COMPUTER-ASSISTED ANALYSIS
OF QUALITATIVE DATA

Paper prepared for the Discussion paper series
of the LSE Methodology Institute

by
UDO KELLE
University of Bremen

August 1997

Contents

1. Context

2. Conceptual issues

3. Techniques of computer-aided qualitative analysis
   3.1 Discovering differences, commonalities and relations between text segments
   3.2 Developing typologies and theories
   3.3 Examining hypotheses

4. Methodological benefits and problems related to computer-use in qualitative research

5. Literature
1  Context

Although software for handling textual data had been available since the mid 1960s, it was not until the early 1980s that qualitative researchers discovered that the computer could assist them in working with their data (KELLE 1995; 1f.). Before that, programs for textual analysis like The General Inquirer had attracted only a limited group of experts in the field of quantitative content analysis. The reluctance of most qualitative researchers to integrate computers into their work certainly reflected a certain distance from the mainstream methodology of quantitative survey and experimental research where, during the 1960s and 1970s, the computer became an indispensable aid. At that time computers were seen by many social scientists as tools for nothing more than the statistical analysis of numerical data (or the quantitative content analysis of textual data). The idea that computers could one day become an indispensable tool for the storage, retrieval and manipulation of text was far away.

This situation was radically changed by the advent of the Personal Computer. Like other 
\textit{hommes des lettres} qualitative researchers discovered rather quickly the enormous possibilities for text manipulation that were offered by the new technology. In the mid 1980s several qualitative researchers with advanced computer knowledge and expertise started, independently from each other, to develop software which could support the analysis of qualitative data. While most of these programs were designed only for the purposes of one specific research project, some packages were put on the market by their developers. Programs like THE ETHNOGRAPH, QUALPRO and TAP started a train of development in the field of computing in qualitative social research. A number of additional software packages, NUDIST, MAX and WINMAX, ATLAS/ti, HYPERRESEARCH, HYPERSOFT (to name but a few) followed in the subsequent years. Nowadays more than 20 different software packages are available which can assist qualitative researchers in their work with textual data, and some of these programs (esp. THE ETHNOGRAPH and NUD*IST) are widely applied in the qualitative community. Their first, sometimes rather awkward and user-unfriendly versions, were rapidly improved and more and complex functions were added. Looking at the scene now, one will even find that these developments have culminated in a race between developers to include as many features as possible in the newest versions of their programs.
Nowadays the field of "computer-assisted qualitative data analysis" can be seen as the most rapidly developing field in the domain of qualitative methodology, with its own "networking projects", conferences and several discussions lists on the Internet.

Given the fact that literature describing such software packages in detail (e.g. TESCH 1990; WEITZMAN, MILES 1995) is always in danger of becoming rapidly out of date, this paper will not concentrate on specific programs, but will discuss more generally techniques of qualitative data administration and data analysis which can be supported by computer programs. Thereby, strong emphasis will be laid on the methodological aspects of computer use in qualitative research.

2 Conceptual issues

The "Operation called Verstehen" (ABEL 1948), the understanding of the meaning of text, can certainly not be performed with the help of an information processing machine, since it cannot be formalised easily (KELLE 1995; 2). However, there are still a variety of mechanical tasks involved in the analysis of textual data. The qualitative research process often generates huge amounts of interview transcripts, protocols, field notes and personal documents, which if not managed properly, can result in ‘data overload’ (MILES, HUBERMAN 1994). Since data analysis and theory construction are closely interlinked in qualitative research, the researcher generates many theoretical concepts in this ongoing process which are often recorded across numerous notebooks, manuscript pages and index cards. Keeping track of the emerging ideas, arguments and theoretical concepts can be a mammoth organisational task.

These problems have been well known for centuries among scholars who have to work with large amounts of texts. A variety of methods for coping with them have been developed, most of them based on of one of the following two basic strategies: (1) the construction of indexes (or "registers", or "concordances") of various kinds, and (2) the inclusion of cross references in the text. Both these techniques can help with an important task of data management: to draw together all the text passages which have something in common. Before the advent of computers, ‘cut-and-paste’ techniques were the most widely used methods in qualitative research for organising data material in this way -- researchers ‘cut up field notes, transcripts and other materials and place data
relating to each coding category in a separate file folder or manila envelope’ (TAYLOR, BOGDAN 1984:136; see also LOFLAND, LOFLAND 1984:134).

To computerise such tasks, a non-formatted textual database has to be built up. Unfortunately, standard software, like word processors or standard data base systems, is usually only of limited use for the construction of such databases, since it does not support those techniques of data management which are needed to structure non-formatted textual data, for instance

1. the definition of pointers containing index words together with the "addresses" of text passages which can be used to retrieve indexed text segments,

2. the construction of electronic cross references with the help of so-called "hyperlinks" which can be used to "jump" between text passages which are linked together.

All software packages developed especially for qualitative research are based on one or both of these techniques. Furthermore, current versions of programs like THE ETHNOGRAPH, HYPERRESEARCH, HYPERSOFT, MAX, NUDIST or ATLAS/ti contain a variety of additional features:

1. Facilities for the storing of the researchers’ comments ("memos") which can be linked to index words or text segments.

2. Features for defining linkages between index words.

3. The use of variables and filters so that the search for text segments can be restricted by certain limitations.

4. Facilities for the retrieval of text segments with specified formal relations to each other (e.g. text segments that appear with a certain specified maximum distance of each other)

4. Facilities for the retrieval of quantitative attributes of the database.

Drawing on examples from research practice I shall discuss how such techniques can be used to support the analysis of qualitative data.
3 Techniques of computer-aided qualitative analysis

The use of computers in qualitative research cannot be regarded as one single method which can be followed in a step by step manner. Furthermore, it comprises a variety of different -- straightforward and very complex -- techniques. Certainly the right choice for one of these techniques can only be made in regard to the researcher’s methodological background, his or her research questions and the research objectives.

A terminological caveat has to be made here: the term "Computer-aided qualitative data analysis" is certainly misunderstood if one sees software packages like THE ETHNOGRAPH, ATLAS/ti or NUD•IST as capable of performing "qualitative analyses" in the same way as SPSS can perform analyses of variance. These software packages are tools to mechanise tasks of ordering and archiving texts and represent software for "data administration and archiving" rather than tools for "data analysis". Consequently, the term "computer-aided qualitative data analysis" as it is used in this paper refers to the interpretive analysis of textual data where software is used for the organisation and management of the data. In the following, different methodological tasks in qualitative data analysis will be presented which can be supported by specific computer programs.

3.1 Discovering differences, commonalities and relations between text segments

After having collected unstructured textual data through fieldwork or open interviewing the qualitative researcher may want to construct "meaningful patterns of facts" (JORGENSON 1989; 107), by looking for structures in the data. This is often conducted by comparing different pieces of data in order to find commonalities, differences or linkages between them. To some degree this process is similar to doing a jigsaw puzzle: the analyst would start by collecting certain pieces of the textual data which are similar in a certain respect. He or she will analyse several parts and their connections, that is the specific way they could be linked or connected to form a meaningful picture. In their famous
monograph about the *Discovery of Grounded Theory* (1967) Glaser and Strauss coined the term "constant comparative method" for this process, whereby "underlying patterns" are discovered through careful and intensive comparison. The central prerequisite for this is "coding", i.e. relating text passages to *categories* that the researcher had either previously developed or which he or she develops *ad hoc*. ‘The analyst starts by coding each incident in his data into as many categories of analysis as possible, as categories emerge or as data emerge that fit in an existing category’ (GLASER, STRAUSS 1967; 105). In practical terms this means "...noting categories on margins, but can be done more elaborately (e.g. on cards). It should keep track of the comparison group in which the incident occurs" (GLASER, STRAUSS 1967; 106).

Most of the software programs for qualitative analysis support this process of categorising and comparing text segments by offering "code-and-retrieve" facilities (Kelle 1995, p.4ff.; Richards, Richards 1995), which allow for

(1) the attachment of "codes" (=index words) to text segments and

(2) the retrieval of all segments from a defined set of documents to which the same code had been assigned (see figure 1).

![Figure 1: Coding and retrieval](image-url)
The comparison of text passages can be aided by attaching variables to whole documents which can be used for *selective retrievals* where the search for text segments is restricted by certain limitations, permitting for instance the retrieval of statements about certain topics from only those interview participants who have a certain characteristic in common. For example, a qualitative researcher investigating the division of household labour among married couples could first of all retrieve all the text segment in which wives talk about housework and contrast them afterwards with text segments on the same topic from interviews with husbands.

The majority of computer programs for qualitative analysis are based on "code-and-retrieve" facilities. In a recent article about methodological aspects of computer use in qualitative research Coffey et al. have warned that the one-sided emphasis on code and retrieve operations may lead to neglecting other computer-aided techniques, esp. techniques offered by hypertext systems (COFFEY, HOLBROOK, ATKINSON 1996). Looking at other hermeneutic sciences (esp. historical and critical biblical exegesis) one can see indeed that indexing (=coding and retrieval) is not always the best tool to support the comparison of text segments (or ‘synopsis’ as it is called in biblical exegesis). The use of cross-references (=hyperlinks) is another important strategy of data administration which could be useful here. Unfortunately, until now there are only few software packages, namely ATLAS/ti and HYPERSOFT, which support the construction of hyperlinks.

### 3.2 Developing typologies and theories

In many qualitative research projects the comparison of text segments leads to the construction of descriptive typologies and to the development of theories. Since qualitative research often starts with the collection of empirical data and theories are developed on this basis, qualitative methodologists have sometimes adopted a naive inductivist model of the research process, assuming that theoretical categories will simply emerge from the empirical material if the researchers free their minds from theoretical preconceptions. Following such views, which are often influenced by the early methodological writings of Glaser and Strauss (see for example GLASER, STRAUSS 1967; 37), qualitative
researchers approach their empirical field without any theoretical concepts whatsoever.

However, one of the most crucial insights of the modern philosophy of science and cognitive psychology is the fact that "there are and can be no sensations unimpregnated by expectations" (LAKATOS 1982; 15). In his later methodological writings Strauss took this "theory-ladenness" of empirical observation into account by proposing a "paradigm model" (STRAUSS, CORBIN 1990; 99ff.). According to Strauss and Corbin, a "coding paradigm" represents a general theory of action which can be used to build a skeleton or "axis" of the developing grounded theory. Glaser, although he has fully repudiated Strauss and Corbin’s concepts in his last book (GLASER 1992), proposed a similar idea: "theoretical codes" represent those theoretical concepts which the researcher has at his or her disposal independently from data collection and data analysis (GLASER 1978).

Such coding paradigms and theoretical codes (which are often implicit at the beginning of an empirical study) can be made explicit by constructing a coding scheme. The following example shows such a code scheme which comes from a research project that studies the transition from school to the labour market (HEINZ 1996; HEINZ et al. 1997). Open interviews were conducted in order to reconstruct the decision processes of school-leavers who entered vocational training courses. In this project the decision processes described by the interviewees were structured according to the following three categories: (1) aspirations, which represent the respondents’ preferences that were used to account for occupational options, (2) realisations, which consist of the actual steps of action that were taken to fulfil realisations, (3) evaluations, which were the respondents’ assessments of the relations between aspirations, conditions and consequences of action. These categories represent the sub-codes (1.1-1.3; 5.1-5.3, 8.1-8.3) shown in Figure 2.

The second type of code categories frequently used for qualitative coding are codes derived from common-sense knowledge. In the interviews with school-leavers all text passages were coded where the interviewee talked for instance about experiences in his/her job, about relevant institutions, about his/her family etc. The main categories (1, 5, 8) shown in Figure 2 represent examples of categories which are drawn from common-sense knowledge.
Both kinds of codes (derived either from common-sense knowledge or from abstract theoretical concepts) that play the most important role at the start of the qualitative research process are either rather trivial or highly abstract. Thus they have something in common: they do not denote well defined empirical events, but serve heuristic purposes. They represent some kind of theoretical axis or "skeleton" to which the flesh of empirically contentful information is added (STRAUSS, CORBIN 1990; KELLE 1994). The research project mentioned above started with structuring the material according to the general categories aspirations, realisations and evaluations and ended up discovering eight different types of biographical aspirations, for example the type "delegation" - young adults who try to delegate the responsibility for their occupational career to the managers of their companies or to the officials at the employment office.

To develop such typologies or theoretical concepts a fine-grained analysis of text segments is necessary in order to find those aspects (or "dimensions") which can serve as criteria for a comparison in order to develop subcategories or subdimensions of the categories already used for coding. This process of dimensionalisation (STRAUSS, CORBIN 1990, p. 69ff.) may be clarified with another example from the same research project. There the respondents’ orientations towards marriage were investigated by first coding text segments
according to whether the topics "marriage" or "family" were mentioned. In a second step text segments of those respondents who regarded marriage as a crucial goal in life were selectively retrieved. The comparison of these text passage lead to the discovery of three different dimensions of this category:

1. marriage was seen by some respondents as the only acceptable form of cohabitation,

2. others viewed marriage as the prerequisite for child-centred family formation,

3. still others regarded marriage as a kind of safeguard. Respondents with this last orientation advanced three different arguments: marriage was seen as (1) offering financial protection, (2) providing a support for the bonding between the partners, or (3) a means to fulfil the expectations of their social environment (parents, relatives etc.).

Through comparison and dimensionalisation, three different sets of code categories were developed in the analysis process: firstly categories which refer to how important the issue marriage was for the respondents; secondly, categories which refer to the orientations towards marriage among those who see marriage as a crucial goal in life, (as the only acceptable form of cohabitation, as a prerequisite for child-centred family formation, as a safeguard); and, thirdly, categories which refer to arguments for why marriage is a safeguard. The hierarchical relation between these sets of categories is shown in Figure 3.
In the terminology of information science the tree structure displayed in Figure 3 can be formally described as a network or graph in which the categories or codes represent the nodes of the graph and the lines between them the edges. Using such a network approach it is possible to expand the basic principle of non-formatted textual database systems in which codes are connected with pointers to text segments (MUHR 1991, 1992). Thereby, it is possible to electronically store the whole structure of the hierarchical typology or graph shown in Figure 3. Consequently, this graph can not only be used to give an account of the emerging typology or theory, but it also allows for rather complex retrieval procedures that follow a long path from one node at one end of the network or graph to a node at the other end.

It should be noted here that graphs may be structured in quite diverse ways: ATLAS/ti and HYPERSOFT are programs that allow the researcher to define all possible linkages between the nodes (permitting the researcher to define "cycles" and "loops"). Other programs (e.g. NUD•IST ) restrict the researcher, for example, to the construction of more tightly structured networks (e.g. hierarchical trees).
3.3 Examining hypotheses

Software for the computer-aided analysis of textual data can also be helpful to aide the refinement of theoretical concepts and the examination of hypotheses. However, one must not forget here that qualitative hypothesis examination is a process quite distinct from statistical hypothesis testing. In the qualitative methodological literature one will find nothing comparable to the precise decision rules which are applied in statistical significance testing. Instead "Testing and confirming findings" (MILES, HUBERMAN 1994: 262) or "verification" (STRAUSS, CORBIN 1990: 108) means in qualitative research: returning to the data (i.e. re-reading one’s transcripts or field notes), or returning to the field (i.e. conducting new observations or interviews), in order to find some confirming or disconfirming evidence. Precise rules that are formulated to inform the researcher, with certainty, about when he or she has to reject or abandon a certain hypothesis are nowhere to be found. Qualitative hypotheses, when they first come into a researcher’s mind, are usually not highly specified and definite propositions about certain facts, but tentative and imprecise, sometimes very vague conjectures about possible relationships. Rather than calling them hypotheses one should call them hypotheses about what kind of propositions, descriptions or explanations will be useful in further analysis. They are insights that "whatever specific claim the successful H(ypothesis) will make, it will nonetheless be an hypothesis of one kind rather than another." (HANSON 1971: 291)

A qualitative researcher investigating gender-specific occupational careers may, for example, develop the hypothesis that there must be a relationship between their interviewees’ orientations towards work and family. To examine this hypothesis complex retrieval facilities can be extremely helpful. Most of the software packages currently available contain such complex retrieval facilities which support the search for co-occurring codes. Thereby, co-occurrences can be defined in various ways:

- Indicated by overlapping or nested text segments to which the codes under investigation are attached, as shown in Figure 4.

- Indicated by text segments that are coded with certain codes (here A and B) that appear within a certain specified maximum distance of each other. If this maximum distance is set at, say, 8 lines, the program would retrieve all
instances where a text segment coded with code B starts within up to 8 lines of the start or end of a text segment coded with code A. (see Figure 5)

- indicated by **sequential ordering** (Code A is regularly followed by Code B), as shown in Figure 5.

---

**Figure 4:** Overlapping and nesting text segments

**Figure 5:** Proximity and sequence of codes

Thus the hypothesis of a relationship between work and family orientations may be examined by retrieving all text segments coded with "work orientation" and "family orientation". Of course, the notion of hypothesis testing would be rather misleading if one understood it as an attempt to "verify" or "falsify" an empirically contentful statement. But this kind of hypothesis examination can
lead to the development of falsifiable statements, for example if one finds that
the interviewees with specific work orientations also show specific orientations
towards the family. Here facilities for locating co-occurring codes are used as a
heuristic device: the objective is to retrieve the original text to which the co-
occurring codes had been attached. Then the researcher investigates the
meaning of a certain co-occurrence by a thorough analysis of the original text.
The interpretative analysis of interview texts forms the basis for the
clarification and modification of the researchers’ initial (general or vague)
assumptions.

Developers of two different software packages, HYPERRESEARCH and
AQUAD, propose a more formal approach towards qualitative hypothesis
examination (see HESSE-BIBER, DUPUIS 1995; HUBER 1995).

When using the hypothesis testing module of HYPERRESEARCH the researcher
formulates his or her hypotheses in the form of "production rules" in which
codes are connected with "if-then" statements: a researcher who has coded
his/her data with codes for "critical life events" and "emotional disturbances"
may wish to examine the hypothesis that critical life events are always or
frequently accompanied by emotional disturbances. S/he could then transform
their hypothesis into a query about all co-occurrences of text segments coded
as "critical life events" with segments coded as "emotional disturbances".
Using HYPERRESEARCH’s hypothesis tester one would formulate the rule

   IF "critical life events" AND "emotional disturbances" THEN ADD
   "life event has caused stress".

If the program finds both the code "critical life events" and the code "emotional
disturbances" in a given document, the hypothesis is confirmed for that
document and the code "life event has caused stress" is added to it.

HYPERRESEARCH only searches for the presence of certain codes within a
given set of documents, and in doing so does not take the precise location of
the text segments into account. In contrast, the program AQUAD helps the
researcher to use information about overlapping, nesting, proximity or
sequence of text segments for hypothesis examination. Taking our previous
example and using AQUAD one would first code the text segments with the
codes "cle" (for "critical life events") and "emo" (for "emotional
disturbances"). Let us assume that during this process the following hypothesis

14
has come to the researcher’s mind "Whenever interviewees talk about critical life events they will also, at the same time, mention emotional disturbances". One can now operationalise "at the same time" as "within a maximum distance of 5 lines in the interview transcript" and run a retrieval that finds all text segments coded with "cle" where a text segment coded with "emo" also occurs within a maximum distance of 5 lines. Looking at the result of such a retrieval shown in Figure 6 one can for example see that in the interview "bioss1" the association of "cle" and "emo" occurs only once (at line 102), while in interview "bioss2" there are 5 text passages where text segments coded with these codes are very close to each other.

<table>
<thead>
<tr>
<th>Hypothesis 1 / Codefile</th>
<th>bioss1.cod</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 102 cle - 102 104 emo</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothesis 1/ Codefile</th>
<th>bioss2.cod</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 28 cle - 26 30 emo</td>
<td></td>
</tr>
<tr>
<td>65 70 cle - 72 82 emo</td>
<td></td>
</tr>
<tr>
<td>110 112 cle - 111 115 emo</td>
<td></td>
</tr>
<tr>
<td>220 228 cle - 212 224 emo</td>
<td></td>
</tr>
<tr>
<td>450 452 cle - 456 476 emo</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 6:** result of a co-occurring code search with AQUAD

Thus the co-occurrence of codes (defined as the overlapping, nesting, proximity or sequential ordering of text segments) indicates the presence of critical evidence for or against the hypothesis. Contrary to the first example of qualitative hypothesis examination (concerning the relationship between work and family orientations) the primary goal with this more formal approach would not be to retrieve text but to use the information represented by the codes themselves as a basis for decision making. Similar to statistical significance testing, the decision making process is strictly rule governed. However, there are certain methodological requirements and limitations to such a strategy:

1. The prerequisite of independent testing requires that a hypothesis is not tested with the same empirical material from which it is developed.
2. The hypotheses must be empirically testable, that means they must be precise enough and have empirical content.

3. The codes which are used for hypothesis testing must denote clearly defined phenomena in a reliable and stable way.

4. **Methodological benefits and problems related to computer-use in qualitative research**

Since the advent of the first computer programs that support qualitative research there has been a lively debate about their potential methodological merits and dangers, with discussants expressing great optimism (CONRAD AND REINARZ 1984; RICHARDS, RICHARDS 1991) as well as concerned warnings (AGAR 1991, SEIDEL 1991; SEIDEL, KELLE 1995, COFFEY, HOLBROOK, ATKINSON 1996).

One can identify three aspects frequently mentioned in the literature concerning the benefits of software for qualitative research:

1. By mechanising tedious and cumbersome tasks of data organisation, like searching and copying text segments, computer-use can lead to **greater efficiency**. Thus, software helps to save time and can assist the management of larger samples (KELLE, LAURIE 1995). However, it is crucial to be aware that a simple increase in sample size alone does not necessarily imply that the research findings will be more valid. In qualitative researcher a large sample is usually not regarded as valuable in itself. However, multiple comparisons between purposefully selected cases are crucial for a qualitative study to identify patterns and to develop categories. An increase in sample size may therefore add greater breadth to the scope of the analysis. However, there is also a real danger with software for textual data management of being overwhelmed by the sheer volume of information that becomes available when using computer technology. The amount of time and effort required to prepare the data and enter it into the program is not inconsiderable and increases in tandem with sample size. One should therefore be aware that the potential benefits
of a larger sample size may be outweighed by the extra costs in time and effort required for data preparation and data entry.

2. The use of software packages can make the research process more **systematic** and **explicit**, and therefore more transparent and rigorous while systematising procedures that previously had been unsystematic and enabling researchers to codify exactly how they analyse their data (CONRAD AND REINARZ 1984: 4-6). Thus computers could add trustworthiness to a methodology which had always suffered from the reputation of seducing the researcher into employing an unsystematic, subjective or journalistic style of inquiry.

3. By ridding the researcher of boring and cumbersome mechanical tasks, software for textual data management can free up time that can be spent on more creative and analytic tasks. Thus computer programs can enhance the researcher’s creativity, by allowing him or her to experiment and to "play" with the data, and to explore the relationships between different categories more thoroughly (LEE, FIELDING 1995)

Warnings about the potential methodological dangers of computer use often relate to the possibility that the use of computers could alienate the researcher from their data and enforce analysis strategies that go against the methodological and theoretical orientations qualitative researchers see as the hallmark of their work. Furthermore, concern is often expressed that the use of computer programs may impose a certain methodology on the user. Such worries were especially fuelled by the observation made by Lonkila that user’s guides as well as the methodological writings about software for qualitative data management give the impression of a strong influence of Grounded Theory (LONKILA 1995, p.46). But grounded theory and computer-aided qualitative analysis also share some very problematic aspects, as Lonkila points out: both overemphasise coding and in doing so neglect other forms of textual analysis, especially the kind of fine-grained analysis employed in discourse analysis. Coffey et al. have warned that the strong ties between "code and retrieve" software and grounded theory methodology may inspire a new orthodoxy in qualitative research (COFFEY, HOLBROOK, ATKINSON 1996). However, a closer look at methodological backgrounds of the developers gives the clear impression that different programs have been developed on the basis of rather different theoretical and methodological conceptions. The same holds
true for the users: In a meta analysis of empirical studies Lee and Fielding found that 70% of a sample of qualitative studies performed with the help of computers show no explicit relation to Grounded Theory (LEE, FIELDING 1996: 3.2). Thus, the frequent mentioning of grounded theory maybe explained by the fact that proponents of the Grounded Theory approach belong to those very few authors who try to describe in detail the analytical procedures applied in qualitative research. Consequently, it is not surprising that developers of software that supports qualitative analysis who are searching for a methodological underpinning usually draw on the methodology of Grounded Theory as one of the most well-known and most explicit approaches in qualitative analysis.

Lee and Fielding (1991, p.8) have linked the fear of the computer taking over the analysis to the famous literary archetype based on Mary Shelley’s 19th century novel "Frankenstein. Or: the new Prometheus". Starting from empirical investigations among qualitative researchers who use software for working with textual data, they come to the conclusion that the fear of computer programs as a kind of "Frankenstein’s monster" is often overemphasised: in practice researchers tend to cease using a certain package rather than submitting themselves to the logic of a software program totally different from the logic of inquiry these researchers want to employ.

Other concerns mentioned quite frequently in current debates refer to the danger that the use of computers could alienate the researcher from their data (AGAR 1991; SEIDEL 1991; SEIDEL, KELLE 1995). As with the danger of the computer program taking over analysis, this methodological hazard is also often related to coding. Seidel and Kelle argue that the distinction between two different modes of coding is crucial to avoid serious alienation from the data: codes can either have a referential function, that means they represent signposts to certain text passages. Or they can have a factual function, that means they are used to denote certain facts. The first type of coding is characteristic for an open and inductive style of inquiry employed by an interpretive analysis of textual data in the tradition of hermeneutic and interactionist approaches. The second type relates to a deductive style of textual analysis in the tradition of classical content analysis. By using certain procedures within software for textual data management analysts can -- without realising -- confuse the two modes of coding: they can involuntarily switch from using the referential function of codes (that means from collecting text segments that refer in a
broad and general way to a number of somewhat vaguely defined concepts) to treating codes as if they were representations of factual information. Seidel and Kelle call this the danger of *loosing the phenomenon by reifying the codes*: the analyst starts to work exclusively on his or her codes and forgets about the raw data, although the necessary prerequisite for doing so has not yet been fulfilled: there is only a loose coupling between a code and a piece of data instead of a well-defined relation between a code and a phenomenon, since the code was not attached to denote a certain discrete event, incident or fact, but only to inform the analyst that there is interesting information contained in a certain text segment, related to a topic represented by a code. This danger of losing the phenomenon and reifying the codes is especially prevalent with the "hypothesis testing facilities" described in section 3.3: by seeking to "test hypotheses" without having observed the necessary prerequisites, that is by applying strict rules to vague and "fuzzy" codes, researchers can easily produce artefacts.

Thus, enhanced facilities for coding and retrieval offer fascinating new possibilities for analysts to "play" with their data and thereby help to open up new perspectives and to stimulate new insights. But combining methodologies of theory building and of theory testing should not seduce us into simply "mixing" or even "confusing" them.

5. Literature


Editorial
The LSE Methodology Institute’s Discussion Papers are an opportunity for visitors to the School, members of staff, and invited speakers to the Institute seminars to put forward an argument on qualitative methodology. The paper may be at an early stage inviting a swift first round of reviews. The papers are internally reviewed before they are accepted and then distributed within and outside the LSE for further discussions with the authors.
In this series we encourage contributions that propose ideal-typical descriptions of particular procedures for qualitative data collection and/or analysis, be these text, image or sound based. In an ideal world typifying comprises a discussion of

• the underlying concepts,
• the strength and weaknesses of the method,
• its comparison to similar approaches,
• a discussion of good and bad use of the approach through using criteria such as reliability, transparency or others,
• one or two exemplary results obtained with the method,
• around 5,000 words of length

Martin Bauer
London, October 1996

Contact the author
Udo.kelle@uni-vechta.de

Dr Udo Kelle
Institute for Interdisciplinary Gerontology
University of Vechta
PO Box 15 53, D-49364 Vechta
Germany
List of papers in the series

no 1 The Narrative Interview: Comments on a technique of qualitative data collection
(Martin Bauer, LSE, October 1996)

no 2 Determining the Central Nucleus of Social Representations
(Celso Pereira de Sa, Rio de Janeiro, November 1996)

no 3 Word Associations in Questionnaires: A practical guide to design and analysis
(Wolfgang Wagner, University of Linz, Austria, February 1997)

no 4 Computer-Assisted Analysis of Qualitative Data (Udo Kelle, University of Bremen, Germany, August 1997)

no 5 The episodic interview (Uwe Flick, Hannover, November 1997)

no 6 Types of Category in the Analysis of Content (Bradley Franks, LSE, February 1999)

no 7 Counter-Transference in Social Research: Georges Devereux and Beyond
(Alain Giami, INSERM, Paris, June 2001)

no 8 Nomothesis, Ideographia and Bemetology
(Gerhard Fassnacht, University of Bern, Switzerland, June 2004)

no 9 Empirical Phenomenology: An Approach for Qualitative Research
(Patrick Aspers, Dept. of Sociology, Stockholm University, November 2004)