How and what to code

Authors of this page: Graham R. Gibbs and Celia Taylor
Affiliation: University of Huddersfield
Date written: 30th June 2005
Updated 19th Feb 2010
Online QDA Web Site,
[onlineqda.hud.ac.uk/Intro_QDA/how what to code.php]

Coding

Coding is the process of combing the data for themes, ideas and
categories and then marking similar passages of text with a code
label so that they can easily be retrieved at a later stage for
further comparison and analysis. Coding the data makes it easier
to search the data, to make comparisons and to identify any patterns that require further
investigation.

Codes can be based on:
- Themes, Topics
- Ideas, Concepts
- Terms, Phrases
- Keywords

found in the data. Usually it is passages of text that are coded but it can be sections of an audio or
video recording or parts of images. All passages and chunks that are coded the same way – that is
given the same label – have been judged (by the researcher) to be about the same topic, theme,
concept etc.

The codes are given meaningful names that gives an indication of the idea or concept that underpins
the theme or category. Any parts of the data that relate to a code topic are coded with the
appropriate label. This process of coding (associating labels with the text, images etc) involves close
reading of the text (or close inspection of the video or images). If a theme is identified from the data
that does not quite fit the codes already existing then a new code is created.

As the researcher reads through their data set the number of codes they have will evolve and grow as
more topics or themes become apparent. The list of codes thus will help to identify the issues
contained in the data set.

Approaches to starting coding

It is possible to start coding with themes identified from a priori ideas such as pre-existing theories or
just to let new codes emerge from your data set as you read it (grounded theory).

A priori codes

These can be identified from a range of sources:
- Previous research or theory
- Research or evaluation questions you are addressing
- Questions and topics from your interview schedule
- Your gut feeling about the data or the setting

Video on a priori coding using pen and paper.

Grounded codes

Grounded codes emerge from the data because you put aside your prejudices, presuppositions and
previous knowledge of the subject area and concentrate instead on finding new themes in your data.

What to look for when you are coding

Most typically, when coding, researchers have some codes already in mind and are also looking for
other ideas that seem to arise out of the data. When coding in this second, open minded manner,
Charmaz (writing in the grounded theory tradition) suggests you ask the following questions about the
data you are coding:
- What is going on?
- What are people doing?
- What is the person saying?
- What do these actions and statements take for granted?
- How do structure and context serve to support, maintain, impede or change these actions and
  statements?” (Charmaz 2003: 94-95)

A more detailed list of the kinds of things that can be coded are Table 1 below. The examples of each
type will be descriptive because it makes it easier to explain the phenomena. However, when you
are coding it is advisable to move from descriptive codes to more analytic ones as quickly as possible.

See the discussion in the next section.

<table>
<thead>
<tr>
<th>NO.</th>
<th>WHAT CAN BE CODED</th>
<th>EXAMPLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Behaviours, specific acts</td>
<td>Seeking reassurance, Bragging</td>
</tr>
<tr>
<td>2</td>
<td>Events – short once in a lifetime events or things people have done that are often told as a story.</td>
<td>Wedding day, day moved out of home for university, starting first job</td>
</tr>
<tr>
<td>3</td>
<td>Activities – these are of a longer duration, involve other people within a particular setting</td>
<td>Going clubbing, attending a night course, conservation work</td>
</tr>
<tr>
<td>4</td>
<td>Strategies, practice or tactics</td>
<td>Being nasty to get dumped, Staying late at work to get promotion</td>
</tr>
<tr>
<td>5</td>
<td>States – general conditions experienced by someone or found in</td>
<td>Homelessness “I’ll never meet</td>
</tr>
</tbody>
</table>
Some good articles or texts on coding.


During coding it is easy to generate a huge number of codes, particularly if you are using CAQDAS software. You need to manage your codes. Later on you may decide to refine them. See sorting the codes.

<table>
<thead>
<tr>
<th>1</th>
<th>Ideas – general wisdom experience of people of either ill or organisations</th>
<th>Implementers – in rare cases you can expect anyone better at my age setting for someone who is not really suitable</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Meanings – A wide range of phenomena at the core of much qualitative analysis. Meanings and interpretations are important pairs of what directs participants actions.</td>
<td>The term ‘chilling out’ is used by young people to mean relaxing and not doing very much</td>
</tr>
<tr>
<td>a.</td>
<td>What concepts do participants use to understand their world?</td>
<td>Jealousy &quot;I just felt why did she get him&quot;</td>
</tr>
<tr>
<td>b.</td>
<td>What norms, values, and rules guide their actions</td>
<td>A PhD is referred to as ‘a test of endurance’ (because finishing a PhD is a challenge)</td>
</tr>
<tr>
<td>c.</td>
<td>What meaning or significance it has for participants, how do they construe events what are the feelings</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Participation – adaptation to a new setting or involvement</td>
<td>About new neighbours “In my new house I have to keep my music down at night as the neighbours have young children”.</td>
</tr>
<tr>
<td>8</td>
<td>Relationships or interaction</td>
<td>Seeing family “Now my sister lives in the next road she visits more and we’ve become much closer.</td>
</tr>
<tr>
<td>9</td>
<td>Conditions or constraints</td>
<td>Lose of job (before financial difficulties, moving away (before lost contact with old friends)</td>
</tr>
<tr>
<td>10</td>
<td>Consequences</td>
<td>Confidence gets dates, positive attitude attracts opportunities</td>
</tr>
<tr>
<td>11</td>
<td>Settings – the entire context of the events under study</td>
<td>University, work place, housing data</td>
</tr>
<tr>
<td>12</td>
<td>Reflective – researcher’s role in the process, how intervention generated data</td>
<td>Probing question “How did you feel when he said that?”</td>
</tr>
</tbody>
</table>

| Table 1. Types of phenomena that can be coded (Adapted from Bogdan and Biklen, 1992; Strauss, 1987; Mason, 1996; and Gibbs, 2006) |

### Constant comparison

Many writers make suggestions about the ways you can approach your data so that you remain open minded about what can be coded and start to notice significant patterns in the data. Perhaps the most famous are those made by the grounded theorists.

The most common procedure they recommend is ‘constant comparison’. What this means is that every time you select a passage of text (or its equivalent in video etc.) and code it, you should compare it with all those passages you have already coded that way, perhaps in other cases. This ensures that your coding is consistent and allows you to consider the possibility either that some of the passages coded that way don’t fit as well (and might therefore be better coded as something else) or that there are dimensions or phenomena in the passages that might well be coded another way altogether. But the potential for comparisons doesn’t stop there. You can compare the passage with those code in similar or related ways or even compare them with cases and examples from outside your data set altogether.

Here is a short animation (3 mins) about sorting coloured beads that illustrates well the process of constant comparison. After an initial sort into piles (an initial categorisation) additional beads are sorted into the existing piles but as there are more beads (cases) to deal with it becomes clear that some piles can be divided into two (codes can be divided or sub-codes can be created). A simple metaphor, but entertaining!

For example, Strauss and Corbin (1990) devote a chapter called ‘Techniques for Enhancing Theoretical Sensitivity’ to examining some of the ways they recommend qualitative analysts might use to ensure that they look carefully at the data and explore all its dimensions (Strauss and Corbin, 1990 pp.75-95). This includes systematic comparisons and far out comparisons. In the former, you think about all the ways in which some phenomenon you have found in the data can vary and be treated and seen differently by people. For example, if your respondent has been talking about the way her parents continued to give her financial support after she had left them and set up her own home, you can connect this with all the other ways that parents might support their children, financial, emotional, child care, finding employment, setting up personal contacts, housework, do-it-yourself, gardening and many others. This allows you to think or discover new ways of coding the experiences this respondent and others in your study might have had.

In the case of far out comparisons, the comparison is made with cases and situations that is similar in some respects but quite different in others and may be completely outside the study. For example, still thinking about parental help, we might make a comparison with the way coaches help sports men and women. Reflecting on the similarities and differences between coaching and parental relationships might suggest in some other dimensions to parental help, like the way that coaches get paid for their work but parents don’t.

### Other techniques to identify themes and codes

Ryan and Bernard in a recent paper (2003b) suggest a number of ways in which those coding transcripts can discover new themes in their data. Drawing heavily on Strauss and Corbin (1990) they suggest these include:

- **Word repetitions** – look for commonly used words and words whose close repetition may indicated emotions
- **Indigenous categories** (what the grounded theorists refer to as in vivo codes) – terms used by respondents with a particular meaning and significance in their setting.
- **Key-words-in-context** – look for the range of uses of key terms in the phrases and sentences in which they occur.
- **Compare and contrast** – essentially the grounded theory idea of constant comparison. Ask, ‘what is this about?’ and ‘how does it differ from the preceding or following statements?’
- **Social science queries** – introduce social science explanations and theories, for example, to explain the conditions, actions, interaction and consequences of phenomena.
- **Searching for missing information** – essentially try to get an idea of what is not being done or talked about, but which you would have expected to find.
- **Metaphors and analogies** – people often use metaphor to indicate something about their key, central beliefs about things and these may indicate the way they feel about things too.
- **Transitions** – one of the discursive elements in speech which includes turn-taking in conversation as well as the more poetic and narrative use of story structures.
- **Connectors** – connections between terms such as causal (‘since’, ‘because’, ‘as’ etc) or logical...
and reorganisation of codes can form the basis for some key analysis of the data.

Organising codes into a coding frame or coding list

As well as marking the transcript or field notes to show what is coded as what, you should keep a separate list of the codes you have constructed and against each one write a short description. Next time you find a passage that you think can be coded with an existing code, you can see if it exists in your frame or list and if it does, check with the definition to be sure that it does fit there. If you can't find an appropriate code (there isn't one or the text doesn't fit with the definitions) then you can create you own codes for each line of text.

Eventually, you will have a large number of codes and you will find it necessary to sort them into some sort of order or into groups. One way to do this is using a hierarchy. You may find several codes group together as types or kinds of something. In that case move them closer together and put them either in a list of their own, or make them sub-codes a major code the kinds or types of which they all represent.

Exercise A

This short text is by a young woman who discovered she was pregnant. Read the 10 lines of text and decide which of the listed codes describes what is happening in each line. Go to Exercise A.

Interactive Exercise A - Amanda

This short text is by a young woman talking about leaving home. Read the 10 lines of text and then create you own codes for each line of text. Go to Exercise B.

Interactive Exercise B - Karen

This short text is of a young women talking about leaving home. Read the 10 lines of text and then create you own codes for each line of text. Go to Exercise B.

Training video on coding

This is a short presentation showing how to interpret a short piece of text and develop codes and how to mark the text with those codes...

Analytic/Theoretical coding

These are codes based on the analytical thinking by the researcher about why what is occurring in the data might be happening.

Exercise video from a lecture on coding that includes discussion of the move from descriptive to analytical codes.

Videos of a lecture by Graham R Gibbs on approaches to grounded theory. (March 9 2010.) Five short videos (7 to 22 minutes long) examine the core elements of the Glaser and Strauss approach and examine the three stages recomended by Strauss and Corbin.

A short, 3 minute video produced by the Open University as part of their Exploring Psychology course. It is No 5 Analysing an Interview.

Types of coding

Whatever techniques you adopt, the tendency, when you start coding is to create codes that are some kind of summary or précis of the text you are examining. This kind of coding is called descriptive coding because it essentially forms a summary description of what is in the transcript or text. An essential part of QDA is that you move on to develop codes that go beyond description and start to categorise and analyse the data. Two examples of mainly descriptive coding.

Descriptive coding

This is when coding is used to describe what is in the data. Here are two exercises to help demonstrate descriptive coding.

Exercise video from a lecture on coding that includes discussion of the move from descriptive to analytical codes.

Videos of a lecture by Graham R Gibbs on approaches to grounded theory. (March 9 2010.) Five short videos (7 to 22 minutes long) examine the core elements of the Glaser and Strauss approach and examine the three stages recomended by Strauss and Corbin.

A short, 3 minute video produced by the Open University as part of their Exploring Psychology course. It is No 5 Analysing an Interview.
Non- Hierarchical coding (flat coding)

A non-hierarchical arrangement of codes, like a list, there are no sub-code levels.

- Close, generalised friendships
- Sporting friendships
- Sports club members
- Work friends
- Making new friends - same sex
- Making new friends - different sex
- Losing touch with friends
- Becoming sexual relationships

(Remember, in a real code list, you should also include short definitions.)

Hierarchical coding (tree coding)

A hierarchical arrangement of codes, like a tree, a branching arrangement of sub-codes. Ideally, codes in a tree relate to their parents by being 'examples of...', or 'contexts for...' or 'causes of...' or 'settings for...' and so on.

- Friendship types
  - Close, generalized
  - Sporting
  - Club
  - Non-club
  - Work
- Changes in Friendship
  - Making new friends
  - New same sex friends
  - New different sex friends
  - Losing touch
  - Becoming sexual relationship

A short video from a lecture on coding that includes discussion of hierarchical coding frames.

Interactive Exercise C - Hierarchical coding

This exercise involves putting a list of codes into a hierarchical arrangement. Go to Exercise C.

Applying new codes

As you code the data you are likely to create new codes, you therefore need to go back and check the units of data you coded previous to creating this code. This is to check if there is any more data that should be coded at the newly created node.

Diagram to show how new codes should be applied to previously coded data.

It is important to apply all your codes to the whole data set because:
- Something you come across later on may change how you want to code the data
- You may not have noticed a new pattern in the data until you had coded a number of interviews
- If you only code the data units after you created the node, any previously coded data units will not be included in the searches involving this code

It is a lengthy process and you may find that previously coded data does not need to be coded at the newly created code but you could not know this without checking.

Memos and codes

It is important to keep written notes that are meaningful to you, during your coding process. These notes are often called memos. A major use for memos is to record longer definitions of the codes and to note any analytic thoughts you have about the significance and relationship to others of the code in question.

Typically, information you could incorporate into the memo about a code will include:
- why you have created the code
- some detail of what the code is about and what the coded text reveals
- why you have changed a code (for instance re-named it)
- thoughts and questions about the analysis that occur to you as you code

Memos are essential if you are working in a team and sharing the coding of the data so colleagues know why you have coded the data in that way. See Memos in Writing as Analysis.

Summary

- Coding involves categorising and indexing sections or chunks of your data,
- Codes can come from theory and explanations 'outside the data' and/or 'emerge from the data',
- Data formats that can be coded ranges from transcribed text to video,
- Coding often starts by being descriptive but needs to becomes analytical,
- Any new codes created should be applied to the whole data set (previously coded units of data),
- Memos should be used to record your thoughts and ideas about your codes during the process.

Five short videos from a lecture on coding.

References