A tlas.ti for qualitative data analysis

BRIGITTE SMIT

BRIGITTE SMIT is a lecturer in the Department of Education Management and Policy Studies at the University of Pretoria. She teaches Comparative Education and Qualitative Research Methodology. She is also a senior researcher in the Centre for Evaluation and Assessment. Her research interests include teachers and change, and the human and emotional side of educational change. She conducts computer-assisted data analysis training workshops in qualitative research for academics and advanced post-graduate students.

Abstract

In this article I discuss a variety of theoretical and conceptual dimensions of Atlas.ti (Muhr, 1994, 1997a, 1997b). I draw on my own Atlas.ti experiences (Smit, 2001), and discuss advantages and disadvantages of using computer-aided qualitative data analysis software (CAQDAS) (Lee & Fielding, 1995). How can the quality of a research project be enhanced and how will the end product be affected? These are some questions I intend to answer.

Firstly, I introduce qualitative data analysis in general. Secondly, I discuss the relevance of computer-aided qualitative data analysis software in qualitative research and how Atlas.ti, focuses on coding procedures, (cf. Miles & Huberman, 1994; Dey, 1993) which supports a grounded theory approach for data analysis. Thirdly, I illustrate some facets of my Atlas.ti project and explain some technical aspects, such as the VISE principles, visualisation, integration, serendipity and exploration as the main strategic modes of operation that may enhance the quality in the data analysis.

Introduction

This article discusses how qualitative data analysis can be supported by utilising computer-aided software, Atlas.ti (Muhr, 1994, 1997a, 1997b). Atlas.ti is a powerful workbench for qualitative data analysis, particularly for large sections of text, visual and audio data. This software offers support to the researcher during the data analysis process, in which where texts are analysed and interpreted using coding and annotating activities. Furthermore, it provides a comprehensive overview of a research project, which is called the Hermeneutic Unit (HU) in Atlas.ti, and it facilitates immediate search and retrieval functions. This programme also has a network-building feature, which allows one to visually connect selected texts, memos, and codes by means of diagrams. In the past researchers have analysed their data by hand, using for instance cut-and-paste and colour coding methods. Today, researchers use data analysis software, which are handy tools for managing and analysing great amounts of raw data. The guiding question in this article is therefore, how can the quality of data analysis be enhanced by using Atlas.ti? In response to this research question, I discuss some general notions regarding qualitative data analysis, then I move on to the computer-aided qualitative data software discussion and how this links to a grounded theory approach. Lastly, I illustrate some facets of my Atlas.ti project (Smit, 2001) and highlight some technical concerns.
Qualitative data analysis

Data analysis in qualitative research is an ongoing, emerging and iterative or non-linear process. Before one begins with an analysis, data are transcribed, which simply means that texts from interviews, observational notes or memos are typed into word processing documents. These transcriptions are then analysed either manually or with computer programmes, such as Atlas.ti. To analyse, literally means to take apart, words, sentences, and paragraphs, which is an important act in the research project to make sense of, interpret and theorise such data. This is done by organising, reducing and describing the data. Schwandt (1997, 4) in this regard maintains that an analysis ought to be rigorous, systematic, disciplined, and carefully methodologically documented. According to Alasuutari (1995, 7) data analysis in qualitative research also refers to "reasoning and argumentation that is not based simply on statistical relations between 'variables', by which certain objects or observation units are described." In other words, when using qualitative analysis as a means to explain or make sense of the inquiry, we do not use as evidence the frequencies or the quantities with which something occurs, but rather elicit meaning from the data, in a systematic, comprehensive and rigorous manner.

Various authors have described these analysis processes – Tesch (1990), Dey (1993), Miles and Huberman (1994, 112), and Silverman (1997, 2000). Tesch (1990, 95) has identified some principles appropriate for most types of qualitative research analysis, which have guided my project (Smit, 2001):

- Qualitative analysis takes place throughout the data collection process. As such the researcher will reflect continuously on impressions, relationships and connections while collecting the data. The search for similarities, differences, categories, themes, concepts and ideas forms part of the continuous process.
- An analysis commences with reading all the data and then dividing the data into smaller more meaningful units.
- Data segments or units are organised into a system that is predominantly derived from the data, which implies that the analysis is inductive.
- The researcher uses comparisons to build and refine categories, to define conceptual similarities, and to discover patterns.
- Categories are flexible and may be modified during the analysis.
- Importantly, the analysis should truly reflect the respondents' perceptions.
- The result of an analysis is a kind of higher-order synthesis in the form of a descriptive picture, patterns or themes, or emerging or substantive theory.

The process of qualitative data analysis is also described in detail by Dey (1993, 10) in terms of meanings, which are mediated through language and action and tied to a particular context. This is contrary to quantitative data, which deal with numbers that may appear powerful. They do, however, mean little if they are not based on meaningful conceptualisations. Dey (ibid.) states that data, which are collected, are "produced" by the researcher. This involves selecting data, the techniques of data collection and the transcriptions through note taking and tape recordings. This will affect what in the end constitutes data for the purpose of research.

As referred to earlier, to analyse literally means to break into bits and pieces, or to break down the data, which Miles and Huberman (1994) label as "coding" and Dey (1993) refers to "categorising". Dey (1993, 30) describes data analysis as "a process of resolving data into its constituent components, to reveal its characteristic elements and structure". One of the aims of
an analysis is to describe the data as well as to describe the objects or events to which the data refer. Sometimes more than descriptions are needed, and interpretations, explanations, or predictions are required. The how, why and what needs to be answered and that is done through the analysis, moving beyond the initial description, transforming the data into something it was not.

Descriptions form the basis for the analysis, and the analysis forms the basis for further description. Data is broken up in order to classify it. Concepts are created in classifying the data; the connections are made between the concepts, which in turn provide the basis for a fresh description. To describe means to set forth in words, to recite the characteristics of a person, object or event. The primary steps in the qualitative analysis are the so-called "thick" (or information rich) description (Denzin & Lincoln, 2000, 15, Merriam, 1998, 29), which includes information on the context of an act – the intentions and meanings that organise action. Furthermore, meaning cannot be reduced entirely to a personal matter. Meaning is inherently ambivalent and context-dependent; hence one cannot rely on the respondents' intentions as an incontestable guide to interpretation. Respondents or participants perceive and define situations – according to their understanding of their own motivations and of the contexts in which they act. Neither motivations nor contexts are self-evident, and allowance has to be made for the usual mix of ignorance and self-deception, delusions, fantasies and even lies. Inconsistencies and contradictions in humans do exist. In addition, social forces such as obsequiousness (obedience) towards power, pressures for conformity, and fears of embarrassment and conflict can also distort behaviour and motivations. Pure rational accounts of respondents' intentions cannot be expected or even hoped for.

Descriptions of meaning are the basis for the analysis and are done by the researcher. The role of the researcher in the qualitative analysis refers particularly to awareness of bias and preconceived ideas, since assumptions may blind the evidence of the data. Significantly, as Dey (1993, 64) argues, "... the danger lies not in having assumptions but in not being aware of them." Dey (1993, 36) also refers to the fact that "... qualitative analysis is usually concerned with how actors define situations, and explain the motives which govern their actions". In analysing these actions, the researcher wants to ensure that this relates to intentions of the actors involved.

In practical terms, once the audio interview recordings of were transcribed into text, the reduction and analysis began. In essence, I read the transcriptions while listening again, edited where necessary, and loaded text into the Atlas.ti computer software. Data were then classified, a process that involved breaking up data into bits and bringing it together again in a new way. This was a process of assigning data to categories or classes and identifying formal connections between them (Dey, 1993, 275). It is an important step in the analysis, for without classifying data there is no way of knowing what is actually analysed and no meaningful comparisons can be made. Classifying data is an integral part of the analysis, which lays the conceptual foundations upon which interpretations – which make action meaningful to others – and explanations are based. Classification is not neutral and it is done for a purpose, guided by the research objectives.

Once the data had been classified, regularities, variations and peculiarities were examined and patterns identified. Dey (1993, 227) defines this as the process of identifying substantive connections by associating categories or linking data. Correlations or relations between different categories can be studied and a picture of the data can be built, which will both be clearer and more complex than the initial impressions. In other words, the different parts of the puzzle can be fitted together. Dey (1993, 94) writes that although people usually think in
generalities, they live in detail. As such, words are employed by people to convey ideas, but when the ideas are grasped they forget the words.

To sum up, the core of qualitative analysis is a twofold task – firstly to select a bit of data and secondly to assign it to a category, a process called coding (Dey, 1993, 57). In essence, any research is an exercise in selection processes, and the researcher needs to realise that the analysis is ultimately concerned with human situations and social processes. The so-called facts are produced through conceptualisation and "facts" per se have been merely manufactured. Facts, therefore, depend on the researcher's perceptions, which are shaped by his/her thinking.

Why Computer-Aided Qualitative Data Analysis Software? (CAQDAS)

I worked with a large amount of unstructured textual data – namely interview data and open-ended questionnaire data – and was faced with what Kelle (1995, 1-17) describes as serious data management problems which could not easily be solved by the use of standard database systems. Although such programmes can be used to fulfil one of the central tasks of qualitative data management – the retrieval of relevant segments of text – they nevertheless impose serious limitations. They require that text segments and coding schemes be defined before the data are entered, which contravenes the inductive categorisation strategy preferred by most qualitative researchers.

Since the mid-eighties a variety of non-formatted textual database systems have been developed for qualitative research. Programmes like the Ethnograph, WINMAX, ATLAS/TI, NUD.IST, NVivo Kwalitan or HyperRESEARCH all use similar data structures to assist the organisation and management of textual data: the addresses (e.g. in terms of line numbers) of text segments (which the researcher can define freely) are stored as pointers together with the names of the codes allocated to these segments. With such software, unstructured textual material can be organised by attaching codes to certain text passages. Richards and Richards (1994, 447) write that computer-aided qualitative data analysis software (CAQDAS) is now widely applied in the qualitative community. The first generation of "code-and-retrieve programmes" only mechanised widely used cut-and-paste or indexing techniques but did not change their underlying logic or offer analytic features, which could not be employed using manual methods. This situation changed as more and more complex coding and retrieval facilities were added to these programmes, which were promoted by their developers as a means of qualitative "theory building" and "hypothesis testing".

Kelle (1995, 62) indicates that the first code-and-retrieve programmes linked codes to text segments by using pointers. Similar data structures can be used to define linkages between codes themselves. Since theoretical categories, such as super codes, can be more or less closely related to the codes used to organise the data material, the idea emerged that the structure of a theory developed in a qualitative project could be represented through a "network" of codes. Code-based theory-builder programmes – for instance NUD.IST, NVivo, Hypersoft or Atlas.ti – contain features, which support the construction of networks of code categories.

To sum up, HyperQual2, Kwalitan and QUALPRO are code-and-retrieve programs. Atlas.ti, HyperRESEARCH, NUD.IST, NVivo, Ethnograph, and winMAX are code-based theory builders. Atlas.ti has an additional feature, a graphical network builder (cf. Weitzman, 2000, 809).
Grounded theory in CAQDAS

Coding played an important part in my analysis, and I needed to establish where and how it originated. Coding as a grounded theory strategy has been incorporated within software applications. This is evident in the close relationship between the processes of coding and the use of computers. According to Lonkila (1995, 42), at the heart of grounded theory is a very detailed and explicit coding of texts. In computer assisted qualitative data analysis, coding is conceived as attaching keywords to text segments. The development of Atlas.ti has been strongly influenced by grounded theory (cf. Muhr, 1994, 1997a, 1997b). This does not imply that this software may only be used in an analysis that uses a grounded theory approach. Acquaintance with grounded theory did, however, facilitate my analysis. According to Coffey, Holbrook and Atkinson (1996), coding should not be overemphasised, considering that a large part of the qualitative research consists of interpretation and hermeneutic analysis. On the one hand, grounded theorising entails far more than only the process of coding; on the other hand, computer aided analysis software can offer far more than a code-and-retrieve function. Facilities such as attaching analytic memoranda to specific points in the text are also offered.

Coding of data is central both to grounded theory and to most of the programmes developed specifically for qualitative analysis. Coding in grounded theory is, however, more complex than just attaching labels to text segments, and isolating and naming categories. In fact, coding means "how to dimensionalise them and discover their conditions, consequences, and associated interactions and strategies. The distinctive feature of coding in grounded theory is striving towards theory building" (cf. Lonkila, 1995, 42). Although my research project (Smit, 2001) does not claim all the tenets of grounded theory – that is, aiming toward theory development – it was important for me to become acquainted with the same theory, particularly since the data analysis was conducted with the computer. Using a programme such as Atlas.ti to conduct the analysis does not necessarily have to be in line with grounded theory methodology, but the theory does explicate detailed procedures for coding and memo writing, facilitating a way of working with the data (cf. Lonkila, 1995, 50).

In my project (Smit, 2001), I used open coding, axial coding and selective coding grounded theory strategies, as described by Strauss and Corbin (1998, 55-143):

Open coding

Open coding refers to naming and categorising phenomena through close examination of the data. Data are broken down into discrete parts, which are compared and questioned with "what, where, how, when and how". In other words, open coding fractures data into concepts and categories. Then data are compared and similar incidents are grouped together and given the same conceptual label. The process of grouping concepts at a higher, more abstract level is termed categorising. Labels are then attached to the segments of texts. Put more accurately, coding "represents the operations by which data are broken down, conceptualised, and put back together in new ways. It is the central process by which theories are built from data" (Strauss & Corbin, 1990, 57). The product of labelling and categorising are concepts, which form the basic building blocks in grounded theory construction. Strauss and Corbin (1998, 120-121) suggest that open coding can be done line-by-line, which is time-consuming but most generative; this is the manner in which I coded. Particularly at the beginning of the research, categories can be quickly generated. Coding can also be done by sentence or paragraph or by perusing the entire document, depending on personal preferences.
Axial coding
Axial coding is that part of the analytic process in which the researcher puts the parts of the data identified and separated in open coding back together in new ways to make connections between categories or the codes. In this way the complexity of the context is brought back into the picture. The focus lies with the relationship between categories or codes. In Atlas.ti, this is referred to as linking codes. According to Strauss and Corbin (1998, 124), axial coding looks at how categories crosscut and link. Categories are related to their subcategories to form more precise and complete explanations of the phenomena. In coding, a category stands for a phenomenon, such as a problem or an issue or an event that has been defined by respondents as being significant. Texts do give clues as to how categories do or do not relate. The actual linking of categories does not take place descriptively but on a conceptual level, which implies that text is converted into concepts. The analysis here takes place on two levels – the actual words used by the respondents and the conceptualisation of these words by the researcher. Through questions such as, where, how, when, why and who relationships can be uncovered among categories. Working with the actual data, relationships may not always be so evident. Linkages between categories may be subtle and implicit or hidden. Therefore it is helpful to use some scheme, a paradigm to sort out and organise emerging connections.

Selective coding
Strauss and Corbin (1998, 143 ff) write that selective coding involves the process of selecting one, main core category and relating the other categories to it. It implies the process of integrating and refining categories. In Atlas.ti, I was able to create code families and also to rename codes, redefining codes in other words.

Memos
Punch (1998, 206-208) explains that memos assist the researcher to think, to make decisions or to interpret while analysing the data. Throughout the process of data analysis it is a good idea to engage in memoing, i.e. recording reflective notes about what one is learning from the data. Code memos, theoretical memos and operational memos can be distinguished. Code memos relate to open coding, whereas theoretical memos relate to axial and selective coding. Operational memos contain directions relating to the evolving research design. One of the helpful objects in Atlas.ti is called memos.

An Atlas.ti project
Lee and Fielding (1995, 29-40) comment that some qualitative researchers seem to be reluctant to fully exploit the new possibilities offered by complex coding and retrieval, as investigations among users of CAQDAS show. I found the computer-aided data analysis software extremely helpful. CAQDAS contains strategies for complex retrieval of codes, which can be applied in two different ways: The search for co-occurring codes can be used as a heuristic device. Here the objective is to retrieve the original text to which the co-occurring codes have been attached. The meaning of a certain co-occurrence was investigated by a thorough analysis of the original text. I utilised Atlas.ti “The knowledge workbench” (Muhr, 1994, 1997a, 1997b), which offered the support needed, facilitating activities involved in text analysis and interpretation, particularly selecting, coding, annotating and comparing noteworthy segments. Atlas.ti renders a code-and-retrieve function and provides support for theory building by facilitating connections between codes to develop higher-order classifications and categories, formulating propositions that imply a conceptual structure that fits the data. Although the underlying logic of coding and searching for coded segments differed little from the manual techniques, such as cut–and-paste or colour coding or discourse analysis (in a critical paradigm), the speed and the
comprehensiveness of these searches was an undoubted benefit. Furthermore, the software could cope with multiple and overlapping codes without losing the context. Codes were also combined using Boolean logic forming super codes with operators such as 'and, or, not'.

In my *Atlas.ti* project (Smit, 2001), I worked with 7 interview and questionnaire transcripts, which are called *primary documents* in *Atlas.ti*. I highlighted some 541 text segments, referred to as *quotations*, which yielded 684 labels, which are called *codes* in the software. I then grouped the codes into 16 code clusters, i.e. *code families*. Many *links* were established and a variety of *networks* was created. I would recommend researchers to submit a CD-ROM, as an http file to access the audit trail of the data analysis.

In Box 1, I illustrate a small section of the data analysis and explain the concepts. The *Hermeneutic Unit*, (HU) in *Atlas.ti* refers to the complete project or research for example a thesis or a dissertation. The file reference indicates the location where the project is saved. The word *Super* refers to the researcher who actually does the analysis, and time and date are given for further reference. *Codes-quotations list* means that this particular information shows a particular *code*, with the relevant *quotation*, that is the verbatim evidence given by the respondent. The *code-filter: PT* shows that this particular list was filtered by using all the *primary text*, also referred to *primary documents*, which simply means all the interviews. P1 would then represent the first interview. 1:7 stands for the first interview, 7th code, and quotation in line 54-58, coded by the Super, the researcher. The {1-1} refers the number of codes, and how often this code has been linked to another.

**Box 1: Data analysis and concept explaining**

```
HU: PhD Education policy Change
File: [C:\Program Files\Scientific Software\ATLAS\TEXTBANK\PhD Education policy Change]
Edited by: Super
Date/Time: 12/18/00 04:56:11 PM

----------------------------------------
Codes-quotations list
Code-Filter: PT

Code: effect: (discipline) loopholes for the lazy teacher {1-1}

P 1: INTerview1atlas.txt - 1:7 (54:58) (Super)
Codes: [effect: (discipline) loopholes for the lazy teacher] [effect: all will pass irrespective of knowledge levels]

... some loopholes for a lazy teacher, very much so, because it does not go out of the work that is done, it goes about ...
... and each child is on a different level and it does not matter if this child only knows one thing about water and the other child knows five things, they are both a pass.
```

Figure 1 illustrates a view of a focused network on emotions and feelings that are linked to change. Box 2 shows coded data from an interview.
**Primary documents: ONE**

*File name: file://c:/my documents/interview1atlas.txt*  
87 quotations

**Box 2 Example of coded data from an interview**

**Codes (116):**

- effect: (discipline) loopholes for the lazy teacher, effect: all will pass irrespective of knowledge levels, effect: slack attitude, effect: teacher can get away with a lot, effect: we do not know what is expected, emotional response: teacher is under pressure, emotional response: teachers are overloaded, emotional response: teachers are uptight, emotional response: breaks down the relationship between teachers and children, emotional response: careful responses since it could cost my job, emotional response: classroom is a nightmare, emotional response: despondency: do less, because I cannot win, emotional response: gets your back up a little bit, emotional response: I have a definite problem with it, emotional response: I have never felt this before, emotional response: imposing, emotional response: insecure of what will happen to our jobs, emotional response: it makes me uncomfortable, emotional response: job: will I have one next year, emotional response: little information is coming through, emotional response: lots of stress, emotional response: my rebellion, don't impose the structure at the cost of teacher uniqueness, emotional response: older staff feel threatened, emotional response: some older staff are keen to get new ideas, emotional response: stress becomes visible through lack of enthusiasm, emotional response: teacher does not feel professional, emotional response: teacher uniqueness is ignored, emotional response: teachers are under pressure, emotional response: tension and pressure, emotional response: told not to worry; yet still fear, emotional response: to p: do not more than I have to, emotional response: unsure of what is to happen to their school, emotional response: we are not very sure about our situation, emotional response: what does the future hold for us in 5 years, emotional response: will I be able to cope?, feeling insufficient, ft: representative to workshop, ft: teachers are not well equipped, ft: the representative workshops with staff, ft: workshop had some good ideas, ft: workshop on 2005 is only information, ft: workshop on 2005 was attended, ft: workshop was brief, ft: workshop was one day only, ft: workshoped the workshop with staff, further training: I had good training in group work, further training: no communication after the workshop, government to pay salaries, group work is okay, group work: before or after something new, group work: brainstorming, group work: difficult, group work: independent work in different groups, group work: individual is still important, group work: never in the middle of something, group work: noisy, group work: older staff perceive this as play, group work: research activities, group work: sharing, group work: we see brainstorming in our groups, information comes via other schools, information from Gauteng Department of Education, information from other schools cause unconscious forms of stress ...
A tlasti and the "VISE" principle: Textual and conceptual analysis

As explained in the previous section, grounded theory has exerted a particularly strong influence on the qualitative analysis programmes Atlas.ti, Nud.ist, and NVivo, which can be seen in the structure of the programmes. According to Pandit (1996), there are two modes of data analysis within Atlas.ti, namely on the textual level and on the conceptual level. The textual level focuses on the raw data and includes activities such as text segmentation, coding and memo writing. The conceptual level focuses on framework-building activities such as interrelating codes, concepts and categories to form theoretical networks.

The methodological details regarding the analysis of a project could be available as an addendum or on CD-ROM to facilitate access to the raw data, which is often difficult to trace after the transformation process into a final presentation. Coded interview data, code lists, networks, code families and memos of textual and conceptual levels can be accessed via an assigned CD-ROM. A demo version of Atlas.ti can be submitted with project, or the web browser at <http://www.atlasti.de> can be accessed in order to explore the data, using the software.

The last components, which need clarification, are the main principles of the Atlas.ti methodology and are termed "VISE", which refers to Visualisation, Integration, Serendipity and Exploration.

Visualisation in Atlas.ti refers to the direct supportive role of the way humans think or plan. Complex properties and relations between objects are visual and keep the researcher focused on the data. Although the researcher works with the detail, the integrated whole of the project is always within direct reach, in the "Hermeneutic Unit". Serendipity stands for an intuitive approach to the data, browsing through the data as the researcher makes relevant discoveries, but without a forceful search. The process of getting acquainted with the data uses an exploratory, discovery-oriented approach.

Transcribed texts are opened in a "hermeneutic unit" where all the data, codes, memos and diagrams that belong to the analysis are stored. An analysis commences on the "textual level", which implies that the researcher works mainly with the texts or documents. In open coding text segments are marked, codes are assigned and memos written. Explanatory commentaries may also be written. Lists of codes and memos may be sorted in a variety of ways. One way to sort codes or memos may be according to groundedness, which means according to the series of text passages assigned to a code or memo. Another way to sort codes may be according to the conceptual density, which relates to the number of other codes connected with or linked to a particular code.

The researcher who works mainly on the conceptual level – that means mainly with concepts – uses axial and selective coding. In Atlas.ti, this indicates that codes and memos are joined to families. For theory building the researcher defines concepts consisting of codes of higher order – cf. Atlas.ti "super codes" – which are not connected to particular text passages, but to codes. Relations between codes may be represented graphically and can be defined or redefined according to standard logic relations (cf. Muhr, 1997a).

Concluding thoughts

Since hardly any educational qualitative inquiry in South Africa that I am currently aware of has been supported by Atlas.ti, some background of its philosophy and reasoning for implementation appears relevant in order to grasp the rationale for its usage in data analysis. Using computer-aided text analysis, the researcher needs to appreciate that computers are not
capable of comprehending or discerning the meaning of words or constructs. Their real strength and contribution lie in ordering, structuring, retrieving and visualising tasks. In addition, the computer can create order out of a mass of field notes, interviews, codes, concepts and memos. Put simply, computer software can provide assistance to analyse data, but it cannot do the analysis for the researcher (cf. Weitzman, 2000, 805).

References


Tesch, R. 1990. *Qualitative research: Analysis types and software tools.* London: Falmer.
