2010

The CAQDA Paradox: A divergence between research method and analytical tool

Michael Jones
University of Wollongong, mjones@uow.edu.au

Kieren Diment
University of Wollongong, kdiment@uow.edu.au

The CAQDA Paradox: A divergence between research method and analytical tool

Abstract

A wide range of software tools are available to assist researchers with the process of qualitative data analysis. These include tools that emphasise manual handling of data, (e.g. NVivo, Atlas.ti) and tools that provide some automated analysis based on statistical properties of texts (e.g. Leximancer). These tools are enhancing research, making research activities less complex and tedious, and rendering the process more transparent and portable (Dohan et al. 1998; Welsh 2002; Andrew et al. 2007; Jones 2007). The use of these tools in published works over the last five to ten years has become increasingly more evident. However, in many cases, this increase in frequency of use is also masking the actual method of research. Many researchers who use terms like "Data were analysed using NVivo" are using their chosen analytical package as a proxy for actual embedded methods of analysis. It is possible therefore that Computer-Assisted Qualitative Data Analysis (CAQDA) tools are becoming a substitute for actual, and perhaps valid, techniques for research, analysis and discovery. This paper investigates the extent of this problem, examining CAQDA based papers which have been published over the last five years and reporting on their use, or misuse, of methodology. Further, this paper proposes a solution to the problem by adopting a CAQDA technique which utilises a generic style of methodology. A tool used by Quantitative researchers, known as 'R', is available which is a free, open source statistical programming language. Within the last five years R has become the lingua franca for statisticians and applied workers to publish reference implementations for novel quantitative techniques. No such tool with sufficient flexibility exists for qualitative researchers. We describe the initial development of a transparent file format and research process which keeps the researcher close to the data and provides strong safeguards against accidental data alteration. This has two main effects. The transparency of the file format keeps the researcher close to the data, and ensures that the researcher keeps in mind the process used to analyse the data rather than the tool in use. The second effect, also related to the open source, transparent plain text basis of the tool, means that an environment for fostering innovation in qualitative data analysis can be easily provided and freely distributed among workers in the field.

The CAQDA Paradox:

A divergence between research method and analytical tool

Michael Jones and Kieren Diment
Faculty of Commerce,
University of Wollongong,
NSW 2522, Australia
mjones@uow.edu.au

Abstract

A wide range of software tools are available to assist researchers with the process of qualitative data analysis. These include tools that emphasise manual handling of data, (e.g. NVivo, Atlas.ti) and tools that provide some automated analysis based on statistical properties of texts (e.g. Leximancer). These tools are enhancing research, making research activities less complex and tedious, and rendering the process more transparent and portable (Dohan et al. 1998; Welsh 2002; Andrew et al. 2007; Jones 2007). The use of these tools in published works over the last five to ten years has become increasingly more evident. However, in many cases, this increase in frequency of use is also masking the actual method of research. Many researchers who use terms like "Data were analysed using NVivo" are using their chosen analytical package as a proxy for actual embedded methods of analysis. It is possible therefore that Computer-Assisted Qualitative Data Analysis (CAQDA) tools are becoming a substitute for actual, and perhaps valid, techniques for research, analysis and discovery.

This paper investigates the extent of this problem, examining CAQDA based papers which have been published over the last five years and reporting on their use, or misuse, of methodology. Further, this paper proposes a solution to the problem by adopting a CAQDA technique which utilises a generic style of methodology. A tool used by Quantitative researchers, known as 'R', is available which is a free, open source statistical programming language. Within the last five years R has become the *lingua franca* for statisticians and applied workers to publish reference implementations for novel quantitative techniques. No such tool with sufficient flexibility exists for qualitative researchers. We describe the initial development of a transparent file format and research process which keeps the researcher close to the data and provides strong safeguards against accidental data alteration. This has two main effects. The transparency of the file format keeps the researcher close to the data, and ensures that the researcher keeps in mind the process used to analyse the data rather than the tool in use. The second effect, also related to the open source, transparent plain text basis of the tool, means that an environment for fostering innovation in qualitative data analysis can be easily provided and freely distributed among workers in the field.

1. Introduction

There is wide ranging, long held, debate on the assessment of quality on qualitative methods. This is well positioned by Rolfe (2006, 304):

> [T]his issue can be broadly divided into three positions: those writers who wish qualitative research to be judged according to the same criteria as quantitative research; those who believe that a different set of criteria is required; and those who question the appropriateness of any predetermined criteria for judging qualitative research. Of the three positions, the second appears to have generated most debate …

While acknowledging this incertitude, this paper looks at the problem of quality in qualitative research from a fresh perspective. Regardless of what position people take, and indeed, whether there should be assessment criteria for qualitative research, we believe, if a researcher states they are using a particular method, then they should demonstrate with a relevant degree of rigour, that they are in fact using that method for research and analysis. This extends to all types of qualitative research, and in the case of CAQDA, to the tool a person uses for analysis. A tool should not be a proxy for a valid method of analysis. Rather, the tool is merely an extension of the analyst, designed to augment analysis. Researchers must treat their research and analysis according to their prescribed methodology. We posit that this is not always the case. Consequently, we have developed two hypotheses which guide the discussion and analysis in this paper.

1. The primary focus of this paper is to gauge the integrity of research by authors who use CAQDA for analysis. Our claim is that many of these researchers are not using a clear or well articulated methodology for research or analysis. Instead, they are offering the analytical tool as a means of analysis/research, and not appropriately explaining what they have actually done. Our primary hypothesis is:

   H_1: a significantly large proportion of qualitative researchers using CAQDA use their tool as a proxy for actual methods of research or analysis.

2. The second hypothesis proposes that, in a large proportion of cases, researchers do not adequately explain how they are using their CAQDA tool and/or what value the tools adds to analysis. Instead, researchers tend to proceed on the assumption that readers are fluent in the tools they use and the benefits or deficits its use provides. Our secondary hypothesis is:

   H_2: a significantly large percentage of qualitative researchers do not explain, with sufficient rigor, the application and use of their CAQDA tool.

The premise upon which we will make this measurement will be to examine the espoused method of research and analysis, and using content analysis (Krippendorff, 2004) and constant comparison (Glaser & Strauss, 1967), assessing 325 related articles to deduce whether a valid method is used, and to what degree of rigour it is applied.

This research is important because it questions the assumptions and practices of qualitative researchers. It provides a reality check on the quality of research among our peers, and it flags what we allege are the diminishing standards among some researchers, and the consequential effect this has on the reputation of the field and supporting journals. Seale (1999b) supports this proposition and observes a decline in adhesion to philosophical foundations by researchers who trade quality for efficiency. This, if left unchecked, will lead to a spiralling decline in methodological detail resulting in poor research practice, which will ultimately have a debilitating effect on the whole qualitative paradigm.

A second element of the paper will follow this analysis. This section will attempt to outline an initial solution to the problem identified here by endeavouring to remove the bond between the research method and the CAQDA software. Using this tool, tentatively named "TranscriptMiner", researchers are able to embrace a tool which is a-methodological.

| Deleted: provide |
| Deleted: R |

2. Studies on the quality of qualitative analysis

There are many books and articles which espouse how a researcher should conduct qualitative research. Many of these detail, to differing degrees, how a researcher may validate their qualitative approach to attain a level of quality. However, there are few studies which inquire about the actual quality of work conducted, and where flaws are likely to be found. This gap in the literature is even more profound in business and management research.

In a review of several frameworks for assurance of quality, Walsh and Downe (2006, 113) found that researchers must be both specific and explicit with regard to the data they wish to collect and the method they use:

> [S]pecific methods have evolved with different emphases that are particularly suited to particular spheres of investigation. If the culture of an environment is being explored, then ethnography is most appropriate as method. If the focus is on an in-depth exploration of subjective experience, then phenomenology would be suitable. If 'talk' or dialogue is under scrutiny, then discourse analysis is indicated. Where the nature of the particular method used is not recognised by the researchers, there is a risk of a certain fuzziness that may extend to data collection methods and analysis.

Looking at analysis alone and given the variety and divergence of methods available, it is sometimes difficult to document every step (Walsh & Downe, 2006). Qualitative researchers deserve and enjoy liberal amounts of latitude in expressing their research. However, there are some areas of explanation which we find are generic to most methods and should be provided to assure readers a level of confidence. The following is a list of measures we have used to qualify the studies in this research. These measures for confidence of analysis are derived from a number of studies: (Rolfe, 2006; Seale, 1999a, 1999b; Sitzia, 1999; Walsh & Downe, 2006).

1. How explicit is the research approach (e.g. thematic distillation, constant comparative method, grounded theory)?

    This criterion assesses the extent of discussion and justification which addresses the choice and use of method. It is expected that authors cite leading

theorists, present discussion couched in a language appropriate for the method, and make specific reference to how the method is used in their research. For example, researchers using grounded theory would be expected to cite Glaser, or Strauss or Charmaz, etc. They would be expected to at least mention method-specific terms like selective coding, or axial coding, etc.

2. How and why is CAQDA used?

   When analytical tools are used there needs to be a justification of why they are used and an explanation of how the tool has been implemented. It is not sufficient to merely state that a tool is used. For example a researcher who has made use of NVivo would be expected to discuss how coding was conducted. Researchers may also discuss how memos were used. Modelling may also be included.

3. How clearly are coding systems developed and explained? How systematically are these used?

   Invariably, qualitative researchers collect and compare data using a prescribed unit of analysis and a defined schema for coding and comparison. While the systems adopted will be specific to their actual methodology, there will be a system. This system must be made explicit, and discussion must be presented which aims to assure readers of rigid and consistent application of the analytical approach.

4. How well discussed and defined are findings of themes, concepts and categories? Is there sufficient evidence provided to support their conception?

   Themes, concepts and categories (*research elements*) can be arrived at internally following an emergent process from the data analysis, or they can be applied externally, from some previous conceptualisation, usually from literature, a previous study, or the research instrument. Either of these positions is valid. However, readers cannot be expected to take for granted that this is how these research elements have arrived, and that they are in fact representative of the data under analysis. Authors must present their logic, and if possible, show an audit trail which illustrates how the research elements have been conceived and how representative they are.

5. To what degree is anecdotal support of research findings included?

   The problem with authors providing anecdotal support is that the inclusion of quotes from research consumes word count, and this is not always affordable with journal publication. However, such an inclusion provides factual substantiation of the research data and the connection between analytical abstraction and empirical evidence. Papers which provide only high level conceptualisation risk losing contact with their data, and as such, forcing the reader to make their own assumptions about the supposed connection. Good empirically driven papers will have a balanced approach to abstraction and empirical substantiation.

These are included here as the bare minimum that researchers must address. There are many other factors that should also be discussed, for example: Is the research method

congruent with the data being collected and the situation under study? How does the researcher ensure that context is retained as a fundamental component of the coding? Does the author provide a balanced perspective? What sampling strategy was adopted? These many additional permutations have not been included here. Instead, we have focused on the core criteria. The imposition of which we feel does not decrease the methodological flexibility that is characteristic of qualitative analysis.

It is acknowledged that authors are often made to compromise content on the basis of word count. In such cases, communication of results may take a higher priority than justifying a method. So perhaps the baton of responsibility must pass from the author to the journal editors who must ensure that on balance both results and findings, and method are explicated in sufficient detail and quality (Walsh & Downe, 2006).

3. Method

Three databases – 1) Proquest Central, 2) Proquest Asian Business and Reference, and 3) Proquest European Business – were queried to search for all business and management related articles. The selection of articles which were relevant to this research was filtered according to two lists. Table 1 shows the business keywords that were used, and Table 2 shows the list of CAQDA tools that were sought. Some of these tools were removed from the list of search terms due to their similarity to common words with different meanings. They were removed from the search because of the spurious results they affected. These omitted tools are listed in Table 3.

Administration
Organisation/Organization
Leadership
Management
Business

Table 1 – Business keywords used for search terms.

NVivo	NUD.IST	Atlas.ti
MaxQDA	Qualrus	Transana
askSam	Folio Views	MetaDesign
AQUAD	MECA	Inspiration
SemNet	HyperRESEARCH	QCA
GOFER	Metamorph	Orbis
Sonar Professional	The Text Collector	WordCruncher
ZyINDEX	COMPUTER MAX	QUALPRO
HyperQual2	Kwalitan	Leximancer

Table 2 – Names of CAQDA tools which were used for search terms.

The Ethnograph	
Tabletop	
ETHNO	
Martin	

Table 3 – CAQDA tools which were excluded from the search.

Each database was searched according to the resulting combination of search criteria. For example: Proquest Central + Administration + NVivo. The combination of these unique search terms resulted in 405 individual searches. The accumulated results from all searches yielded 587 articles. This number was further reduced after reading, and manually screening, each article to ensure a fit with the research schema. The final number of articles that were included in this study was 325.

Each article was then read and analysed to quantify its quality, using a 5-point Likert scale to rank the paper based on the five criteria discussed above and according to the two research hypotheses. For example if a method is merely motioned it would be allotted a score of 1. If all of the criteria were discussed and the method properly introduced and explained, then a score of 5 would be allocated. Where the method or analysis included quantitative methods, only those aspects relating to qualitative were evaluated.

The following section characterises the data and presents findings relevant to the hypotheses.

4. Findings and Discussion

Characteristics of the data
The journal articles were drawn from the years 2005 to 2009, with the majority published in 2007 and 2008. Table 4 details the breakdown.

Year	Count	%
2005	55	17%
2006	58	18%
2007	97	30%
2008	92	28%
2009	23	7%
	325	100%

Table 4 – Breakdown of Journal Articles.

In total there were 722 authors. The majority of these came from the UK (27%), followed by the USA (26%) and Australia/New Zealand (21%). The complete breakdown is graphed in Figure 1.
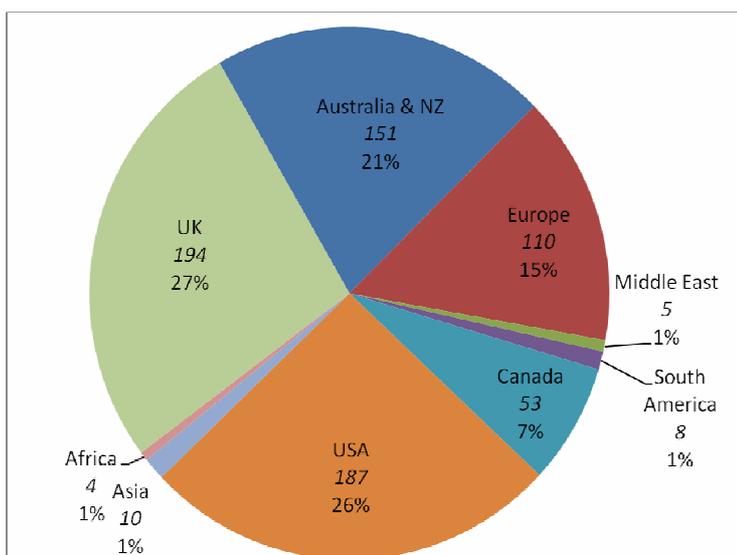
Figure 1 – Breakdown by Author Location (Country/Region *Count* percentage).

The methods of analysis used in these papers were predominantly case study (28%) and grounded theory (20%). However, a shockingly large proportion of papers (21%) did not explicitly state a method, or provide sufficient description of their method for the reader to deduce a valid research approach for the paper. The analytical software of choice by the majority of authors was NVivo or Nudist with 72% of the paper count. Tables 5 and 6 show these breakdown statistics. Table 7 provides a cross-tabulation between the software and the method of research showing what software is most commonly associated with which method. From this analysis it seems Atlas.ti may be more popular for use with grounded theory and NVivo more popular for case study analysis.

Case Study	Content Analysis	Ethno-graphy	Grounded Theory	Phenom-enology	Thematic analysis	Unclear	Other
28%	7%	6%	20%	3%	5%	21%	9%

Table 5 – Research Methods used.

Atlas.ti	HyperRESEARCH	Leximancer	MaxQDA	NVivo/Nudist	Other
20%	2%	2%	2%	70%	2%

Table 6 – CAQDA Software used.

	Case Study	Content Analysis	Ethno-graphy	Grounded Theory	Phenom-enology	Thematic analysis	
Atlas.ti	8.6%	0.9%	4.1%	10.0%	1.4%	1.8%	**26.8%**
HyperRESEARCH	-	1.4%	-	0.5%	-	-	**1.8%**
Leximancer	0.9%	-	-	-	0.5%	-	**1.4%**
MaxQDA	0.9%	-	0.5%	0.5%	-	-	**1.8%**
Nudist	7.7%	3.2%	0.5%	4.1%	-	1.4%	**16.8%**
NVivo	22.3%	5.9%	3.6%	12.3%	3.2%	4.1%	**51.4%**
	40.5%	**11.4%**	**8.6%**	**27.3%**	**5.0%**	**7.3%**	**100%**

Table 7 – A Cross Tabulation showing the correlation between research method and CAQDA software. *'Unclear' methods and 'Other' categories of method and CAQDA have been omitted.*

Hypothesis 1 – do researchers use CAQDA as a proxy for actual methods of research?
Analysis finds that in this sample of 325 articles, researchers do not provide sufficient explanation of the research method they have used. In total, 236 papers (73%) were allocated a Likert score of 1 or 2, meaning that the author(s) had not sufficiently explained their method of research or analysis. Only 50 articles (15%) were considered to have appropriately addressed this element of the paper.

Hypothesis 2 – do researchers rigorously explain their use of the CAQDA tool?
Similarly, analysis on the second hypothesis finds that researchers do not provide sufficient explanation of how they use their CAQDA tool, or what value it adds to the research and analysis. In total, 234 papers (72%) scored a Likert value of only 1 or 2. Only a minority of papers in the sample 37 (11%) actually discussed how their tool worked and its various nuances.

Discussion
Most papers typically pass off their poor methodological discussion using general terms and language. For example the following extracts illustrate popular approaches used to veil a valid discussion on method. In each case, these extracts comprise the entire methodological discussion. *In order to conceal the identity of these authors these extracts have not been referenced, however a list of references can be provided by the primary author on request.*

1. *The QSR NUD*IST Vivo (NVivo) software package was used to manage the data. Using NVivo, transcripts were coded according to themes and analysed using a constant comparison approach (Glaser 1992). Phenomena were labelled and categories were discovered which were then analysed in terms of their properties and dimensions. From these concepts theory is able to be built (Strauss & Corbin 1990). Code notes were written from the open coding procedure; these were our initial thoughts about important themes, and possible relationships and issues that seemed important to the participants.*

   The adoption of a grounded theory approach is implicit in this statement, and is drawn from references to Glaser and Strauss & Corbin. However, the actual process of analysis is vague. A person unfamiliar with grounded theory would

not understand the term 'constant comparison' and there is no clear connection between *theory building* and *Strauss & Corbin*, nor how theory is actually built. In this case it is not improbable that an unweary reader could be led to believe that NVivo is the method of analysis.

2. *Using atlas.ti software, the interviews were catalogued, cleansed, and qualitatively analysed. The data confirmed ...*
   The project managers told stories about their knowledge and learning experiences in projects. From these, it was possible to extract common concepts and solutions that they deployed and to aggregate these into principles for practitioners.

   The reader is left making many assumptions here. Interview data, it seems, undergoes a form of thematic analysis through the intervention of Atlas.ti. Just how this transformation happens is not clear from this description of method. In this case (1) the research approach is not explicit, (2) there is no mention of why and how Atlas.ti is used, (3) there is no explanation of a coding system, and (4) while themes are somewhat discussed elsewhere in the paper, no explanation of arrival is given.

3. *Initial analysis of the data into rudimentary categorizations was performed using Atlas.ti software. The research results were then further distilled through subsequent discussions among members of the research team, re-examination of the transcript data and reorganization of the initial theme categories. Thematic development was a fluid process, involving ongoing reflection on the contents of the transcripts, inclusion of narrative accounts directly from the interview transcripts, and the search for deeper meanings and linkages of the emergent themes in light of the researchers' own ...*
   backgrounds and other relevant literature.

   Here too, the CAQDA software seems to do all of the work. The method, which is espoused to be thematic analysis, is barely referred to. A person who is familiar with the method is likely to infer the information needed to understand what is going on. However, we as authors must be prepared to appeal to a wider audience. We should not affect a language which works to alienate our readership. This is a problem realised with this article.

In the next section of the paper we introduce a potential solution to this problem. This tool, provisionally named TranscriptMiner provides a means to achieve an a-methodological approach to qualitative research. Use of TranscriptMiner would remove the presumption of CAQDA having an embedded methodology, just as the use of specific general purpose statistical software packages are not presumed to particular quantitative methods.

Deleted: R

Deleted: R

Deleted: . J

Deleted: SPSS or Excel

Deleted: is

Deleted: to be attached

Deleted: a

5. Why do we need a comprehensive open source qualitative data analysis tool?

A major development in the world of quantitative analysis in recent years has been the development of the R project (Ihaka & Gentleman, 1996), which is an open source

programming language for statistics. Within the last 5 years, R has moved from being a relatively obscure open source project to the main way of providing reference implementations for novel statistical techniques. Along the way it has been widely adopted in industry, especially in the areas of biotechnology and finance (F. Leisch, Core R developer, personal communication). The open source nature (where the software can be freely redistributed on the condition that the source code must be provided by anyone redistributing the software) of this project is extremely important as it provides researchers with a level playing field, and a standardised approach to foster innovation, and provide a standardised framework for replicable and verifiable methodology.

At present there is no equivalent open source project which aims to provide an equivalent facility for qualitative data analysis despite some small scale or specialised software interface for code and retrieve analysis or content analysis (see http://en.wikipedia.org/wiki/Computer_Assisted_Qualitative_Data_Analysis_Software for the very short list). However none of these projects are of sufficiently ambitious scale to be able to replace or enhance existing commercial qualitative data analysis tools. Perhaps one of the reasons for this is that the open source projects either focus on a graphical user interface for code and retrieve processes and/or focusing on a single specialised aspect of the work. However, the real problem that needs to be solved in order to provide comprehensive infrastructure for open source qualitative data analysis tools is providing an appropriate file format, and the associated data structures to provide a framework for interrogating the project's corpus.

The well known qualitative data analysis software packages are exclusively commercial, closed source off the shelf software packages. To the best of our knowledge almost all packages use closed proprietary file formats, and interoperability between different packages is mostly unavailable, and/or discouraged. As a result of this, changing qualitative analysis packages is hard, and export from one package to another package, or just to home grown tools will almost certainly result in data loss. This barrier to use of alternative software causes significant inertia, reduces researcher freedoms, and may cause reduced detail of reporting of methodology as detailed in the first part of this paper. One possible exception to this is the XML export facility provided by Atlas.ti which claims that a project's complete data can be exported to XML. However, as no other qualitative package supports this format yet, this interoperability is limited. Additionally, for technical reasons relating to the XML specification (especially the need for balanced tags), XML will never be a good file format for simultaneously human and computer readable file formats for qualitative data analysis. The reason that this is important is described below.

A simple minimalist architecture
Rather than XML, we propose that a combination of SGML (Standardised General Markup Language – a superset of XML), standard text based configuration files, and use of the standard features of a modern operating system's file system. Together, these can provide the basis for the data structures needed for computer assisted qualitative data analysis.

The goal of the development of a standard set of tools for computer assisted qualitative data analysis should be to provide a human readable and writable file format that can be used at the level of a reasonably computer literate researcher. This

enables the user interface portion of the tools to be taken care of by existing tools, and the developers can then concentrate exclusively on the logic required for data entry and data retrieval. The requirement for a minimal level of computer literacy is especially important for this cohort, as there is little or no history of computer programming skills being widespread among qualitative researchers. We describe the structure developed to date which provides this simple structure for qualitative data analysis.

The SGML based file format is quite simple. Codes are placed in curly brackets as follows:

```
{optional_label:code_name}Text being coded is here
{/optional_label:code_name}
```

A more concrete example is in this fictional conversation between Fred Flintstone and Barney Rubble:

```
Fred:

{q:how}Hi Barney, how are you?

Barney:

I'm pretty good Fred{/q:how}, {t:activity}have you been to work
today at all? {/t:activity}
```

Where the "q:" prefix indicates a question, and the "t:" prefix indicates a theme. It's then a simple matter to key each theme and question in a text file. For example the questions text file might look like this:

```
Smalltalk {q:smalltalk}

    How are you? {q:how}

    What do you think of the weather? {q:weather}

Bigger questions {q:big}

    What is the meaning of life? {q:life}

    What is the airspeed velocity of an unladen swallow?
{q:swallow}
```

A corresponding themes file would key the themes in a separate file. This is clearly a very light weight representation of the kind of code tree structures found in commercial qualitative data analysis software. By using a programmer's text editor, we can have the relevant coding files on screen at the same time as the transcript file, and record keyboard macros which mean that coding can be done extremely quickly while minimising the risk of error.

As well as this coding, we need to be able to specify searches and retrieve from them. To this end the second author has developed a set of programmer's libraries, and a web application to provide document search and code retrieval functionality. The

software library is available from http://github.com/singingfish/Text-TranscriptMiner, and the web based search interface is available from http://github.com/singingfish/Text-TranscriptMiner-Web. The programming language Perl has been chosen for this work due to its flexibility, very large library support and particular strengths with text processing. The programmer's library is backed by automated tests which mean that as well as ensuring the reliability of the software with text data, the same testing framework, which is based on receiving expected output from known input, new tests can be written to describe particular research procedures.

The web application is intended to run on the local machine, not to be accessible over the internet. The point of the search and retrieve process is to be able to provide a unique "address" for each chunk of coded document. This along with the use of version control (see below) provides a convenient mechanism for relating memos to specific parts of the document. With the search function being web based, it's quite possible to use links to ensure that memos for example are closely bound to the specific parts of the documents that they arise from. Clearly when we are looking at the links between documents, HTML in combination with Javascript provides us with a highly structured, well understood document model that provides much pre-existing technology which can be used to help understand the content and structure of a set of documents.

While the library and the web application are in very early stages of development, they illustrate that simple, powerful robust and replicable processes can be implemented very easily to provide an open source framework for qualitative data analysis. As is usual with software projects, the production of quality code is a smaller problem compared to writing end user documentation, and making it simple for external contributors.

Tracking a project's history.
The final piece of the puzzle, and something that is neglected by all other qualitative analysis is the facility to record the history of a project. Fortunately with our emphasis on text based infrastructure, and ready appropriation of programmer's tools, we can use the version control utilities already in use by open source software developers. Without going into detail, this allows us to record the complete history of a project, including transcription, the coding process, and correction of any errors along the way. This provides an audit trail which allows the researcher to understand themselves the evolution of their thought processes during data analysis, as well as a transparent record which can be used by collaborators to understand the process underlying the analysis of the data. This is generally achieved by browsing the differences between one revision and others (the technical term is a *diff*) either examining line by line or word by word differences. As well as this, due to the roots of version control software in the open source community they are designed to be used for the purposes of collaboration, so multiple people can work on the same project (and even the same document) at the same time. Concurrent changes from different users can be merged together, and policies and procedures can be put in place for dealing with conflicts (i.e. two or more researchers making edits to the same portion of the same document). Clearly this approach has much potential to inform and clarify the collaborative process.

Multimedia

We have avoided discussing qualitative analysis of multimedia resources. There are two reasons for this. Firstly, the vast majority of qualitative research summarise aspects of these kinds of sources with text, and the ability to use a good programmers text editor as transcription software by playing embedded audio or video, while providing semi-automatic attribution of speakers and time stamps in the document is very simple. Similarly clickable image maps provide the ability to annotate images, and there is a range of pre-existing open source solutions to perform this kind of task inside the web browser. Therefore the multimedia capabilities of qualitative data analysis software are a secondary consideration. The larger part of the problem is to provide a common-sense structure for data management, which can be simultaneously be used by both humans and researchers.

6. Conclusion

Qualitative analysis runs the risk of losing its position as a valid paradigm of research if researchers don't take care to be rigorous when applying and explaining their research methods, especially when CAQDA are used. The latter case is particularly important because researchers often fall into the trap of using the CAQDA tool as a proxy for actual embedded means of analysis. The research conducted here finds that almost as many as 75% of CAQDA based papers do not explain or apply research methods rigorously enough, and instead rely upon the reader to make metal leaps and assumptions to deduce a method of research and analysis.

As a solution to the problems presented here, we offer TranscriptMiner as a potential analytical framework. One of the benefits of this approach, which is appropriate here, is that TranscriptMiner is a-methodological, and as such cannot create a situation where the software can become a proxy for an actual research method.

Deleted: R

Deleted: R

7. References

Andrew, A. and E. Peter (2007). "Structuring qualitative enquiry in management and organization research." Qualitative Research in Organizations and Management 2(1): 62.

Dohan, D. and M. Sanchez-Jankowski (1998). "Using Computers to Analyze Ethnographic Field Data: Theoretical and Practical Considerations." Annual Review of Sociology 24(1): 477-498.

Glaser, B. G., & Strauss, A. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research. New York: Aldine.

Krippendorff, K. (2004). Content Analysis: An Introduction to Its Methodology. Thousand Oaks, Calif.: Sage Publications.

Jones, M. (2007). "Using Software to Analyse Qualitative Data." Malaysian Journal of Qualitative Research 1(1): 64-76.

R Development Core Team (2008). R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rolfe, G. (2006). Validity, trustworthiness and rigour: quality and the idea of qualitative research. Journal of Advanced Nursing, 53(3), 304-310.

Seale, C. (1999a). Quality in Qualitative Research. Qualitative Inquiry, 5(4), 465-478.

Seale, C. (1999b). The quality of qualitative research. London :: SAGE.

Sitzia, J. (1999). How valid and reliable are patient satisfaction data? An analysis of 195 studies. International Journal of Qualitative Health Care, 11(4), 319-328.

Walsh, D., & Downe, S. (2006). Appraising the quality of qualitative research. Midwifery, 22(2), 108-119.

Welsh, E. (2002). "Dealing with Data: Using NVivo in the Qualitative Data Analysis Process." Forum: Qualitative Social Research 3(2): 1-7.