Breadth and depth method: Step 1

Overviewing the archived qualitative data and constructing a corpus

- We've likened the first step of our breadth-and-depth method to an archaeologist's aerial survey. If you think of a great view you might have from a plane where you can see lots of interesting features that you might like to investigate and explore as places to dig into. At this stage we're not interested in the detail of the data but in flying systematically across the data landscape to get an overview.
- The aim of this step is to look across an archive or several archives, to find and review potential sources of academic data - of an appropriate nature, quality and focus to suit your research topic – and then to construct one large corpus from these data sets. Creating a corpus involves bringing together multiple qualitative data sets. You may wish to synthesise data sets in their entirety, or you may decide to select parts of a number of data sets, for instance by research participant characteristics or by method of data collection.
- During this step, decisions are made about what to include in/exclude from what will become your newly assembled corpus. When reviewing potential sources for inclusion, the researchers' own questions set the criteria, the topic of study, and geographical or linguistic context to be sought in the meta data that the archive offers concerning the projects therein.
- Working with large volumes of qualitative material requires careful organisation; a process for which there are three key stages:

1. Data audit

This involves exploring comprehensively the volume, nature and type of data available in your selected data sets and the format it is stored in. At this point the data remains in the original data sets. The information you record will depend on your approach to sampling, analysis and your study design. It could include: the number and type of files available; details of any embargoed materials; the number and type of cases; the availability of contextual materials; the geographical spread of the sample and the keywords assigned to each case. This part of the process will give you a sense of the state of the data, nature of the original projects, an understanding of the structure of cases within the archive and an insight into what your data assemblage might look like. It can also reveal anomalies in the data landscape such as gaps in contextual material or unusual files.

2. Data management

Computer software to aid qualitative analysis is essential in the management of large volumes of qualitative data. While you might organise data in Excel or a standard data base, it is convenient if the software you plan to use in the next step is adopted to store and manage data (see next step). We have trialled a variety: the commonly-available analysis package NVivo (Server), a freeware text analysis package (Antconc) which only work with plain text, bespoke programing in the language R, and the specialist text analysis and conceptual mapping software Leximancer (v4.50). The chosen package must permit the management and retrieval of data with ease. There are two essential aspects to data management: (i) the harmonisation of file names to aid retrieval; and (ii) the reorganisation of files from their original data sets into new groupings dependent on the substantive focus and chosen unit of analysis for cases. It is at this point that the individual datasets are merged and re-organised into one corpus. It may also be necessary to reformat files to suit the program that you plan to use.

3. Data storage

Large volumes of qualitative data can require significant storage space. We used password protected computers/external hard drives. Cloud-based options are problematic in terms of data security and may breach the terms of archive agreements and/or institutional ethical approval.

SOME USEFUL RESOURCES:

- Dr Georgia Philip: Working with qualitative longitudinal data http://bigqlr.ncrm.ac.uk/2017/06/26/guest-blog-10-dr-georgia-philip-working-with-qualitative-longitudinal-data/
- Dr Anna Tarrant: Reflections from the Men, Poverty and Lifetimes of Care study http://bigqlr.ncrm.ac.uk/2016/03/09/assessing-the-feasibility-of-secondary-analysis-within-and-across-two-qualitative-longitudinal-datasets-reflections-from-the-mplc-study/
- Dr Rebecca Taylor: The challenges of computer assisted data analysis for distributed research teams working on large qualitative projects http://bigqlr.ncrm.ac.uk/2017/09/18/guest-blog-11-dr-rebecca-taylor-the-challenges-of-computer-assisted-data-analysis-for-distributed-research-teams-working-on-large-qualitative-projects/

Activity: Constructing a corpus

Visit the Timescapes Archive or the UK Data Archive. Use the contextual and metadata to select two data sets (or part data sets) you would like to bring together.

- What kinds of questions might you ask of the data sets?
- How useful is the contextual and metadata available?

If possible, download the data and review the files. Consider what kind modifications to file names and formats or reorganisation of file folders you might want to make so that your corpus is ready for the next stage.

Breadth and depth method: Step 2

Approaches to Breadth Analysis using 'Data Mining' Tools

Note the irony/tension of terminology – mining suggests digging deep - but we suggest many of the techniques are metaphorically better described as doing a near-surface-level breadth analysis rather than depth analysis.

Basic techniques

These often start from approaches that are familiar to students of corpus linguistics. https://wmtang.org/corpus-linguistics/corpus-linguistics/ and include computational procedures that reduce text to bags of words or attend to the order, sequence and co-location of words. The basic procedures include

Counting frequencies: Counting the occurrence of words in particular texts or the entire set of texts. Note that further preparatory steps are often taken to make this basic procedure more useful for social science research such as: removing 'stop words' that are judged not to carry any particular value for the exercise (indefinite articles, conjunctions, pronouns, proper names etc.); lemmatisation; ignoring or removing words that have multiple meanings (adding them to 'stop words'). You may also wish to distinguish the words of an interviewer from the words of an interviewee, and, perhaps, exclude the former from frequencies or the other techniques listed. Lists showing the frequencies of word occurrence are a standard 'text mining' starting point.

Concordance, key words in context– looking at the words that occur immediately around, literally before and after, any particular word

Co-location and proximity - various forms of algorithms to spot frequency of co-location and proximity of words as a clue to 'themes' or 'topics' or 'discourse'

Measuring 'keyness', the relative rate of use of word, spotting the exceptional frequency or infrequency in comparison to usage in a corpus that is being used as the standard reference point. (Scott, M. (1997). PC analysis of key words - and key key words. *System*, 25(2), 233-45.) Rather than using a corpus or in addition to it, sometimes relative searches are conducted comparing rates of use using a dictionary of specific terms drawing on research documenting their use in a population. Some research of this sort uses 'emotion words'.

Software products

Software that you are already familiar with could be used for some of these purposes, for example, Word combined with Excel or NVivo but this is likely to involve a lot of work and there are somewhat easier ways.

AntConc (http://www.laurenceanthony.net/software/antconc/) free and does all of the above

Wordsmith is a standard commercial product that generates wordlists and keyness. It costs £50 although you can download a free trial version4. http://www.lexically.net/wordsmith/

Wmatrix was developed by University of Lancaster has additional functionality, also costs £50 a year or is free to Lancaster students. http://ucrel.lancs.ac.uk/wmatrix/; http://ucrel.lancs.ac.uk/usas/

You can develop your own program using the *open source programing language R and Python* to use algorithms searching through the co-location and proximity of words to try to map meaningful topics

IRAMUTEQ that provides users with statistical analysis on text corpus and tables composed by individuals/words. http://www.iramuteq.org/

And then there are many commercial products that are using algorithms searching through the co-location and proximity of words to try to map meaningful topics such as Leximancer.

Leximancer https://info.leximancer.com/

Some computer aided qualitative analysis software integrate elements of web based data mining

DiscoverText https://**discovertext**.com

WebQDA https://www.webqda.net/?lang=en

Activity: Using Antconc

Download the free software:

http://www.laurenceanthony.net/software/antconc

Using the pre-prepared data set you have been given or that you yourself have prepared (the files must be in plain text format) and the list of stop words you have been given (or the list of stop words that you have identified)

- Look at word frequencies using 'word list' with and without stop words
- Try looking at concordance and moving back and forward between file view
- Try a keyness analysis and then look at keywords in context moving back and forward between different 'views' – file view, key word in context view

Note we are using Antconc for this exercise not because we wish to promote this particular software but because it is free, well documented and relatively fast and easy to learn and teach.

Breadth and depth method: Step 3

Preliminary Analysis

- Preliminary analysis is the third step in our breadth-and-depth method. This is where we begin to move from breadth to depth by sampling the features identified using text mining in step 2.
- Drawing on our archaeological metaphor we have likened this step to shallow 'test pit' sampling using the keywords and themes identified in step 2. The idea is to dig deep enough to show whether there is anything of interest or not, but not to go into the data in great depth.
- Your research questions guide identification of particular themes or keywords as more interesting than others. Use these as your starting point. We suggest that you also complete this step sampling some keywords or themes that, on the face of it, seem less attention-grabbing but of some possible relevance before moving onto step 4.
- At this stage you will be exploring, relatively quickly given the volume of data, sample extracts of material that contain keyword(s) or themes of interest from step 2. It is important to keep the context in which the data was generated in mind. Your focus will be on a cursory reading of the extracts; enough to enable you to decide what material you would wish to include in/exclude from your analysis. We suggest exploring extracts of around 200 words to gain a clear sense of whether the material is of relevance to your research questions.
- Any themes or concepts that seem ambiguous should be eliminated at this stage. It is an iterative process so you may need to go back to step 2 to conduct some more mapping.
- You could, at this stage, be faced with hundreds of extracts. You'll need to think about your sample size and this will be informed by your sampling logic, for example:
    - Theoretical
    - Purposive
    - Realist
- The sampling logic will be determined by your research design, as well as practical considerations of time and resources.
- You may not find anything of particular interest the first time and you might need to go back to step 2 to look again at other themes.
- Whilst it might be tempting to do so it is important not to delve into the data in any great depth. Rather, the emphasis of step 3 is to identify the key places across the corpus in which to dig deeper during step 4.

SOME USEFUL RESOURCES:

- Baker, S.E. and Edwards, R. (2012) *How many qualitative interviews is enough?* Discussion Paper. NCRM.
- Emmel, N. (2013) *Sampling and Choosing Cases in Qualitative Research: A Realist Approach*. Sage, London.

Breadth and depth method: Step 4

General approaches to conducting in-depth analysis

Qualitative analysis involves immersion in data that is sensitive to its context and to multi-layered complexity. In-depth analysis focuses on rich detail to represent intricate social realities and produce nuanced social explanations.

Qualitative data analysis involves breaking down and organising data in order to address your research aims and answer your research questions. It involves, variously:

- reading and re-reading your data over and again
- searching for, identifying and describing patterns of meaning and processes
- comparing and contrasting meanings and processes
- searching for relationships between meanings and processes

There is a diversity of qualitative analytic strategies and in-depth techniques that illuminate, variously, social meanings, subjectivities, activities, processes, constructions and discourses. Examples include:

- Thematic analysis, focusing on recurrent themes in the data.
- Framework analysis, organising data to identify commonalities and differences across and within cases.
- Grounded analysis, inducting meaning from the specifics of data
- Narrative analysis, focusing on how research participants construct and sequence the stories they tell.
- Conversation analysis, examining the procedures that speakers use to communicate.
- Discourse analysis, mapping ways of knowing about a situation, behaviour or group of people.

Which analytic technique, or combination of techniques, and focus are adopted is determined by the researchers' epistemological stance, conceptual approach, substantive concerns, and the pragmatics of the form/s of data.

SOME USEFUL RESOURCES:

There are discussions of analysing large amounts of qualitative data on the BigQual website, in a series of guest blog posts, including:

- Sian Lincoln and Brady Robards on narrative analysis http://bigqlr.ncrm.ac.uk/2017/11/08/guest-post-12-dr-sian-lincoln-and-dr-brady-robards-facebook-timelines-young-peoples-growing-up-narratives-online/

See also:

- Edwards, R. and Weller, S. (2012) Shifting analytic ontology: using I-poems in qualitative longitudinal analysis, *Qualitative Research* 12(2): 202-217.
- Phoenix, A., Boddy, J., Edwards, R. and Elliott, H. (2017) 'Another long and involved story': narrative themes in the marginalia of the Poverty in the UK survey, in R. Edwards et al. (eds) *Working With Paradata, Marginalia and Fieldnotes*, Cheltenham: Edward Elgar.

There are also textbooks on qualitative analysis, including:

- Braun, V. and Clarke, V. (2013) *Successful Qualitative Research: A Practical Guide for Beginners*, London: Sage.
- Coffey, A. and Atkinson, P. (1996) *Making Sense of Qualitative Data: Complementary Research Strategies*, London: Sage.