire text. For theory-formation what is of particular importance is the development of relationships between the axial categories and the concepts that are related to them in terms of their formal and content aspects. The axial category is developed in its temporal and spatial relationships, in relationships of cause and effect, in means–ends relationships and in terms of argumentative and motivational connections. The hypothetical relationships in axial coding must be repeatedly checked in a deductive procedure, using new data material. To explain the relationships between categories that relate to partial aspects of social action, Strauss's coding paradigm has proved to be of value (Figure 5.13.1).

The following example, in which 'pain' has been selected as the axial category, will illustrate the coding paradigm: 'If I've drunk too much (context), I get (condition) a headache (phenomenon/axial category). Then I take an aspirin (strategy). After a while it's better (consequence)' (taken from Strauss and Corbin 1990: 98).

The *phenomenon* denoted by the axial category is, for example, an event or a fact. The actions of an individual as well as interactions between different people revolve around the phenomenon. The following questions make easier the choice of axial category: What do my data refer to? With what are the actions and interactions in the data actually concerned? *Causes* or causal conditions contribute to the occurrence or development of the phenomenon, for instance, a broken leg (= cause) leads to pain (= phenomenon). It is important here to clarify the properties of the cause. For the example given, this would mean asking: What kind of fracture? Simple or compound? And so on. With causes a distinction must sometimes be made between the subjective view, as it may be presented, for example, as a speaker's perspective in an interview text, and the view of the researcher. *Causes* are normally only valid in a particular set of *conditions*, and here what is of particular importance for the formation of an action-related theory are the conditions that promote or restrict the possibilities for action or interaction. Under *contextual conditions* are included particularly time, place and duration. And among *intervening conditions* we find the social, political and cultural environment and the individual biography. *Actions* and *interactions* have two properties. (1) They are processes and have a sequence, and it is therefore appropriate to ask about sequences and temporal course of action. (2) They are goal-oriented and are often performed for particular and specifiable reasons, for which reason one may refer to (interactional) strategies or tactics.

'Goal-oriented' should not be confused with (conscious) intention. For the purposes of the analysis a functional mode of observation is preferred that disregards intentions. Strauss and Corbin (1990: 104) offer the following example. In an investigation into the self-consciousness of children a field observation was analysed. A child throws a glass of milk onto the floor and is rebuked by its mother in the presence of other children. It was not a conscious intention of the mother that the child's self-consciousness

Table 5.13.1 Coding families (adapted from Glaser 1978: 75–82)

Coding families	Concepts	Examples
The Six Cs	Causes, contexts, contingencies, consequences, conditions	… of pain suffering
Process	Stages, phases, phasings, transitions, passages, careers, chains, sequences	Career of a patient with chronic pain
The Degree Family	Extent, level, intensity, range, amount, continuum, statistical average, standard deviation	Extent of pain suffering
Type Family	Types, classes, genres, prototypes, styles, kinds	Kinds of pain – sharp, piercing, throbbing, shooting, sting, gnawing, burning
The Strategy Family	Strategies, tactics, techniques, mechanisms, management	Coping with pain
Interactive Family	Interaction, mutual effects, interdependence, reciprocity, symmetries, rituals	Interaction of pain experience and coping
Identity-Self Family	Identity, self-image, self-concept, self-evaluation, social worth, transformations of self	Self-concepts of pain patients
Cutting-Point Family	Boundary, critical juncture, cutting point, turning point, tolerance levels, point of no return	Start of chronification in the medical career of pain patient
Cultural Family	Social norms, social values, social beliefs	Social norms about tolerating pain, 'feeling rules'
Consensus Family	Contracts, agreements, definitions of the situation, uniformity, conformity, conflict	Compliance

should suffer from the rebuke (here the interactional strategy), but rebuking can be coded here as a strategy.

Actions and interactions lead to particular consequences. Strauss (1987: 57) recommmends that care be exercised in applying the coding paradigm to linguistic peculiarities in the data: researchers should regard keywords such as 'because', 'since', or 'owing to' as indicators of causal conditions. Consequences of actions are often indicated by means of expressions such as 'as a result of', 'and so', 'with the result that', 'the consequence was', 'consequently'.

As a further stimulus in *axial coding*, an overview of theoretical framing concepts may be used, or so-called *coding families*. The C-family (causes, contexts, consequences, conditions, etc.) corresponds to the coding paradigm described above. For Glaser (1978: 74) this coding family is central to the analysis of social events (the 'bread and butter theoretical code of sociology') (see Table 5.13.1).

Selective coding

In this phase the researcher is particularly active as an author on the basis of the categories, coding notes, memos, networks and diagrams so far developed. As a starting point for establishing the main phenomenon of the analysis it is advisable to look at coding lists, summarizing memos and representations of networks. The main phenomenon is described as the *core category* and is possibly already present in the formulation of the research question of the particular investigation. Admittedly it must sometimes occur in the research process that a different phenomenon than originally assumed will take on central importance for the issue in question. There are indeed such shifts in a research perspective in the course of data collection and interpretation, which lead to new and surprising discoveries. For this reason *grounded theory* recommends asking repeatedly, in the course of an investigation, which phenomena are central and formulating appropriate theory-memos.

If a number of well-worked-out axial categories are available we may assume that the central phenomenon has been captured in its essential aspects – otherwise it is necessary to return to earlier phases in the research process. In the practice of research there are two possibilities. (1) One of the axial categories includes the central phenomenon and is therefore suitable as the core category. The candidate for the

core category is characterized by its formal relationships with all the other important categories and occupies a central position in the network of terms. (2) It often proves to be sensible to give a central location to a phenomenon to which more than a single axial category relates. In such a case it is necessary to detach oneself from the axial categories, and to formulate a new category which comes about by means of summarizing or reformulating one of the existing categories.

Frequently investigators experience difficulties in sticking to the central proposition of the investigation because of the 'surfeit of important details'. Here one should ask what 'story' the data tell. The researcher will summarize in a few sentences the results of the investigation for an interested reader. Guiding questions for this kind of record are: What is the issue here? What have I learned from the investigation? What is central? What relationships exist? The main story revolves around a core category, unfolds this in a concise way and shows relationships with other important categories. After determining the core category, its properties and dimensions, other relevant categories are related, systematically and in a schematically oriented manner (for example, in the sense of the coding paradigm), to the core category. Once the relationships to the main categories have been formulated, their particular properties and dimensions may be compared with regard to regularities and patterns.

An example of selective coding

In an investigation of the psychological reaction to the nuclear accident at Chernobyl (Legewie et al. 1989), it was possible to discover the following pattern: in experiencing a threat to one's own physical health and life expectancy what was decisive was whether age was an important constituent in a person's self-image. 'Young' people (not in the sense of biological age, but in the sense of a self-attributed property, or 'subjective age') saw themselves in this respect as far more threatened than 'old' people. This statement could only be made after a systematic comparison of combinations found no evidence of the combinations 'young' + 'no threat' and 'old' + 'severe threat'. The example demonstrates how gaps within a theory (such as defective specification, or defective grounding of the statements in the data) may, through a systematic procedure, be discovered, reviewed and ultimately eliminated.

The degree of generalizability of a theory developed in this way depends, at least in part, upon a process of abstraction that permeates the entire research procedure. The more abstract the formulation of the developed categories – in particular the core category – the more widely the theory may be applied. But, in addition, the time and energy invested in its development will also increase, because ultimately the route from the data to the relatively abstract categories must be documented in every detail. A grounded-theory is testable by again confronting the theoretical propositions, as hypotheses, with reality. For social and, in particular, historical phenomena there are limits to this, because the social conditions cannot be reproduced at will nor very precisely.

2 LIMITATIONS OF THE METHOD

The character of grounded theory as a *Kunstlehre* (art) renders its learnability more difficult, and makes particular demands of investigators in respect of their creativity. The requirement – which seems initially to be liberating – that one should distance oneself from existing theories and allow the theory to grow out of the data, often causes insecurity among students. Particularly in respect of decisions about the transition points between the different phases of coding, there are scarcely any fixed rules (Flick 2002: 185). The pragmatic direction, in terms of which data collection and analysis is complete when theoretical saturation is reached (that is, no new aspects can be incorporated into the theory), is hardly adequate for beginners. From this it again becomes clear how important teamwork and research supervision are in the context of this method.

3 DEVELOPMENTS AND PERSPECTIVES

While Barney Glaser withdrew from active research in the 1980s, Strauss developed the approach further and devoted himself in particular to a didactic orientation in order to make the method teachable and learnable (Strauss 1987; Strauss and Corbin 1990). Glaser (1992) accuses Strauss in this respect of having abandoned the original idea of allowing the theory

to 'emerge' in favour of 'forcing' theoretical structures. His criticism was particularly directed at the *axial coding* paradigm. In the first comprehensive publication on *grounded theory* it was vitally important to Glaser and Strauss (1967) that the method be adapted to particular questions and circumstances. Adaptations or systematic further developments in the procedure are to be found in Breuer (1996), Flick (1996) and Charmaz (1990). Breuer supplements the grounded theory approach for his own questions by the use of transference and counter-transference in the psychoanalytical sense (see **5.20**). Flick (1996), in his investigation of psychology and technology, proceeds on the basis of Moscovici's (1984) concept of social representations. On the assumption that in different groups different views of technology will be found, groups are pre-selected for investigation. In that way *sampling* is limited to the selection of cases that differ between the groups. Charmaz (1990) takes 'thick' presentation of cases as a starting point for theory development.

A further development of grounded theory is also to be seen in the improvement in practical analysis through the use of specific computer programs (see **5.14**). Programs such as ATLAS.ti (Muhr 1997) support the task of analysis and make quality control possible by ensuring that the analytical process of individual researchers or complete teams can be documented and reproduced in every detail.

FURTHER READING

Glaser, B. G. (1978) *Theoretical Sensitivity*. Mill Valley, CA: The Sociology Press.

Glaser, B. G. and Strauss, A. (1967) *The Discovery of Grounded Theory. Strategies for Qualitative Research*. Chicago: Aldine.

Strauss, A. and Corbin, J. (1990) *Basics of Qualitative Research*. Newbury Park, CA: Sage.