6.16 Grounded theory, coding and computer-assisted analysis

Graham R Gibbs, *Reader in Social Research Methods, University of Huddersfield*

For much qualitative research analysis of social reality requires that we understand the 'symbolic world' of those we study, that is, the meanings and interpretations that people apply to their experiences. The researcher may start with some understanding of the phenomena, even before finishing the collection of data, but this understanding expands during the course of the project by continually checking data against interpretations until we are satisfied we have a full grasp of the phenomena.

The procedures most researchers use to manage and prepare data for such analysis are quite straightforward although time consuming. These involve:

i. Compiling the corpus of data (fieldnotes, transcripts, images, audio and video recordings). Despite the growing popularity of images, audio and video data in research, most researchers still transcribe their interviews and observations as they find this form of textual data much easier to use in the subsequent stages of analysis.
ii. Detailed reading of the transcripts (or listening to the recordings or looking at the video and images) both to gain an impression of their content as a whole and to begin to generate ideas, hunches, categories and themes that interpret the phenomena. At this point many researchers find it helpful to write a précis of each case.
iii. Explicitly searching for categories and patterns in the data; marking the data with category (or 'code') labels. In this process of coding the researcher marks passages of text (or parts of images) as being examples of things represented by the code names.
iv. Constructing thematic outlines using the codes to lay out the sequence in which topics will be considered.

In the past these procedures involved the physical manipulation of data (literally cutting up transcripts and sorting the cuttings into sets of associated extracts kept in folders or files) but nowadays the process can be conducted using software (although some still prefer 'manual' methods, especially for small-scale studies). Whatever approach is used, good data management is key. Qualitative data are voluminous; Miles and Huberman's (1994) evaluation of innovation in a school system produced over 3,000 pages of fieldnotes. It is vital that the analysis is not partial (leaving out significant data or giving more weight to some data) and is comprehensive (does not omit key phenomena) but above all is worked up from the data rather than reflecting a pre-existing idea supported by highly selective examples. The danger of what Silverman (2001, pp 222-3) refers to as anecdotalism is rife as there is always a temptation to give undue weight to particularly colourful, surprising or intriguing examples in the data even if they are not actually common, typical or theoretically significant.

Grounded theory

The most commonly used analytic approach which also most closely reflects the emphasis on working 'up from the data' is Glaser and Strauss's (1967) *grounded theory*. In the 1960s many researchers undertook qualitative research by using their reading of the literature and theory and

their prior experience in the field to construct an *a priori* list of list of categories which they would then use to mark-up the data. Glaser and Strauss rejected this. For them, not only should the researcher maintain a open mind in order to find new phenomena in the data (the process of *open coding*), but far from 'testing' existing ideas (or hypotheses) the analyst should take an inductive approach i.e. be concerned with constructing analytic codes and categories from the data and discovering new theory. For Glaser and Strauss it was important to maintain *theoretical sensitivity* by immersing oneself in the data and staying open to new ideas, interpretations and ultimately theories which would be grounded in the data. This sensitivity enables the researcher to generate or discover ways of categorizing the data.

The principle way of ensuring that the coding is thorough and consistent is the *constant comparative method*. The researcher begins by examining 'incidents' recorded in fieldwork or mentioned in interviews, incidents being discrete acts or the expression of an attitude by respondents (Becker and Geer, 1960). Each is coded into as many theoretical categories as possible (open coding). Categories may be derived from terms used by participants (characterising aspects of their work for example), from interpretations developed by the analyst and sometimes from existing analytic constructs (found, for example, in the literature). Before assigning a code to a passage of text or other data, the method requires that we re-examine all incidents or passages previously coded using the same category. This recursive approach ensures that theoretical properties of categories are fleshed out as analysis proceeds. It helps identify different categories and the relationship between categories. Thus, conceptualisation emerges during coding and this is aided and recorded by the writing of '*analytic memos*' on the meaning and significance of the code or category. Such memos specify the properties of the codes, define relationships between categories and identify gaps, such as potential related codes.

There is a tendency for initial coding approaches to produce very descriptive codes, coded passages that are decontextualised (devoid of meaning because they are removed from their context) and, above all, too many codes. Therefore, most grounded theorists and researchers adopting similar coding approaches describe different levels or forms of coding at different stages of the analysis and recommend that the analyst moves from descriptive codes to more conceptual or theoretical codes whilst at the same time reducing the number of key codes being used. Glaser (1978) and Charmaz (2006) refer to two stages, initial, open or substantive coding and later, theoretical coding. Strauss and Corbin (1998), who describe three stages, open coding, axial coding and selective coding, have developed a range of hints, suggestions and heuristics about how the codes may be refined from stage to stage. Glaser (1992) has criticised this for being too mechanistic and forcing theories onto the data rather than allowing them to emerge, but most researchers find steps and heuristics very helpful for guiding their own practice. One procedure that most grounded theorists agree on is that of dimensionalising codes - recognising that they have properties or dimensions. Open coding often develops a number of related codes that can be gathered together as types, forms, occurrences, consequences, causes or settings for other, more abstract, analytic and theoretical codes. Such grouping reduces the number of codes and creates more theoretical categories.

In the later stages of analysis the researcher concentrates on clarification, simplification and 'reduction' of the theory to limit the number of categories germane to the emerging theory. Ongoing analysis can now focus just on those categories and they become '*theoretically*

saturated', that is the dimensions and properties of the category include all the variations found in the data. It becomes immediately apparent whether a subsequent incident requires modification of the category. Where it does, the category is coded and compared to existing categories, if not, the incident need not be coded.

Over the years there has developed a range of distinct accounts of grounded theory, typified by the disagreements between the originators Glaser and Strauss. Such differences may account for the fact that, while the majority of qualitative researchers claim adherence to grounded theory, inquiry into their working practices reveals substantial variation and deviation from the steps laid down in the canonical texts (Fielding and Lee, 1998). As long as the research is based on thorough and exhaustive analysis, this probably does not matter much, but as Seale (2001, pp 658-9) notes, it all too often leads to a superficial 'pattern analysis' that simply reports major themes without any attempt to develop theory or investigate more subtle relationships and processes. A key recent development has been the proposal by Charmaz (2006) of a constructivist grounded theory. She has argued, pace Denzin and Lincoln (1994), that the essential principles of grounded theory are compatible with the view that theory rather than simply being discovered, is actually constructed by the analyst during analysis by engaging, interacting and interpreting the respondent.

Other analytic approaches

The popularity of the coding method can be gauged by the fact that several other approaches to qualitative analysis have embraced many of the key ideas of grounded theory. Thus the *template analysis* approach (King 2004, Crabtee and Miller, 1992) and *interpretative phenomenological analysis* (IPA) (Smith, Flowers and Larkin, 2009) borrow many procedures from grounded theory. In addition, both take a more explicit case by case approach that owes much to another, earlier approach, *analytic induction*. According to this method, developed by Znaniecki (1934), following the initial identification of a phenomenon and the inductive generation of a hypothesis about it that should apply to all the cases, the researcher engages in repeated rounds of checking each case against the hypothesis. When a case does not fit, either the hypothesis is tweaked to make it fit, or if that is not possible the researcher revises the definition of the phenomenon to which the hypothesis applied. Analytic induction has not been much used in recent decades, partly because of the recognition of its logical flaws (Hicks, 1994) but also because it was overtaken by the popularity of coding approaches. Nevertheless, although template analysis and IPA use coding they have retained a case-by-case, iterative development of initial coding as a key mechanism for the evolution of analysis.

There has been a revival of interest in pure case-by-case approaches prompted by Ragin's (1987) 'qualitative comparative analysis'. The method compares outcomes across multiple cases. It uses mathematics to systematically identify 'universal conditions' which are always present in particular combinations when the phenomenon occurs. Much of this is now done with the assistance of dedicated software programs and recent developments have allowed for the inclusion of fuzzy sets which encompass outcomes that are present only to a certain degree.

Computer-assisted qualitative data analysis

Many qualitative researchers now use dedicated, computer-assisted qualitative data analysis (CAQDAS) software to assist their analysis. Most such software supports the coding approach to

analysis by including facilities to code and revise the coding of data, organise the coding scheme and retrieve coded data. Data which may be analysed includes text (including tables and pdf documents), images, audio or video recordings.

The programs support the annotation of data and the writing and linking of analytic memos. Many now include a range of text retrieval or textual analysis functions which range from sophisticated text searching (multiple terms, searching for similar words or words with the same root, and searching in defined passages) through to concordance, word frequencies, keyword in context and linked cluster diagrams. Most also support the integration of variable-based data (also called attributes or families), such as quantitative and categorical data about cases, respondents or settings which can be linked to the qualitative data. Several programs now support the integration of GIS (geographical information systems) data.

A key function in most programs is a sophisticated retrieval facility for coded passages, images etc. At its simplest this means the researcher can quickly retrieve all the data coded in a particular way – the equivalent of the paper-based researcher looking through all the data collected in one folder or file. But the programs go much further than this in allowing searches for both Boolean and non-Boolean combinations of codes and codes and variables. Boolean searches use terms such as 'and', 'or' and 'not' so one can undertake a search for and retrieve data that is coded at one code 'or' at another code, or data at one code 'and' which is in a case that has a certain attribute or variable value. Non-boolean searches allow for the combining of codes using terms like 'near' or 'overlapping'. Such retrievals are very quick and easy to do and allow the analyst to undertake a rapid investigation of a wide range of relationships between codes and between codes and variables to investigate things like gender and age variations or the co-incidence in the data of several ideas and concepts or talk about those ideas. Using software does not guarantee good quality analysis, but at least it provides the tools so that researchers can undertake well-organised, thorough, well-documented and grounded analyses.

Question for discussion

- In what ways can a coding scheme be developed and how can the researcher tell when codes are theoretically saturated?

References

Becker, H. and Geer, B. (1960) 'Participant observation: the analysis of qualitative field data', in R.N. Adams and J. Preiss (eds) *Human organisation research*, Homewood, IL: Dorsey Press.

Charmaz, K. (2006) *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. London, Thousand Oaks, New Delhi: Sage.

Crabtree, B.F. and Miller, W.L. (eds) (1992) *Doing qualitative research*. Newbury Park, London: Sage.

Denzin, N.K. and Lincoln, Y.S. (eds) (1994) *Handbook of Qualitative Research*. London: Sage.

Fielding, N. and Lee, R.M. (1998) *Computer analysis and qualitative research*, London: Sage Publications.

Glaser, B. (1978) *Theoretical sensitivity*, Mill Valley, CA: Sociology Press.

Glaser, B. (1992) *Emergence vs. forcing: Basics of grounded theory analysis*, Mill Valley, CA: Sociology Press.

Glaser, B. and Strauss, A. (1967) *The discovery of grounded theory*, Chicago, IL: Aldine.

Hicks, A. (1994) 'Qualitative comparative analysis and analytic induction: the case for the emergence of the social security state', *Sociological Methods and Research*, vol 23, pp 86-113.

King, N. (2004) 'Using Templates in the Thematic Analysis of Text', in C. Cassell and G. Symon (eds), *Essential Guide to Qualitative Methods in Organizational Research*. London: Sage. pp. 256-270.

Miles, M. and Huberman, A. (1994) *Qualitative data analysis: An expanded sourcebook*, Beverly Hills, CA: Sage Publications.
Ragin, C. (1987) *The comparative method: Moving beyond qualitative and quantitative strategies*, Berkeley, CA: University of California Press.
Seale, C.F. (2001) 'Computer-Assisted Analysis of Qualitative Interview Data', in J. F. Gubrium and J. A. Holstein (eds), *Handbook of Interview Research: Context and Method*. Thousand Oaks, Ca: Sage. pp. 651-670.
Silverman, D. (2001) *Interpreting Qualitative Data. Methods for Analysing Talk, Text and Interaction*. London: Sage.
Smith, J.A., Flowers, P. and Larkin, M. (2009) *Interpretative Phenomenological Analysis: Theory, Method and Research*. London: SAGE.
Strauss, A.L. and Corbin, J.M. (1998) *Basics of Qualitative Research: Techniques and Procedures for developing Grounded Theory*. Thousand Oaks: Sage.
Znaniecki, F. (1934) *The method of sociology*, New York, NY: Farrar and Rinehart.

Further reading

On qualitative software and its relationship to qualitative analysis:

Lewins, A. and Silver, C. (2007) *Using Software in Qualitative Research: A Step-by-Step Guide*. London: Sage.
di Gregorio, S. and Davidson, J. (2008) *Qualitative Research Design for Software Users*. Maidenhead: Open University Press: McGraw-Hill.

On the procedures of qualitative data analysis:

Gibbs, G.R. (2007) *Analyzing Qualitative Data*. London: Sage.
Boeije, H. (2010) *Analysis in Qualitative Research*. Los Angeles: Sage.
Bernard, H.R. and Ryan, G.W. (2010) *Analyzing Qualitative Data: Systematic Approaches*. Los Angeles: Sage.
Bryant, A. and Charmaz, K. (eds) (2010) *The SAGE Handbook of Grounded Theory*. Los Angeles: Sage.

Website resources

For those interested in qualitative software: http://caqdas.soc.surrey.ac.uk
For those interested in qualitative analysis: http://onlineqda.hud.ac.uk/