

## More Than Just Coding? Evaluating CAQDAS in a Discourse Analysis of News Texts

*Katie MacMillan*

**Key words:**

computer-assisted  
qualitative data  
analysis software,  
CAQDAS,  
discourse  
analysis,  
qualitative  
research methods,  
media analysis

**Abstract:** Computer assisted qualitative data analysis software (CAQDAS) is frequently described as a tool that can be used for "qualitative research" in general, with qualitative analysis treated as a "catch-all" homogeneous category. Few studies have detailed its use within specific methods, and even fewer have appraised its value for discourse analysis (DA). While some briefly comment that CAQDAS has technical limitations for discourse analysis, in general, the topic as a whole is given scant attention. Our aim is to investigate whether this limited interest in CAQDAS as a qualitative tool amongst discourse analysts, and in DA as a research method amongst CAQDAS users, is practically based; due to an uncertainty about research methods, including DA; or because of methodological incompatibilities. In order to address these questions, this study is based not only on a review of the literature on CAQDAS and on DA, but also on our own experience as discourse analysts putting some of the main CAQDAS to the test in a media analysis of news texts.

### Table of Contents

- [1. Introduction](#)
- [2. CAQDAS and DA](#)
- [3. A Practical or a Methodological Concern?](#)
- [4. Discourse Analysis: The Methods](#)
  - [4.1 Corpus linguistics, concordances and CAQDAS](#)
  - [4.2 Critical discourse analysis and CAQDAS](#)
  - [4.3 Conversation analysis and CAQDAS](#)
- [5. A Practical Application: Smoking Guns](#)
  - [5.1 Importing news texts](#)
  - [5.2 BOOLEAN searches](#)
  - [5.3 Coding segments](#)
- [6. Discussion: So Far No Smoking Guns](#)
  - [6.1 Methodological compatibility?](#)

### [References](#)

### [Author](#)

### [Citation](#)

## 1. Introduction

The following study emerges from a two year project,<sup>1</sup> with the aim of testing and developing methods for the analysis of media content. The project's overall concern is with the substantive topic of the presentation of news and representation of power, politics, and policy in the media, and it is with this focus

1 This work is supported by the UK Economic and Social Research Council's project H333250014 "Assessment And Development Of New Methods For The Analysis Of Media Content" as part of the ESRC's Research Methods Programme.

that we applied a number of methods, including discourse analysis (DA); content analysis; and frame analysis to media news materials. Given the increasing popularity of computer assisted qualitative data analysis software (CAQDAS) within qualitative research, our interest was also in evaluating its impact on the research community (MACMILLAN & KOENIG, 2004), as well as its value as a research tool for various different analytical methods (KOENIG, 2005). [1]

CAQDAS is frequently described as a tool that can be used for *all* qualitative research in general (RICHARDS, 1995; CARVAJAL, 2002), with qualitative analysis treated by some researchers as a "homogeneous" category (COFFEY et al., 1996) which needs no further explanation. With functions designed to categorise data as though the researcher were building a "theory" of research (RICHARDS & RICHARDS, 1994) through a system of codes, all CAQDAS programmes are compatible with "Grounded Theory" (GT), where theory is that which "emerges" from out of the data (GLASER & STRAUSS, 1967). The majority of practical studies which discuss method at all discuss GT as the framework for a CAQDAS "analysis". Few studies detail using CAQDAS as a tool for other qualitative methods, and fewer still examine its use in discourse analysis. [2]

A small number of descriptions make a passing comment on the lack of importance of qualitative software to discourse analysis. However, apart from suggesting technical limitations within the programmes, they fail to attend to why this might be. Given that we had used CAQDAS with a number of studies, involving the comparative analysis of media news, and including frame analysis and content analysis, we were keen to investigate why so few studies make use of CAQDAS in discourse analysis, and to detail the extent to which using the programmes would help, or hinder analyses. [3]

## 2. CAQDAS and DA

At the beginning of 2004 a question was posted on "*Qual-Software*" (qual-software@jiscmail.ac.uk), an online discussion forum which provides the opportunity for researchers to discuss qualitative software and its use within qualitative analysis. "*Qual-Software*" is a well-established forum, with questions and observations often stimulating informed and lively debate from experienced and "novice" researchers. On this occasion, however, the researcher's question received no online response. Personal communication with the researcher posing the question later confirmed that no responses at all had been received from other list members, either online or off-line. The topic that had met with this unusual silence was whether anyone had any experience of using the qualitative software programme "NVivo" in discourse analysis. [4]

Two months later we asked a similar, although less specific, question to this list. Our concern was to find out whether researchers using computer-assisted qualitative data analysis software had any practical experience of using any of the programmes with discourse analysis. We also posed this question online to members of the Loughborough "Discourse And Rhetoric Group" (DARG)—a large group with a following of discourse analysts both within the UK and abroad—to

judge the extent to which discourse analysts used any kind of CAQDAS in their work. While no one could offer any first hand experience of working with CAQDAS and discourse analysis, the responses we did receive—requests for information of our findings—encouraged us to pursue the matter further. [5]

### 3. A Practical or a Methodological Concern?

According to GIBBS et al. (2002, p.6.33), "[t]here are some forms of qualitative research where there is little use of CAQDAS. This is true of approaches like narrative, conversation analysis, biography and discourse analysis". The reason for this, the authors suggest, is technical, with programmes not sufficiently advanced to handle tasks such as the complex transcription notation needed for some forms of DA. However, as we shall discuss below, since the few discourse analysis studies to use CAQDAS *do* tend to use it mainly for technical and organizational tasks, GIBBS et al's explanation doesn't give us the whole picture. A more substantial explanation would not only have to consider the practical application of CAQDAS, but would also examine the theoretical implications of using a software programme in a fine grained analysis of texts. [6]

Our aim, therefore, was to not only consider the extent to which a limited interest was practically based, but also because of a basic uncertainty about the research methods, including DA, and, more profoundly, because of methodological incompatibilities. Following on from our initial enquiry, we decided to investigate the extent to which CAQDAS can be useful in managing data for a discourse project, by putting it to the test ourselves. [7]

There are a wide variety of software packages for qualitative analysis designed to do the practical tasks of sorting, storing, editing and coding, traditionally done by hand (WEITZMAN, 2000; see also WEITZMAN & MILES, 1995). In general CAQDAS are useful for organising, categorising, and searching data<sup>2</sup>, particularly when this involves the cumbersome process of managing large quantities of text. CAQDAS can, as the acronym suggests, *assist* in research using certain kinds of qualitative methods, most particularly Grounded Theory, and qualitative content analysis.<sup>3</sup> They can also, to a lesser degree, be of use in frame analysis (KOENIG, 2004). They are not, however, able to *do* analysis per se (CARVAJAL, 2002; COFFEY et al., 1996), and they are not suitable for all methods. [8]

Despite this there are still frequent misunderstandings of the role of CAQDAS in research. One of our earlier studies (MACMILLAN & KOENIG, 2004) drew attention to the tendency for both software developers and researchers to be almost entirely uncritical in their appraisal of the software, aside from the usual practical comparisons of type. Unrealistic expectations about what CAQDAS can do tend to contribute to a myth amongst researchers that the programmes are a

---

2 For a description of what CAQDAS can, and cannot do, see [http://www.lboro.ac.uk/research/mmethods/research/software/caqdas\\_primer.html](http://www.lboro.ac.uk/research/mmethods/research/software/caqdas_primer.html). For an evaluation of the main CAQDAS programmes see also <http://www.lboro.ac.uk/research/mmethods/research/software/caqdas.html>.

3 See [http://www.lboro.ac.uk/research/mmethods/research/software/caqdas\\_primer.html](http://www.lboro.ac.uk/research/mmethods/research/software/caqdas_primer.html).

"method" in itself, with little understanding shown of the multiplicity of disciplines within qualitative research. It is hardly surprising, therefore, that there seems to be little known about discourse analysis, let alone an understanding that within this broad area of language study, there is a diversity of approaches with different epistemological roots, and very different methodologies. [9]

#### 4. Discourse Analysis: The Methods

Aside from analysing the use of language in talk and texts, some of the methods within discourse analysis have little else in common, but, in general, can be defined as a "set of methods and theories for investigating language in use and language in social contexts" (WETHERELL et al., 2001b, p.i). Approaches include discursive psychology; conversation analysis and ethnomethodology; critical discourse analysis and critical linguistics; and sociolinguistics.<sup>4</sup> For all methods of DA context is important in collecting and analysing data (although the focus on context varies between approaches). [10]

In discursive psychology the focus is on talk as action (EDWARDS, 1997), rather than a reflection of action. Discursive psychologist Jonathan POTTER (2000, p.31) describes discursive psychology (DP) as focusing "on the production of versions of reality and cognition as parts of practices in natural settings". As such, POTTER argues, it is "offered as one potential successor to cognitivism" (p.31). In a discussion of questions addressed by DP, POTTER explains that the crucial point of interrogation is what cognition *does* (ibid., p.35). In looking at memory, for example, DP is concerned with what memory does in interaction—how a version of the past is constructed in order to sustain an action. [11]

Conversation analysis (CA) has its roots in ethnomethodology (GARFINKEL, 1967), and, broadly, examines the methods people use to make sense of their everyday social world. From this perspective, as with DP, talk is not "merely *about* actions, events and situations, it is also a potent and *constitutive part of* those actions, events and situations" (POTTER & WETHERELL, 1987, p.21). Unlike ethnomethodology, however, CA examines "the minutiae of naturally occurring conversations represented in verbatim transcript" (ibid., p.81), looking at accounts in context, and in terms of sequential organization, in order to identify systematic properties in talk. [12]

For other forms of DA, such as critical linguistics (CL) and critical discourse analysis (CDA), the central concern is with social conditions, rather than discursive action. Roger FOWLER (1991, p.5), in a discussion of the "different goals and procedures" of different branches of linguistics, describes CL as an "enquiry into the relations between signs, meanings and the social and historical conditions which govern the semiotic structure of discourse". CDA is concerned with "understanding the nature of power and dominance" and how "discourse contributes to their production" (VAN DIJK, 2001, pp.301-2; see also

4 See WETHERELL et al. (2001a, p.6) for a description of six "more or less distinct" discourse traditions, as mentioned above, but also including Bakhtinian research; and Foucauldian research.

FAIRCLOUGH, 1995; FOWLER, 1991; VAN DIJK, 1995). Critical linguistics and critical discourse analysis differ from traditional linguistics in that textual context is crucial—with the text "not the sentence (or the word, or the sound)" important as "the basic unit" of analysis (KRESS, 2001, p.35). [13]

Within linguistics different strands of the discipline have different aims and different procedures. Traditional approaches, for example, treat language as a set of precise rules which must be adhered to in order to facilitate efficient communication. This perspective, which builds on existing assumptions about language, focuses on the structure of language units (including sounds), and conventionally involves using invented sentences to illustrate how these rules work—a method which tends to be disconnected from ordinary talk and social context (DE BEAUGRANDE, 1996). [14]

Sociolinguistics, however, studies the social aspects of language, looking at how social relationships are organised linguistically within different communities. This approach looks at different language situations in different cultures and communities (e.g. LABOV, 1972), and seeks to identify patterns or rules within conversation. Studies in sociolinguistics and the ethnography of communication (e.g. GUMPERZ & HYMES, 1972) use data from corpus based materials, questionnaires, and interviews, as well as from natural settings, including tape recorded conversations, and observation. [15]

Most linguistic studies, which use computer analysis programmes tend to use quantitative concordance programmes. The example below (YATES, 2001) is one of the few studies we found which used both a concordance programme to run a statistical count of word use, and CAQDAS to group kinds of talk into various categories. The study is based within sociolinguistics, but draws on a method of linguistics known as corpus linguistics (CL) in the handling of large quantities of data. [16]

#### **4.1 Corpus linguistics, concordances and CAQDAS**

CL is a linguistic approach which, with its interest in counting or measuring linguistic features, lends itself readily to the search, count, and code facilities of computer programmes. Unlike the rationalist approaches of traditional linguistics, CL (as with sociolinguistics) tends to use naturally occurring language data—"based on large samples of language use that the researchers hope are representative of general language practices across a group, culture or even a society" (YATES, 2001, p.94). [17]

The data, the "corpus" of materials, is a body of text frequently available in machine-readable form (MCENERY & WILSON, 1996). In CL "concordance programmes" are used to find words in the context of text segments, to list them in order, and to calculate word frequency. For his study YATES compared his computer-mediated communication (CMC) data, collected from the Open

University computer conferencing system, with the Lancaster-Oslo/Bergen corpus<sup>5</sup> of written British English, and the London- Lund corpus of spoken English<sup>6</sup> to decide whether his own corpus of texts was more like written or spoken language. [18]

In such studies computer programmes are viewed as necessary to handle the "words and phrases in a corpus of several million words" (YATES, 2001, p.111). YATES used a software programme that showed concordances and key-words-in-context (KWIC), displaying lists of all occurrences of target words (and including surrounding words). This way he was able to conclude that CMC, statistically, is more like written than spoken language. [19]

Atlas.ti was used to search for specific key words, and to then code the data, creating categories of identity which were then listed according to gender, age, nationality, assertion and so forth. This study comfortably and unproblematically uses CAQDAS to categorise the use of the terms "I am" and "I'm" within a large corpus of materials. The resulting analysis ranks top categories, comparing various methods of communication in a way that nicely echoes the initial concordance-derived statistical analysis, with computers most usefully employed to list and count word frequencies, and to count word categories. The main function of the CAQDAS here is no different from most of the studies examined—its value to the analysis is in its ability to segment and count the data. [20]

Although Diction, a Windows programme which incorporates dictionaries to search for a number of categories and sub categories (ALEXA & ZUELL, 1999), was designed to be used in the analysis of political speeches, as with the above programmes, it cannot facilitate fine grained analysis.<sup>7</sup> This particular software summarises content by categorising texts into a number of variables, to enable the comparison of data against standardised scores. [21]

Attempts to "link" content analysis and discourse analysis (WILSON, 1993), have so far, unfortunately, been of more benefit to quantitative linguistic analysis than to qualitative discourse analysis. For example, in a discussion on "CLAWS" (the Constituent Likelihood Automatic Word-tagging System), the software developed by Lancaster University, Andrew WILSON describes how the programme can be used to categorise speech in a number of ways, using a probability matrix derived from large bodies of tagged and manually corrected texts. The value for discourse analysis, according to WILSON, is in being able to identify speech modifiers, as well as allowing the identification of attributes assigned by the speaker to an object. This, however, is more of a concern for traditional social psychology, than for methods such as discursive psychology, for which social

---

5 This corpus consists of 500 individual texts distributed across 15 text categories.

6 The L-L corpus consists of 100 spoken texts.

7 Likewise, the now abandoned software CETA (Computer-Assisted Evaluative Text Analysis), although designed to help analyse language, worked by parsing of texts into nuclear sentences, in order to make predications about "meaning objects" (e.g., people, institutions, concepts, events) or about the relationship between meaning objects (see EVANS, 2001; see also CUILENBERG et al., 1988).

texts are in themselves a "potent, action-orientated medium", and should not be treated as a secondary route to "attitudes, events or cognitive processes" (POTTER & WETHERELL, 1987, p.60). From this perspective we cannot unproblematically identify attributes as assigned by the speaker to the object. Descriptions should be examined for what they do, not what they say. [22]

#### 4.2 Critical discourse analysis and CAQDAS

The above discussions, while clearly concerned with the analysis of language, use the software either for content analysis, or quantitative linguistic analyses. As we have stressed, for most DA studies, including CDA, there has been little interest in using CAQDAS. As FOWLER and KRESS (1979, p.198) emphasise in their discussion of examining texts in CDA, any tool, or method, which creates distance by lifting discourse out of context, to consider them in isolation "would be the very antithesis" to approaches within this field. [23]

While this view, according to Gerlinde HARDT-MAUTNER (1995), continues to be valid, and applicable to software programmes use in general, HARDT-MAUTNER argues for an approach which combines the quantitative methods of corpus linguistics with the qualitative methods of CDA, in order to make up for what "is usually lost in terms of breadth" (ibid., p.3) in a qualitative study. HARDT-MAUTNER's thoughtful discussion acknowledges that while coding distances the researcher from the text, gathering quantitative information can be the first step towards providing insights on the research question. Once again, however, there is a clear division between the quantitative and qualitative elements of the study, with software programmes used to measure instances, and remaining separate from a textual analysis where the complete texts are examined in detail for the "full ideological significance" (ibid., p.9) of expressions. [24]

When it comes to analysing discourse, DA, in general it seems, has little need for the support of CAQDAS or other more quantitative programmes. Teun VAN DIJK, editor of "The Handbook of Discourse Analysis" (1985), and author on a number of critical texts on discursive racism, ideology, and the media, argues that while CAQDAS may provide the means for a large scale quantitative search, and quantities of categories, this it is not sufficient in itself to count as a method of analysis, and that a "deep qualitative analysis" on a smaller selection of data will "generally yield much more insight" (*Personal communication*, 2004). [25]

It is important to be aware of the difference between a discourse analysis which uses quantitative methods to count specific lexical items, and a discourse analysis which studies how language works, by examining its use in context. In the above studies software programmes (designed for qualitative or for quantitative purposes) were used alongside a qualitative analysis, or in linguistic analysis to code and count instances of lexical use. None took an entirely qualitative approach to discourse analysis. In the following section we look at how computer programmes are used practically, to manage data, and to provide links between sound files and text, in an entirely qualitative analysis of talk. [26]

### 4.3 Conversation analysis and CAQDAS

With an emphasis on order, in the identification of conversational "devices", and in the "highly organised nature of ordinary talk as sequential social action" (EDWARDS & POTTER, 1992, p.28), conversation analysis is one of the few exceptions to use CAQDAS as part of detailed qualitative analysis. Once again, however, the value of the software is not in assisting the analysis, but in handling the data. Some software programmes, for example, have been used in CA to store and play audio/video recordings in conjunction with transcriptions, and to enable the researcher to move between visual, sound and textual data. [27]

Paul TEN HAVE (1991; 1998) has written a number of practical descriptions of CA, as well as some useful general remarks about CAQDAS. His interest in computer software, qualitative research methods, and the analysis of doctor-patient interaction, led him to examine the role of Ethnograph in taped medical consultations, and while he cautions that the "analysis of conversational materials can never be *automatic*", software can at times be useful in a routine for "*computer assisted analysis*" (1991, p.2, original emphasis). [28]

Ethnograph is designed to manage textual data, including field notes, and transcripts, and, like most software programmes, can be used for data searches, retrieving "segments" of text, coding, and memo writing.<sup>8</sup> TEN HAVE looks at some of the practical considerations of using this software with transcriptions of the detail that CA requires. Although he found serious technical limitations with the particular software he used, and was at times restricted by the task of assigning codes, TEN HAVE's appraisal is fairly positive. He asserts that CAQDAS can offer CA some help in managing a research project, by offering the opportunity to judge whether the analysis has achieved a "satisfactory level of consistency" (ibid., p.8) throughout the process of coding, memo writing, and note taking. Even a major problem with context is treated pragmatically. TEN HAVE found that Ethnograph, as with many software programmes, is unable to allow an analysis of sequential structures, and as such segments of text become fragmented and isolated during the process of coding. In an attempt to solve this, TEN HAVE worked out his own system of coding, and a way of linking each utterance to the context in which it was spoken. [29]

Not all researchers, however, have the time or the expertise to design a system of coding which will maintain a link with the transcript extract and the textual context. A possible solution may lie in the programme chosen to do the task. Sequentiality in context can be maintained, according to Brian TORODE (1998), by Code-a-Text<sup>9</sup>, a software programme which works with both texts and audio-visual recordings. In his analysis of evaluations in narrative story-telling TORODE discusses the reasons why this programme might be particularly appropriate for CA, including its ability to use fully formatted texts which can be edited without disturbing the existing coding. TORODE argues that this programme makes a

8 See <http://www.qualisresearch.com/>.

9 See <http://www.code-a-text.co.uk/>.



valuable contribution to CA, by allowing the researcher to extend the range of questions that can be asked about narratives in conversation; but more particularly because it has a function which allows direct links from the transcript segment to the sound or video recording. [30]

Despite TEN HAVE's optimism for computer assisted analysis, and TORODE's enthusiasm for Code-a-Text in particular, the general tone is one of caution when using CAQDAS in conversation analysis. According to Clive SEALE, although Code-a-Text, and Atlas.ti are able to store and replay sound recordings, and to provide a link between the text and the recording, the "more popular CAQDAS packages are unable to support many of the things conversation analysts wish to do" (2000, p.165). Such tasks would include working with the main transcription system used in CA, as developed by Gail JEFFERSON (ATKINSON & HERITAGE, 1984). This form of notation uses symbols throughout the transcription to denote types of speech utterances (e.g. pauses, rising intonation, emphasis, word extension, loud, soft, and inaudible talk), and cannot be transferred into any of the main qualitative data analysis programmes. [31]

In summary, the studies we examined which used qualitative software with discourse analysis were all either doing entirely quantitative linguistic analysis, or discourse analysis in conjunction with statistical analysis, or were using the programmes practically, to hold and display text, images and sound files. With the exception of CA, none of the studies reviewed were doing a fine grained analysis of texts, in accordance with the theory and methods described above. Our challenge was to conduct the kind of study we had been searching the literature for—to use CAQDAS throughout a DA of texts, from the construction of research questions, to data collection, to reading and note taking, to coding, and to analysis. At what point would the CAQDAS assist or inhibit the research? We had identified the practical value of using CAQDAS for managing data with some approaches, would we be able to do more than this, and to fulfil the promise of one of the developers of the QSR products, NUD.IST and NVivo, that "[a]ll researchers working in the qualitative mode will be clearly helped by some computer software" (RICHARDS, 1995, p.105)? [32]

## **5. A Practical Application: Smoking Guns**

Our approach to discourse analysis is one that has developed out of various perspectives including ethnomethodology and conversation analysis, sociology of science, and rhetoric, and their applications to topics in social and general psychology (ANTAKI, 1994; BILLIG, 1987; 1992; EDWARDS, 1997; EDWARDS & POTTER, 1992; POTTER, 1996; POTTER & WETHERELL, 1987). It is an inductive approach which, in general, avoids the use of coding categories or interpretative schemas, in favour of tying analytic claims closely to the details of the texts. While we knew that CAQDAS can search, organise, and code data, we were also aware of claims that the software can do "more than just coding" (COFFEY et al., 1996). It was this "more" that we next set out to test. [33]

For the following enquiry three CAQDAS programmes (MaxQDA, NVivo, and Qualrus) were selected, simply because they were ranked high amongst some of the main programmes currently used in qualitative research, and were already being evaluated by our project for other tasks<sup>10</sup>. The aim in this study was not to compare the programmes, however, but to ensure that if problems were encountered, we could then explore whether this might be an issue for qualitative software in general, or with specific programmes, by setting each programme the same task. [34]

Our practical evaluation of CAQDAS began in the early stages of conducting a study on the historic and discursive usage of the phrase "smoking gun" in the UK and US news media (BILLIG & Macmillan, 2005). Of particular interest was how the metaphor originated in its modern sense, and how it progressed from metaphor to idiom. This entailed detailed qualitative analysis, examining the use of the term during the Watergate Crisis, the so-called "Iran-gate Crisis" and more recently in relation to the search for Saddam Hussein's alleged weapons of mass destruction. For the purposes of this study we had collected all news items on "smoking guns" over a twenty year period (1983 – 2003) from the LexisNexis<sup>11</sup> news archive. This had produced a data set of approximately 2000 news stories. [35]

Although the success of the discourse study "is *not* in the least dependent on sample size" (POTTER & WETHERELL, 1987, p.161, original emphasis), and analysis can be successfully carried out on small samples, discourse analysis is still extremely "labour-intensive" (ibid., p.161). There was no expectation, of course, that a software programme could analyse the "smoking gun" stories for us, but our large data set did offer the opportunity to assess the extent to which CAQDAS could help us to "manage" our data, or produce a framework which might assist our analysis, and to consider whether the operational functions were significantly more useful than those in "Windows". [36]

### 5.1 Importing news texts

After collecting the data, our first test was to measure how quickly we could import the news items into NVivo, Qualrus, and MaxQDA—our chosen qualitative software programmes. Learning to operate any CAQDAS is inevitably time consuming (WELSH, 2002), and while Qualrus and MaxQDA are fairly easy to operate "intuitively", NVivo is clumsier, more difficult, and therefore more time consuming for the novice user. [37]

With any qualitative software programme data has to be "prepared" in a particular way. In DA investigation begins the moment the analyst starts to read the text. With CAQDAS, however, data must be tidied into relevant and searchable

---

10 For a detailed evaluation of these programmes, and other main CAQDAS, see our website at <http://www.lboro.ac.uk/research/mmethods/index.html>.

11 This professional database holds data on UK and EU case law & legislation, company information, market research data and news coverage. The process of retrieving news, however, is at times problematic, producing unreliable results such as information repetition and information loss (DEACON, 2004; SOOTHILL & GROVER, 1997).

segments first. This means that texts have to be made ready by changing the format of the file. With NVivo and Qualrus the required format for texts is "text" (.txt) or "rich text format" (.rtf), and with MaxQDA this is .rtf only. .Txt can recognise .html in principal, but is inadequate for the task, while .rtf cannot recognise html documents, documents with tables, or texts with "headers" and "footers" at all. Unfortunately, much of our data from LexisNexis media texts contained the name of the newspaper in a header. This meant that each of the many "headers" within a document of texts had to be located and deleted before the text was ready to be imported. [38]

The quickest way to save each story in LexisNexis is as a search consisting of 200 stories (the size limit for a LexisNexis document) or less. The document of up to 200 stories then has to be imported into the chosen software package and reconverted into single stories for analysis. Text splitting turned out to be problematic. There are many programmes designed to split texts, but most work on the principal that the user is able to define the size of all "chunks" (e.g. a specified number of lines) before activating the function. This is understandable, but clearly of little use when splitting media texts, since all stories will be of varying sizes.<sup>12</sup> [39]

After importing the data and splitting the texts into single documents, we were then ready to examine the news stories for an overview of how often the phrase "smoking gun" was used, and in which kinds of context, and to begin formulating our research questions. This involved reading through all 2000 stories, noting what the story was about, and which "smoking gun" terms were used. Although all stories were now accessible in CAQDAS, this made the task of reading the texts no easier. In fact it was still more pleasant to print out the news items, and to read through them with a highlighter pen to hand! [40]

## 5.2 BOOLEAN searches

We had by now identified that there was a number of different expressions incorporated in the phrase "smoking gun", including "*a smoking gun*", "*the elusive smoking gun*", "*the so called smoking gun*", "*smoking guns*", "*no single smoking gun*", and "*no smoking guns*". While, of course, the "find" function of a Windows programme can quickly and efficiently find such phrases within a text (and therefore in context), and can highlight them in bold and/or italics (if required), we decided that this was a good opportunity to see whether the Boolean search function of the various CAQDAS programmes could improve on this. [41]

When carrying out a Boolean search the text is examined for relevant keywords with the three most commonly used search terms "and", "or", and "not". Our initial, inexperienced Boolean searches were extremely time consuming, with a first search (in NVivo) of all the documents taking 30 minutes, and the computer crashing at the end of the search. Since we knew that stories containing the terms we were looking for existed, we were surprised that our first attempts at

---

<sup>12</sup> For advice on splitting files see  
<http://www.lboro.ac.uk/research/mmethods/resources/preparation/index.html>.

Boolean searches produced nothing. Questions posed to one of the developers of NVivo produced two possible solutions to our problem—either to perform a "text search" instead of a Boolean search, or to code the document before searching it. [42]

Coding in the early stages of analysis is a way of creating a broad overview of the data, and can have a place in most analyses, as long as the "codes" are directly related to the research questions (POTTER & WETHERELL, 1987). However, the danger of coding according to the capabilities of the software is that the researcher is steered towards treating the data in terms of categories, and as such something that can be given significance through counting, dividing, and sub dividing. Coding becomes the method of analysis, rather than a way of managing the data. For our evaluation categorising news stories was a way of indicating a relevant story, and briefly summarising this relevance. This was as useful in CAQDAS, as inserting a "Comment" in Microsoft Word, or writing in the paper margin might be. [43]

### 5.3 Coding segments

There are practical and methodological problems with coding a document in order to later retrieve coded segments. The software programme can only do what you ask it to do, which means that it can only retrieve segments exactly as they have been coded. Practically this is a long, drawn out process for such a small part of the research. In order to assign codes comprehensively throughout our data we ideally needed to work through all 2000 extracts. This means that, methodologically, coding is given priority before the researcher is familiar with the data, or before coding has been identified as necessary for analysis. [44]

A number of the main programmes have been designed with Grounded Theory in mind<sup>13</sup> and as such work systematically to build "theory" out of the data. In practical terms this works through the creation of hierarchical categories and an orderly application of codes to text. After a first coding, the software is organised in such a way as to encourage further coding of the data into more precise categories, and beyond, until the researcher has what she perceives to be the issues under scrutiny in a nutshell. [45]

While counting instances is not particularly useful in general for discourse analysis, it did, on this occasion, provide us with background historical information of the use of the term "smoking gun" over time. Our first codes were designed to categorise, broadly, *when* the term "smoking gun" was used (which period of time), and *what* political context it occurred in. The *how* of its use was the later task of a more detailed analysis. In this way we were able to locate the regular use of the idiom within controversial political situations. Using CAQDAS for this very basic content analysis was certainly do-able, and, once the data had been imported, was relatively trouble-free. It was not, however, strictly necessary. In fact, in a comparison of the time taken to code and count each story using the

---

13 See [http://www.lboro.ac.uk/research/mmethods/research/software/caqdas\\_primer.html](http://www.lboro.ac.uk/research/mmethods/research/software/caqdas_primer.html).

software, and the time taken to do this manually, simply by reading the story, and making notes, CAQDAS came off worst. Bearing in mind the time taken to load the data into the software, and to learn to operate the programme, this task turned out to be more efficient, more reflective and much more enjoyable done by hand. [46]

Since the intention was to evaluate CAQDAS, as well as to create a broad overview of "smoking gun" use in the press, we next developed a second set of codes to look at

- what the stories were concerned with (*e.g. co-operation failure; sufficient evidence*);
- what kind of expressions were used (*for example, "the elusive smoking gun", "the so called smoking gun", "no smoking guns"*);
- which actors were speaking (*for example, PM Tony Blair; President George W. Bush; UN spokesperson Hans Blix*);
- and the tone of the news item (*e.g. defensive; evasive*). [47]

Assigning codes to sections of the documents meant we were still no closer to doing a discourse analysis of texts, although it did have the (shoe box) benefit of holding stories within categories that we could later retrieve quickly. Part of this frustratingly time consuming process involved getting to grips with operating the software, and attempting to find solutions to a number of practical problems. [48]

Examples of practical problems included discovering that the major CAQDAS don't have an undo function; and that NVivo's "coding stripes"—a function heralded as a virtue of the software (WELSH, 2000)—didn't work adequately. We needed this function to display the codes (including newspaper names, assigned as codes) given to particular segments of particular news items, so that we could see a rough summary of the stories at a glance. However, using the function froze the computer screen within a minute of activating the coding stripes, and lost all work on coding. This happens, apparently, whenever the user edits with the Coder, and is a fault clearly known by the developers, with the QSR website advising the user, rather unhelpfully, to simply to turn off the coding stripes when coding (see <http://www.qsr.com.au/support/faq/faq.asp>, FAQ NV-137). [49]

At first the editing tools in Qualrus didn't work.<sup>14</sup> We then discovered that, contrary to the "QualrusHelp" claim that Qualrus could handle "source files of varying sizes, including very large ones", Qualrus also recommended that large documents be divided into smaller sections. We therefore decided to test this ourselves, by importing over 350 media stories into Qualrus, and, for this trial, not splitting the texts into separate stories. Our first document contained approximately 80 stories. We coded approximately 11,500 words (about 13 stories) before reaching the end of the scroll bar. The whole document *had* been imported but was no longer visible, or accessible. The immediate solution

---

<sup>14</sup> At the time of testing monitor resolution needed to be set at least at 1024 x 768 in order to operate this function in Qualrus.

seemed to be to return to the original source (once again) and divide it into smaller documents. Once this was done we recoded the new documents, using the main document as a template. However, another option, to update our five month old programme, was given by the technical contact at the Ideaworks Helpdesk. Updating was quick, easy, and worked. No codes were lost. We were then able to view the news document in full. [50]

It is important to remember that throughout this process our focus was on operating the software, and assigning and reassigning codes. Little attention was given to issues of interest from a discursive perspective, including a detailed examination of the rhetorical use of the term "smoking gun" during times of political scandal and crisis, and, most particularly, the ideological effects of the use of the idiom in the search for weapons of mass destruction in Iraq. [51]

## 6. Discussion: So Far No Smoking Guns

Our trial of CAQDAS had to end when we began the analysis proper. If coding is not the framework for analysis, then there is no place for the software in this process, except as a way of moving between the data and the research notes. Working with texts in CAQDAS had consisted of learning to operate the programmes, to overcome technical problems, and to categorise the news items in terms of codes. Although we were able to use the software to count and display our instances of "smoking gun" coded to show a connection with political scandals, this merely confirmed what we had gathered by simply reading through the stories. Our second set of codes, were, in the end, of little value. Assigning categories such as "actor" and "tone" and types of phrase created a content analysis of the texts that showed who said what. Since our main interest was not so much in *who*, *what*, or *how often*, but in *how* such descriptions were used, this had brought us no closer to a discourse analysis of the news media. [52]

Clive SEALE (2000, p.155), in his description of how to use computers in the analysis of "qualitative data", suggests that one of the advantages of CAQDAS is its "speed at handling large volumes of data, freeing the researcher to explore numerous analytic questions" (see also WEITZMAN, 2000). Data management of this kind, according to SEALE, involves sorting texts into categories, or coding segments "which may then be filed and retrieved easily" (*ibid.*). This "speeding up" of data management should not, of course, be confused with a speeding up of analysis. FIELDING and LEE (1998), for example, argue that there is little evidence to show that using CAQDAS shortens the time spent on analysis (see also MANGABEIRA et al., 2004). [53]

In our own case the time spent on the problems we encountered with CAQDAS, and fruitless attempts at inconsequential coding that bore no relation to the finished analysis, considerably *increased* the time we would have spent on the "smoking gun" study had we restricted our research to using the traditional manual methods of examining the data. It had, of course, been useful to locate when the term was used, the political context in which it was used, and how often it was used. It had been useful to highlight sections of text with general coding

descriptions, and also to write memos for certain sections of the text. However, all of these tasks could have been done as quickly and as easily using manual methods, with the help, perhaps of the Word function to insert memo "Comments" in the text. [54]

### 6.1 Methodological compatibility?

CAQDAS are useful for practical tasks in general, such as searching for and retrieving data segments. They are also useful for coding segments of text, for maintaining links between codes, and for providing a framework for materials, from which the researcher can make summary judgements. They can, to a very limited extent, be useful in the early stages of a discourse analysis, to hold texts, to search them, and to assist in rudimentary coding, if such is required. They cannot, however, bring about the kind of organization of materials required for an in-depth, in-context analysis of the level required for a detailed analysis. [55]

By attempting to impose the structure of CAQDAS on our discursive examination of news, we were restricting the scope of our study. This could be avoided either by not using the programmes in discourse analysis at all; or by using them to hold the data, but not to assist analysis. Whether this is worth the time consuming effort of learning to operate the software, and preparing and loading texts, is a matter for each individual project, and researcher. [56]

For DA the material to be analysed has to be understood in relation to its particular discursive, interactional or rhetorical context. This means that its particularities must be studied—it is not enough to consider these as instances of something more general. For this reason, discourse analysis cannot be defined as a universal set of procedures (ANTAKI et al., 2003) to be formalised into a computer package. Instead, discourse analysis always poses new problems which, in their turn, make new demands upon the analyst. In DA the researcher should be in charge of the analysis from the moment the first document is read. Using CAQDAS with DA can, at best, be more time consuming than useful, and at worst, can steer the analyst away from the task of analysis. [57]

### References

- Alexa, Melina & Zuell, Cornelia (1999). A Review of Software for Text Analysis. *ZUMA-Nachrichten*. Spezial Band 5. ZUMA, Mannheim.
- Antaki, Charles (1994) *Explaining and Arguing: The Social Organization of Accounts*. London: Sage.
- Antaki, Charles, Billig, Michael, Edwards, Derek & Potter, Jonathan (2003). Discourse Analysis Means Doing Analysis: A Critique of Six Analytic Shortcomings. *Discourse Analysis Online*, 1. Available at: <http://www.shu.ac.uk/daol/articles/v1/n1/a1/antaki2002002-paper.html> [Date of Access: February 14th 2005].
- Atkinson, J. Maxwell & Heritage, John (1984). Introduction. In J. Maxwell Atkinson & John Heritage (Eds.), *Structures of Social Action: Studies in Conversation Analysis* (pp.1-15). Cambridge: Cambridge University Press.
- Billig, Michael (1992). *Talking of the Royal Family*. London: Routledge.
- Billig, Michael (1987). *Arguing and Thinking: A Rhetorical Approach to Social Psychology*. Cambridge: Cambridge University Press.

- Billig, Michael & MacMillan, Katie (2005). Metaphor, Idiom and Ideology: The Search for "No Smoking Guns" Across Time. *Discourse & Society*, 16(4), 459-480.
- Carvajal, Diogenes (2002). The Artisan's Tools. Critical Issues When Teaching and Learning CAQDAS [46 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 3(2), Art. 14. Available at: <http://www.qualitative-research.net/fqs-texte/2-02/2-02carvajal-e.htm> [Date of Access: February 16th 2005].
- Coffey, Amanda, Holbrook, Beverley & Atkinson, Paul (1996). Qualitative Data Analysis: Technologies and Representations. *Sociological Research Online* 1(1). Available at: <http://www.socresonline.org.uk/socresonline/1/1/4.html> [Date of Access: February 16th 2005].
- Cuilenberg, Jan J., Kleinnijenhuis, Jan & de Ridder, Jan A. (1988). Artificial Intelligence and Content Analysis: Problems of and Strategies for Computer Text Analysis. *Quality and Quantity*, 22, 65-97.
- Deacon, David (2004). Yesterday's Papers and Today's Technology: Digitalised News Archives and Media Analysis. Paper presented to the *Political Communication Research Section, International Association for Media and Communication Research Conference*, July 25-July 30, Porto Alegre, Brazil.
- De Beaugrande, Robert (1996). The Story of Discourse Analysis. In Teun van Dijk (Ed.), *Introduction to Discourse Analysis* (pp.35-62). London: Sage.
- Edwards, Derek (1997). *Discourse and Cognition*. London: Sage.
- Edwards, Derek & Potter, Jonathan (1992). *Discursive Psychology*. London: Sage.
- Evans, William (2001). Computer Environments for Content Analysis: Reconceptualizing the Roles of Humans and Computers. In Orville Vernon Burton (Ed.), *Computing in the Social Sciences and Humanities* (pp.67-87). Urbana: University of Illinois Press.
- Fairclough, Norman (1995). *Media Discourse*. London: Edward Arnold.
- [Fielding, Nigel G.](#) & Lee, Raymond M. (1998). *Computer Analysis of Qualitative Research*. London: Sage.
- Fowler, Roger (1991). *Language in the News: Discourse and Ideology in the Press*. London: Routledge.
- Fowler, Roger & Kress, Gunther. (1979). Critical Linguistics. In Roger Fowler, Bob Hodge, Gunther Kress & Tony Trew (Eds.), *Language and Control* (pp.185-213). London: Routledge and Kegan Paul.
- Garfinkel, Harold (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- [Gibbs, Graham R.](#), Friese, Susanne & Mangabeira, Wilma C. (2002). The Use of New Technology in Qualitative Research. Introduction to Issue 3(2) of FQS [35 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 3(2), Art. 8. Available at: <http://www.qualitative-research.net/fqs-texte/2-02/2-02hrsg-e.htm> [Date of Access: January 27th 2005].
- [Glaser, Barney G.](#) & Strauss, Anselm L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.
- Gumperz, John J. & Hymes, Dell (Eds.) (1972). *Directions in Sociolinguistics: The Ethnography of Communication*. New York: Holt, Rinehart and Winston.
- [Have, Paul ten](#) (1991). User Routines for Computer Assisted Conversation Analysis. *The Discourse Analysis Research Group Newsletter*, 7/3 (Fall: 3-9).
- Have, Paul ten (1998). *Doing Conversation Analysis: A Practical Guide*. London: Sage.
- Hardt-Mautner, Gerlinde (1995). *Only Connect. Critical Discourse Analysis and Corpus Linguistics*. New Technical Papers from UCREL, V 6. Available at: <http://helmer.aksis.uib.no/corpora/1995-3/0114.html> [Date of Access: 20th March 2005].
- Koenig, Thomas (forthcoming). From Frames to Keywords to Frames: Quantifying Frame Analysis. In Sharlene Hesse Biber & Raymond Maitetta (Eds.), *Where Method Meets Technology*. London: Sage.
- Kress, Gunther (2001). From Saussure to Critical Sociolinguistics: The Turn Towards a Social View of Language. In Margaret Wetherell, Stephanie Taylor, & Simeon J. Yates (Eds.), *Discourse Theory and Practice: A Reader* (pp.29-38). London: Sage.



- Labov, William (1972). *Language in the Inner City: Studies in the Black English Vernacular*. Oxford: Blackwell.
- MacMillan, Katie & Koenig, Thomas (2004). The Wow Factor: Preconceptions and Expectations for Data Analysis Software in Qualitative Research. *Social Science Computer Review*, 22(2), 179-186.
- Mangabeira, Wilma C., Lee, Raymond M., & Fielding, Nigel G. (2004). Computers and Qualitative Research: Adoption, Use and Representation. *Social Science Computer Review*, 22(2), 167-178.
- McEnery, Tony & Wilson, Andrew (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Potter, Jonathan (2000). Post Cognitivist Psychology. *Theory and Psychology*, 10, 31-7.
- Potter, Jonathan (1996). *Representing Reality: Discourse, Rhetoric, and Social Construction*. London: Sage.
- Potter, Jonathan & Wetherell, Margaret (1987). *Discourse and Social Psychology: Beyond Attitudes and Behaviour*. London: Sage.
- Richards, Lyn (1995). Transition Work! Reflections on a Three-Year NUD.IST Project. In Robert G. Burgess (Ed.), *Studies in Qualitative Methodology*, V 5 (pp.105-140). Greenwich and London: JAI Press.
- Richards, Thomas J. & Richards, Lyn (1994). Using Computers in Qualitative Research. In [Norman K. Denzin](#), & Yvonna S. Lincoln (Eds.), *Handbook of Qualitative Research* (pp.445-462). London: Sage.
- Seale, Clive (2000). Using Computers to Analyse Qualitative Data. In David Silverman (Ed), *Doing Qualitative Research: A Practical Handbook* (pp.154-174). London: Sage.
- Soothill, Keith & Grover, Chris (1997). Research Note: A Note on Computer Searches of Newspapers. *Sociology*, 31, 591-596.
- Torode, Brian (1998). Narrative Analysis Using Code-a-Text. *Qualitative Health Research*, 8(3), 414-432. See also [http://www.tcd.ie/People/Brian.Torode/WEB\\_ONE/BT99\\_Narrative\\_W32.pdf](http://www.tcd.ie/People/Brian.Torode/WEB_ONE/BT99_Narrative_W32.pdf) [Date of Access: January 24th 2005].
- Van Dijk, Teun A. (Ed.) (1985). *Handbook of Discourse Analysis* (4 vols.). London: Academic Press.
- Van Dijk, Teun A. (1995). Discourse Semantics and Ideology. *Discourse and Society* 5(2), 243-289.
- Van Dijk, Teun A. (2001). Principles of Critical Discourse Analysis. In Margaret Wetherell, Stephanie Taylor, & Simeon J. Yates (Eds.), *Discourse Theory and Practice: A Reader* (pp.300-317). London: Sage.
- Weitzman, Eben A. (2000). Software and Qualitative Research. In Norman K. Denzin, & Yvonna S. Lincoln (Eds.), *Handbook of Qualitative Research* (pp.803-820). London: Sage Publications.
- Weitzman, Eben A., & Miles, Matthew B. (1995). *Computer Programs for Qualitative Data Analysis: A Software Sourcebook*. Thousand Oaks, CA: Sage.
- Welsh, Elaine (2002). Dealing with Data: Using NVivo in the Qualitative Data Analysis Process [12 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 3(2), Art. 26. Available at: <http://www.qualitative-research.net/fqs-texte/2-02/2-02welsh-e.htm> [Date of Access: August 10th 2005].
- Wetherell, Margaret, Taylor, Stephanie, & Yates Simeon J. (Eds.) (2001a). *Discourse Theory and Practice: A Reader*. London: Sage.
- Wetherell, Margaret, Taylor, Stephanie, & Yates Simeon J. (Eds.) (2001b). *Discourse as Data: A Guide for Analysis*. London: Sage.
- Wilson, Andrew (1993). Towards an Integration of Content Analysis and Discourse Analysis: The Automatic Linkage of Key Relations in Text. *Unit for Computer Research on the English Language Technical Papers* 3. 11 pages. Lancaster University (unpublished). Available at: <http://www.comp.lancs.ac.uk/computing/research/ucrel/papers/techpaper/vol3.pdf> [Date of Access: August 10th 2005].
- Yates, Simeon J. (2001). Researching Internet Interaction: Sociolinguistics and Corpus Analysis. In Margaret Wetherell, Stephanie Taylor, & Simeon J. Yates (Eds.), *Discourse as Data: A Guide for Analysis* (pp.93-146). London: Sage.

## Author

*Katie MACMILLAN* is a Research Associate in the Department of Social Sciences at Loughborough University, currently examining methods of analyzing media content. Her long term research interest is in knowledge construction—particularly in terms of reflexivity and research; therapy and therapy talk; and media constructions of news events.

Contact:

Katie MacMillan

Social Sciences  
Loughborough University  
Loughborough  
LE11 3TU, UK

Telephone: +44(0)1509 223730

Fax: +44 (0)1509 223944

E-mail: [k.macmillan@lboro.ac.uk](mailto:k.macmillan@lboro.ac.uk)

URL:

<http://www.lboro.ac.uk/research/mmethods/index.html>

## Citation

MacMillan, Katie (2005). More Than Just Coding? Evaluating CAQDAS in a Discourse Analysis of News Texts [57 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(3), Art. 25, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0503257>.