

 SAGE researchmethods

# Thematic Coding and Categorizing

In: Analyzing Qualitative Data

**By:** Graham R. Gibbs

Pub. Date: 2012

Access Date: April 9, 2019

Publishing Company: SAGE Publications, Ltd

City: London

Print ISBN: 9780761949800

Online ISBN: 9781849208574

DOI: <https://dx.doi.org/10.4135/9781849208574>

Print pages: 38-55

© 2007 SAGE Publications, Ltd All Rights Reserved.

This PDF has been generated from SAGE Research Methods. Please note that the pagination of the online version will vary from the pagination of the print book.

## Thematic Coding and Categorizing

Codes and coding [38](#)

The mechanics of coding [41](#)

Data-driven or concept-driven? [44](#)

What to code [46](#)

Retrieving text from codes [48](#)

Grounded theory [49](#)

---

### Chapter objectives

After reading this chapter, you should

- see the central role of coding in qualitative analysis;
- see from the close examination of an example the importance of creating codes that are analytic and theoretical and not merely descriptive; and
- know two techniques that can be used to promote the move from description to analysis: constant comparison and line-by-line coding.

---

### Codes and coding

Coding is how you define what the data you are analyzing are about. It involves identifying and recording one or more passages of text or other data items such as the parts of pictures that, in some sense, exemplify the same theoretical or descriptive idea. Usually, several passages are identified and they are then linked with a name for that idea – the code. Thus all the text and so on that is about the same thing or exemplifies the same thing is coded to the same name. Coding is a way of indexing or categorizing the text in order to establish a framework of thematic ideas about it (see Box 4.1 for a discussion of these terms). Coding in this way enables two forms of analysis.

You can retrieve all the text coded with the same label to combine passages that are all examples of the same phenomenon, idea, explanation or activity. This form of retrieval is a very useful way of managing or organizing the data, and enables the researcher to examine the data in a structured way.

You can use the list of codes, especially when developed into a hierarchy, to examine further kinds of analytic questions, such as relationships between the codes (and the text they code) and case-by-case comparisons. This will be examined in Chapter 6.

---

### Box 4.1 Code, index, category or theme?

When you first come across it, the idea of a code might seem rather mysterious. You probably first think about it in terms of secret codes and ciphers. For others, the association with computer code and programming might come to mind. As it is used here, codes are neither secretive nor to do with programming. They are simply a way of organizing your thinking about the text and your research notes.

Writers on qualitative analysis use a variety of terms to talk about codes and coding. Terms such as indices, themes and categories are used. Each reflects an important aspect of coding. Ritchie and Lewis prefer the term 'index' as this captures the sense in which codes refer to one or more passages in the text about the same topic in the way that entries in a book index refer to passages in the book (Ritchie et al., 2003). In phenomenological analysis, a term that is used instead of codes is 'themes' (Smith, 1995; King, 1998). Again this captures something of the spirit of what is involved in linking sections of text with thematic ideas that reveal the person's experience of the world. Dey (1993) uses 'category', which indicates another aspect of coding. The application of names to passages of text is not arbitrary, it involves a deliberate and thoughtful process of categorizing the content of the text. Coding means recognizing that not only are there different examples of things in the text but that there are different *types* of things referred to.

To add confusion to this, quantitative researchers also use the term 'coding' when assigning numbers to survey question answers or categorizing answers to open-ended questions. The latter is somewhat like qualitative coding, but is usually done in order to count the categorized responses, which is not the prime motivation of qualitative researchers.

The structured list of codes and the rules for their application (their definitions) that result from qualitative analysis are sometimes referred to as a coding frame. Again, this is confusing, since quantitative researchers use this to refer to the listing that tells them what numeric value to assign to different answers in surveys so that they can be counted. For that reason I have avoided the term. Others use the term 'thematic framework' (Ritchie et al., 2003) or 'template' (King, 1998). Here I just refer to the list of codes, or the codebook, a term used by many other analysts. 'Book' suggests something more weighty than just a list and indeed it is good practice that you should keep more than just a list. The codebook is something that should be kept separate from any coded transcripts. It should include not only the current and complete list of your codes, arranged hierarchically if appropriate, but

also a definition for each along with any memos or analytic notes about the coding scheme that you have written.

Coding is easiest using a transcript. It is possible to code directly from an audio or video recording or from rough field notes, but it is neither easy to do this nor is it easy to retrieve the sections of recording or notes that have been coded when you need them. (The exception to this is when you are using CAQDAS and digital video or audio. Then the software makes it much easier to retrieve the sections of video or audio that you have coded.) In fact, a lot of the time, coding is best done with an electronic text file using dedicated analysis software. I shall examine this in Chapter 9, but here I shall explain techniques that can be done with a paper transcript. I actually use both paper-based and computer-based approaches myself. I find that paper allows me the kinds of creativity, flexibility and ease of access that is important at the early stages of analysis. I then transfer the coding ideas into the electronic version of the project in order to continue the analysis. Do not be afraid about using either just paper or just software or both. As long as you make certain preparations (like introducing your data into the software before you produce printed copy to work on), there is nothing to stop you moving, when you want to, from paper to the software. Of course, you don't have to use software at all. For most of the last century, those undertaking qualitative analysis did not or could not use software. Most of the classic studies using qualitative research were undertaken without electronic assistance.

## Code definitions

Codes form a focus for thinking about the text and its interpretation. The actual coded text is just one aspect of that. For this reason it is important that as early as you can you write some notes about each code you develop. In the previous chapter I introduced the idea of writing memos as an important way of recording the development of your analytic thinking. A key function of such memos is to note the nature of a code and the thinking that lies behind it and to explain how the code should be applied or what kinds of text, images, and so on should be linked to the code. Keeping such a record is important for two reasons:

It will help you apply the code in a consistent way. Without having to reread all the text already coded to this name, you will be able to decide if any new text should really be coded there.

If you are working in a team, it will enable you to share your codes with others for them to use and, if they have done the same, to use theirs. It is quite likely, if more than one member of the team is coding, that more than one person will come up with similar coding ideas. Having memos about the codes will enable you to tell if the codes are, in fact, identical or not.

Keep your code memos in one or more word-processing files (so you can easily edit them or print them out) or use large filing cards to record the details. Typically you will need to record:

- The label or name of the code that you have used in marking up and coding the transcript.

- Who coded it – the name of the researcher (not needed if you are working alone).
  - The date when the coding was done or changed.
  - Definition of the code – a description of the analytic idea it refers to and ways of ensuring that the coding is reliable, that is, carried out in a systematic and consistent way.
  - Any other notes of your thinking about the code, for example, ideas you may have about how it relates to other codes or a hunch that maybe the text coded here could actually be split between two different codes (see Box 3.2 for more ideas).
- 

## The mechanics of coding

Those new to coding often find one of the most challenging things to begin with is identifying chunks of text and working out what codes they represent in a way that is theoretical and analytic and not merely descriptive. This involves careful reading of the text and deciding what it is about. In the visual arts the term 'intensive seeing' is used to refer to the way that we can pay close attention to all the things we can see, even the commonplace and ordinary. In the same way, you need to undertake 'intensive reading' when coding. Charmaz suggests some basic questions to ask as you undertake this intensive reading that will help you get started:

- 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, pp. 94–5)

## An example

To illustrate this initial stage, consider the following example. It is taken from a study of carers for people with dementia and is an interview with Barry, who is now looking after his wife, who has Alzheimer's disease. The interviewer has just asked Barry, 'Have you had to give anything up that you enjoyed doing that was important to you?', and he replies:

BARRY

Well, the only thing that we've really given up is – well we used to go dancing. Well she can't do it now so I have to go on my own, that's the only thing really. And then we used to go indoor bowling at the sports centre. But of course, that's gone by the board now. So we don't go there. But I manage to get her down to works club, just down the road on the occasional Saturdays, to the dances. She'll sit

and listen to the music, like, stay a couple of hours and then she's had enough. And then, if it's a nice weekend I take her out in the 10 car.

## Description

At one level this is a very simple reply. In lines 2 to 6 Barry gives two examples of things that he and Beryl used to enjoy together, dancing and indoor bowling, then, without prompting, he lists two things that they still do together, visiting dances at the works club and going out for a drive. So a first idea is to code lines 2 to 4 to the code 'Dancing', lines 4 to 6 to 'Indoor bowling', 6 to 9 to 'Dances at works club' and 9 to 10 to 'Drive together'. Such coding might be useful if you are analyzing interviews with lots of carers and you wanted to examine the actual activities given up and those still done together and compare them between couples. Then retrieving all the text coded at codes about such activities would enable you to list and compare what people said about them.

## Categorization

However, such coding is simply descriptive; there are usually better ways to categorize the things mentioned and there are other things indicated by Barry's text. In analysis you need to move away from descriptions, especially using respondent's terms, to a more categorical, analytic and theoretical level of coding. For example, you can code the text about dancing and indoor bowling together at a code 'Joint activities ceased', and text on works club dances and driving together to the code 'Joint activities continuing'. Assuming you have done the same in other interviews, you can now retrieve all the text about what couples have given up doing and see if they have things in common. In so doing you have begun to categorize the text.

## Analytic codes

Thinking about this suggests another way to code the text. Both dancing and bowling are physical activities involving some degree of skilled movement. Clearly Beryl has lost that, so we could code lines 2 to 6 to the code 'Loss of physical co-ordination'. This code is now slightly more analytic than those we started with, which just repeated Barry's descriptions. Barry does not talk about loss of physical co-ordination, but it is implied in what he says. Of course you need to be careful. This is an interpretation, based, here, on very little evidence. You need to look for other examples in Barry's interview of the same thing and perhaps other evidence in what he says of Beryl's infirmity.

Another thing to notice about this text is the way Barry changes from using 'we' about what they used to do together, to saying 'I' when he turns to the things they do now. This suggests another pair of analytic codes, one about joint activity with a sense of being a couple, the other about activity where the carer is just doing things for his partner. You might code these as 'Togetherness' and 'Doing for'. Note that these codes do not simply code what happened, but rather suggest the way in which Barry thought about, or conceptualized, these things.

Other things you might have noticed about the passage that might be candidates for codes include Barry's rhetorical use of 'Well' in lines 2 and 3. He says it three times. Is this an indication of a sense of resignation, loss or regret? Again, from such a short passage it is not clear. But you might code it 'Resignation' for now and later see if it is consistent with other text of Barry's you have coded to 'Resignation'. It is interesting to note that Barry says he still goes dancing, on his own. A different interpretation of this use of 'well' and the fact that it is the first thing that Barry mentions, is that dancing was a key thing that he and Beryl did together as a couple. You might therefore think that it is a kind of core or central activity of the couple, something that was central to their life together as a couple. Again, it would be useful to examine other carers to see if there are similar defining activities and to see if this identifies any differences between carers. Perhaps carers where the defining activities have been less affected by Alzheimer's are different from those where it has.

In summary, here are the codes that might be used to code the passage by Barry.

*Descriptive codes:* 'Dancing', 'Indoor bowling', 'Dances at works club', 'Drive together'.

*Categories:* 'Joint activities ceased', 'Joint activities continuing'.

*Analytic codes:* 'Loss of physical co-ordination', 'Togetherness', 'Doing for', 'Resignation', 'Core activity'.

Of course, it is unlikely that you would use all these codes to code just one short passage like this, but I have used them here to illustrate the way you need to move from descriptive coding, close to the respondent's terms, to categorization and to more analytic and theoretical codes. Also notice that I have used the codes only once in this short text. Normally, you would look through the rest of the text to see if there are any more passages that can be coded to the same code and do the same with other participants.

How you develop these thematic codes and which of them you focus on will depend on the aim of the research. In many cases, research is driven by funding bodies and what you have agreed with the funders that you will do. For example, if the research on those suffering from Alzheimer's disease was funded by the bodies that provide services to carers, then you might focus on the themes 'Doing for' and 'Joint activities'. On the other hand, if you were doing a PhD on the social psychology of couples, you might focus on 'Core activity' and 'Togetherness'.

## Marking the coding

When using paper, coding is done by jotting the code name in the margin or by marking text with colour (either in the margin or using highlighter pens). Figure 4.1 shows some of these ways of indicating this coding on the transcript. There are boxes with linked names (I used arrows), shading (e.g. with a highlighter pen) and linked code name. The right-hand margin is used with brackets to indicate the lines coded. I have circled or highlighted some key words or terms such as emotive words, unusual terms, metaphors and words used for emphasis.

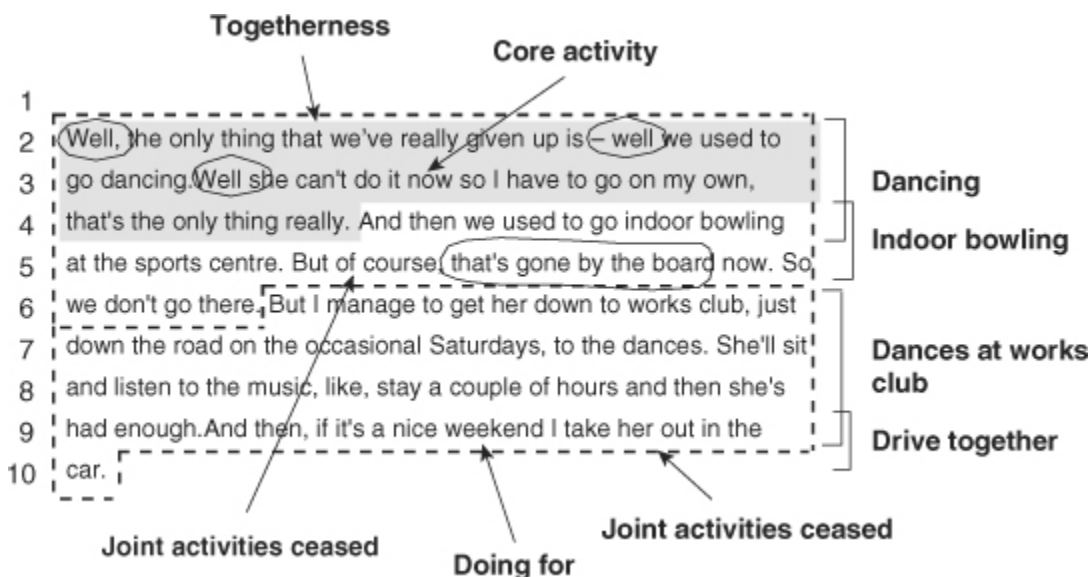
## Data-driven or concept-driven?

The construction of codes in a codebook is an analytic process. It is the building up of a conceptual schema. Although in the illustrations I have discussed the codes were derived from and are grounded in the data, it is possible to build a codebook without initial reference to the data collected.

### Concept-driven coding

The categories or concepts the codes represent may come from the research literature, previous studies, topics in the interview schedule, hunches you have about what is going on, and so on. It is possible to construct a collection of codes in a codebook without, at first, using them to code the data. Such a view is taken by Ritchie et al. (2003) in their advocacy of framework analysis. In framework analysis, before applying codes to the text, the researcher is encouraged to build up a list of key thematic ideas. These can be taken from the literature and previous research but are also generated by reading through at least some of the transcripts and other documents such as field notes, focus groups and printed documents. A similar view is taken by King (1998), who recommends the construction of a template, using similar sources of inspiration, which is a hierarchical arrangement of potential codes. In both King's template analysis and framework analysis, coding consists of the identification of chunks of text that exemplify the codes in this initial list. However, all these authors recognize that the researcher will need to amend the list of codes during analysis as new ideas and new ways of categorizing are detected in the text.

**FIGURE 4.1 Barry's reply with coding**



### Data-driven coding

The opposite of starting with a given list of codes is to start with none. This approach is usually called open coding (see the discussion later in this chapter), perhaps because one tries to do it with an open mind. Of



course, no one starts with absolutely no ideas. The researcher is both an observer of the social world and a part of that same world. We all have ideas of what we might expect to be happening and as social scientists we are likely to have more than most as a result of our awareness of theoretical ideas and empirical research. Nevertheless one can try, as far as possible, not to start with preconceptions. Simply start by reading the texts and trying to tease out what is happening. Such an approach is taken by the advocates of grounded theory (Glaser and Strauss, 1967; Strauss, 1987; Glaser, 1992; Strauss and Corbin, 1997; Charmaz, 2003) and by many phenomenologists in their concept of bracketing – setting aside presuppositions, prejudices and preliminary ideas about phenomena (Moustakas, 1994; Maso, 2001; Giorgi and Giorgi, 2003). But even they accept that a complete *tabula rasa* approach is unrealistic. The point is that, as far as possible, one should try to pull out from the data what is happening and not impose an interpretation based on pre-existing theory.

These two approaches to generating codes are not exclusive. Most researchers move backwards and forwards between both sources of inspiration during their analysis. The possibility of constructing codes before or separately from an examination of the data will reflect, to some extent, the inclination, knowledge and theoretical sophistication of the researcher. If your project has been defined in the context of a clear theoretical framework, then it is likely that you will have some good ideas about what potential codes you will need. That is not to say that they will be preserved intact throughout the project, but at least it gives you a starting point for the kinds of phenomena you want to look for when reading the text. The trick here is not to become too tied to the initial codes you construct.

---

## What to code

The example of coding I have discussed above is very short and specific to one context – caring for those suffering from dementia. What about interviews, notes and recordings on other topics? What other kind of things can be coded? The answer depends to some extent on the kind of analysis you are intending to do. Some disciplines and theoretical approaches like phenomenology, discourse analysis or conversation analysis will require that you pay special attention to certain kinds of phenomena in the texts you are examining.

Fortunately, for a very wide range of types of qualitative analysis that includes much policy and applied research and evaluation work as well as interpretive and hermeneutic approaches, there is a common ground of phenomena that researchers tend to look for in their texts. Some typical examples are listed in Table 4.1. Different authors have a different emphasis, but many of the ideas in the table will be useful to any analysis of texts.

Note that many of the examples in this table are rather descriptive. I have given these because it is easier to illustrate the phenomena with concrete examples. However, as I have suggested above, it is necessary to move from descriptions, especially those couched simply in terms used by participants, to more general and analytic categories. For example, rather than the event ‘Joining a sports club’ you might want to code this text to ‘Activity to make friends’ or ‘Commitment to keeping fit’ or even ‘Identity as a fit person’, which make

reference to the more general significance of this event.

**TABLE 4.1 What can be coded? (with examples)**


---

1.	Specific acts, behaviours – what people do or say.  Avoiding the question. Getting the opinions of friends.
2.	Events – these are usually brief, one-off events or things someone has done. It is not uncommon for the respondent to tell them as a story.  Being rejected at job interview. Moving into a homeless hostel. Finding husband has another woman. Joining a sports club.
3.	Activities – these are of longer duration than acts and often take place in a particular setting and may have several people involved.  Going dancing. Taking a training course. Helping partner with dementia get washed and dressed. Working in a bar.
4.	Strategies, practices or tactics – activities aimed towards some goal.  Using word of mouth to find jobs. Getting divorced for financial reasons. Entering a relationship to get somewhere to live.
5.	States – general conditions experienced by people or found in organizations.  Resignation, e.g. 'At my age it's hard to find work.' Working extra hours to get the job done.
6.	Meanings – a wide range of phenomena at the core of much qualitative analysis. Meanings and interpretations are important parts of what directs participants' actions.  (a) What concepts do participants use to understand their world? What norms, values, rules and mores guide their actions?  The idea of 'on-sight climbing' amongst rock climbers to describe doing a climb without inspection, artificial aids, pre-placed protection or previous practice, with the implication that this is a superior way of doing a climb.  (b) What meaning or significance does it have for participants, how do they construe events, what are their feelings?  Blame, e.g. 'His letter made me feel I was to blame.'  (c) What symbols do people use to understand their situation? What names do they use for objects, events, persons, roles, settings and equipment?  Delivery van referred to as 'the old bus' (affectionately or dismissively). Teaching referred to as 'work at the chalkface' (like work at the coalface, not administration).
7.	Participation – people's involvement or adaptation to a setting.  Adjusting to a new job, e.g. 'I find I have to be careful what I say now, because I know about things before they are finalized.'

---

---

8. Relationships or interaction – between people, considered simultaneously.

Enjoying the family, e.g. ‘. . . they’re 26 and 21 and most boys of that age are married, but mine aren’t and they like to come home, have friends to stay. I like that.’

9. Conditions or constraints – the precursor to or cause of events or actions, things that restrict behaviour or actions.

Firm’s loss of markets (before lay-offs). Divorce (before financial difficulties).

10. Consequences – What happens if . . .

Experience gets jobs, e.g. ‘So what you get is, people that haven’t got no qualification, but have got a few months’ experience are walking into jobs.’

11. Settings – the entire context of the events under study.

Hostel for the homeless. Training college. Day care centre.

12. Reflexive – the researcher’s role in the process, how intervention generated the data.

Expressing sympathy, e.g. ‘It must be hard for you in that situation.’

---

Adapted from Strauss (1987), Bogdan and Biklen (1992), Mason (1996).

---

## Retrieving text from codes

So far I have discussed coding mainly as a way of analyzing the content of the text. However, coding also has another, important purpose, which is to enable the methodical retrieval of thematically related sections of the text. There are several reasons for this:

- You can quickly collect together all the text coded in the same way and read it through to see what is at the core of the code.
- You can examine how, within a case, a coded thematic idea changes or is affected by other factors.
- You can explore how categorizations or thematic ideas represented by the codes vary from case to case, from setting to setting or from incident to incident.

Such retrieval activities will help you develop your analysis and your analytic and theoretical approach. For example, by reading the text you have coded to what might be a rather descriptive code used across several cases, you may discover some deeper, more analytic connection. You can then rename the code and rewrite its definition to indicate this idea, or perhaps create a new code and code relevant text to it.

## Practical retrieval

In order to retrieve the text to do this, you need to have taken some practical measures with your coded transcripts. All these kinds of retrieval are easiest if you are using CAQDAS. I will explore how in Chapter 8.

If you are using paper you will need to do two things:

- Gather together all the text coded with the same code in one place. You should produce many photocopies of your coded transcript so that you can cut up the sheets and store extracts with the same code in separate paper wallets, envelopes or files. If using a word processor, this can be achieved by copying and pasting the text into separate files for each code.
- Tag or label each extract (paper slip or electronically cut-and-pasted text) so that you can tell which document it came from. (If you use line numbers, these will tell you whereabouts in the document it came from. However, note that if you are cutting and pasting in a word processor, line numbering will not be preserved in the copy. In this case it is best simply to add a reference to the original line numbers along with the source tag.) If you have just a few documents, then just a couple of initials at the top of each extract to identify the document will do. But if you have a large number of documents/respondents, then a numbering system will help. A tag consisting of a string of letters or numbers that indicates not only the identity of the respondent but also some basic biographical information (like age group, gender and status) will help identify where the original text came from. You might use something like 'BBm68R' to indicate the interview with Barry Bentlow who is male, aged 68 and retired. Put this tag at the top of each extract or slip.

Such retrieval of the text coded by one code should be kept with any memos about the code so that you can ensure that the definition of the code still makes sense across all the extracts retrieved. If not you may need to recode some of the text or change the code definition. You can also check if any of your analytic ideas recorded in the memo elucidate the text you have retrieved or possibly write more in the memo after examining the retrieved text.

---

## Grounded theory

One of the most commonly used approaches to coding is grounded theory. This approach has been used extensively across a variety of social science disciplines and it lies behind the design of much CAQDAS. Its central focus is on inductively generating novel theoretical ideas or hypotheses from the data as opposed to testing theories specified beforehand. Insofar as these new theories 'arise' out of the data and are supported by the data, they are said to be grounded. It is only at a later stage of the analysis that these new ideas need to be related to existing theory. In their very accessible account of grounded theory, Strauss and Corbin (1990) present many specific ideas and techniques for achieving a grounded analysis. They divide coding into three stages:

*Open coding*, where the text is read reflectively to identify relevant categories.

*Axial coding*, where categories are refined, developed and related or interconnected.

*Selective coding*, where the 'core category', or central category that ties all other categories in the theory together into a story, is identified and related to other categories.

## Open coding

This is the kind of coding where you examine the text by making comparisons and asking questions. Strauss and Corbin also suggest it is important to avoid a label that is merely a description of the text. You need to try and formulate theoretical or analytic codes. The actual text is always an example of a more general phenomenon and the code title should indicate this more general idea. This is the hard part of coding. As you read the text, phrase by phrase, you should constantly ask questions: who, when, where, what, how, how much, why, and so on. This is designed to alert you to the theoretical issues lying behind the text and to give you a sensitivity to the deeper theoretical levels in them.

## Constant comparison

There are also several contrasts one can construct to help understand what might lie behind the surface text. The idea behind these contrasts or comparisons is to try to bring out what is distinctive about the text and its content. All too often we are so familiar with things that we fail to notice what is significant. Think about comparisons all the time as you go through doing your coding. This is one aspect of what is referred to as the method of constant comparison (Glaser and Strauss, 1967). Here are some examples of techniques suggested by Strauss, and Corbin (1990).

### Analysis of word, phrase or sentence

Pick out one word or phrase that seems significant, then list all its possible meanings. Examine the text to see which apply here. You may find new meanings that were not obvious beforehand.

### Flip-flop technique

Compare extremes on a dimension in question. For example, if someone mentions their age is a problem in finding work, try to contrast this with what it would be like for someone very young, just entering the job market, and someone else near the end of their working life. You may discover dimensions or issues you hadn't thought of before, such as the interaction of age and skills. Older people may lack new skills, but young people lack general work experience skills.

### Systematic comparison

Ask a series of 'what ifs' to explore all the dimensions of two phenomena. How do they differ, how do people respond differently? These are to try to stimulate you to recognize what is already there. For example, you can:

- Ask what if the circumstances, order of events, characteristics of the people, places, settings and so on were different.
- Ask how are the events and so on like and unlike others.

- Take a key element and free associate or read the text parts in a different order to try to stimulate ideas of what is in the text.

### **Far-out comparisons**

Take one element of the concept you are examining and think of the most remote or different example of some other phenomenon that shares some characteristics with that concept. Then work through all the other elements of both phenomena to see if they shed any light on the original. For example, you might compare a homeless man with a man who has had an arm amputated. Both suffer loss. Those without limbs experience stigma. Is that the same for the homeless? Those experiencing stigma deal with it by avoiding public places (hiding away), passing it off as others' problem and so on. Do the homeless do the same? Alternatively, you might compare the homeless who talk about their bad luck with gamblers' talk of a run of bad luck. Gamblers overestimate the extent to which they can control events. Is that the same for those looking for a home? In these cases the point about the comparison is to generate more codes that form dimensions, properties or aspects of the original idea.

### **Waving the red flag**

Be sensitive to phrases like 'Never', 'Always', 'It couldn't possibly be that way'. They are signals to look more closely. It is rarely the case that they are actually true. They usually mean things shouldn't happen that way. You need to find out what would happen if that situation actually did occur.

All these are good ways of encouraging more creative and deeper thinking about what is in the text. But in addition to these kinds of imaginative comparisons, it is important to carry out other kinds of comparisons. For example, you should compare what you have just coded with other text you have coded earlier or coded in a similar way. You can also compare the case you are working on with other cases you have researched. As you create new codes and code new text, it is worth checking to see whether text previously coded this way still makes sense now you have done some further coding. This is a matter of making sure you have consistently applied your coding across all the data you have. In some cases such comparisons may lead you to revise the codes you are using and/or the passages you have coded with them.

## **Line-by-line coding**

An approach recommended by many grounded theorists as a first step is line-by-line coding. This means going through your transcript and naming or coding each line of text, even though the lines may not be complete sentences. The idea is to force analytic thinking whilst keeping you close to the data. One of the dangers of coding, and of any kind of qualitative analysis, is importing your own motives, values and preoccupations into the codes and analytic scheme you construct. If you are not careful, your analysis may more closely reflect your own preconceptions and prejudices than the views of your respondents. One of the advantages of line-by-line coding is that it forces you to pay close attention to what the respondent is actually saying and to construct codes that reflect their experience of the world, not yours or that of any

theoretical presupposition you might have. On the other hand, line-by-line coding does not mean you should simply accept your respondents' views of the world. As I have suggested above, try to be more analytic and theoretical in your coding even if this means sometimes that your interpretations differ from that of your participants. Coding should remain grounded in the data in the transcript, but this does not mean it simply reflects respondents' view of things. Looking at the data line by line should stop you 'going native', that is, accepting your respondents' view of the world. You need to reflect that world-view, not accept it.

To illustrate line-by-line coding, consider the short extract in Fig. 4.2. This comes from a longer interview with a homeless man, Sam. (N.B in line 105 I have inserted the word 'term' in square brackets. This was not Sam's word, but it makes it clear what he means by 'long relationships'.) The example shows some initial, line-by-line coding. Some of these codes are still rather descriptive but they reflect the actions that Sam is talking about and some of the ways he sees the world and they prompt some examination of the rest of the transcript for comparisons. Line-by-line coding is just a way to get started and the next step is to develop and refine this coding.

The codes can be grouped in this way:

89 INTERVIEWER

90 Have you stayed in hostels for many years?

91 SAM

92 No, err but I've always moved around... since leaving school. I've  
 93 always been in a partnership, err, I've always seemed to. It's been a  
 94 long-term partnership so I've never been sort of out of partnerships  
 95 so it's not been too bad. For years and years I've lived with people.  
 96 But when I've had domestics and things like that, well you see, I  
 97 left home at fifteen years old and I've never been back to live with  
 98 my mum and dad. I'm one of them sorts of people who don't like  
 99 going and lying about on friends' couches or putting on people. So  
 100 really, yes, if I've had domestics and that, I've gone and slept in  
 101 'car – for days on end sometimes. But really this is my first time  
 102 that I've actually come away from everybody and lived by meself.  
 103 I have been homeless but I've never had a place by meself. I'm just,  
 104 like, one of them sorts of people that doesn't like putting on other  
 105 people. My problem is with me long [term] relationships. I make  
 106 friends easy when I'm in relationship. I get a lot of friends but  
 107 they're friends in that environment and that new place and what  
 108 happens then is I'll break up with her. Me head breaks up and I  
 109 don't like seeing, sort of seeing other person with someone else. So  
 110 I leave district then and move to another area, so it's starting again  
 111 I leave district then and move to another area, so it's starting again  
 112 then from scratch, that's what it is.

*Peripatetic lifestyle*  
*Partnership/relationship*  
*Long-term relationships*  
*Partnerships acceptable. Shared accomm.*  
*Domestics*  
*Chose independence*  
*Characterises self as independent*  
*Not reliant on others*  
*Domestics*  
*Slept in car*  
*Hostel seen as live alone*  
*Never lived on own*  
*Sees self as not reliant*  
*Relationships a problem*  
*Make friends easily*  
*Friends geog. limited*  
*Break up. Mental distress*  
*Jealousy*  
*Leave area. Avoidance*  
*Start again*



**FIGURE 4.2 Interview extract showing line-by-line coding**

89 INTERVIEWER	
90 Have you stayed in hostels for many years?	
91 SAM	
92 No, err but I've always moved around... since leaving school. I've	<i>Peripatetic lifestyle</i>
93 always been in a partnership, err, I've always seemed to. It's been a	<i>Partnership/relationship</i>
94 long-term partnership so I've never been sort of out of partnerships	<i>Long-term relationships</i>
95 so it's not been too bad. For years and years I've lived with people.	<i>Partnerships acceptable. Shared accomm.</i>
96 But when I've had <u>domestics</u> and things like that, well you see, I	<i>Domestics</i>
97 left home at fifteen years old and I've never been back to live with	<i>Chose independence</i>
98 my mum and dad. I'm one of them sorts of people who don't like	<i>Characterises self as independent</i>
99 <u>going and lying about on friends' couches or putting on people.</u> So	<i>Not reliant on others</i>
100 really, yes, if I've had <u>domestics</u> and that, I've gone and slept in	<i>Domestics</i>
101 'car – for days on end sometimes. But really this is my first time	<i>Slept in car</i>
102 that I've actually come away from everybody and lived by myself.	<i>Hostel seen as live alone</i>
103 I have been homeless but I've never had a place by myself. I'm just,	<i>Never lived on own</i>
104 like, one of them sorts of people that doesn't like putting on other	<i>Sees self as not reliant</i>
105 people. My problem is with me long [term] relationships. I make	<i>Relationships a problem</i>
106 friends easy when I'm in relationship. I get a lot of friends but	<i>Make friends easily</i>
107 they're friends in that environment and that new place and what	<i>Friends geog. limited</i>
108 happens then is I'll break up with her. <u>Me head breaks up and I</u>	<i>Break up. Mental distress</i>
109 don't like seeing, sort of seeing other person with someone else. So	<i>Jealousy</i>
110 I leave district then and move to another area, so it's starting again	<i>Leave area. Avoidance</i>
112 then from scratch, that's what it is.	<i>Start again</i>

I have omitted some repeated codes and clarified the names of one or two. All this grouping has done is gathered similar codes together. Looking at this grouping and the original transcript, you might begin to refine the codes. For example, there are a lot of codes about the ending of relationships. For Sam, relationships ending following what he called 'domestics' is clearly closely connected with his moving home and homelessness. The code 'Domestics' is what Glaser and Strauss (1967) refer to as an in vivo code. These are concepts used by the participants themselves to organize and conceptualize their world. Notice, though, that these are concepts, not just the respondent's words. In the case of Sam, 'domestics' clearly refers to some kind of argument or dispute with his partner of the time. His use of the term is itself puzzling. It is redolent of police and legal terms like 'domestic violence' and 'domestic disturbance'. Given that Sam tells us later in the interview that he has been to prison, we might therefore wonder whether these break-ups did involve the police and the legal system. In addition, his break-ups also involve some strong emotions like jealousy, so much so that he feels obliged to move out of the area. Notice in line 109 his use of the metaphor 'me head breaks up'. Again, later in the interview he explains how he has also spent some time in a mental hospital, so the distress is severe. Another key aspect of Sam's view of the world illustrated by this coding is his self-perception. Through repetition, he is clearly at some pains to portray himself as independent, not reliant on others and not someone who exploits his friends. Whether this is so is another matter, but he clearly

sees himself this way and thinks it important that the interviewer does too.

The next step after this initial line-by-line coding is to refine the actual codes and to rearrange them into a hierarchy. Refining serves two purposes. First, you will need to revisit the text to see whether it is better coded another way, for example using different codes coding longer passages, and whether there are examples elsewhere in the same transcript or in other transcripts that need coding using the new codes. It also provides a chance, as I discussed in the example in Fig 4.1, of making initially descriptive codes more analytic. Rearranging the codes into hierarchy will be discussed in Chapter 6.

---

## Key points

- Coding is a fundamental analytic process for many types of qualitative research. It consists of identifying one or more passages of text that exemplify some thematic idea and linking them with a code, which is a shorthand reference to the thematic idea. Having coded you can retrieve similarly coded text and compare how it varies across cases and with text coded in different ways.
- One of the most important issues of coding it to ensure that they are as analytic and theoretical as possible. You need to move away from codes that are simply descriptive and couched in the respondents' views of the world to codes that suggest new, theoretical or analytic ways of explaining the data.
- For some analysts the process of coding is one that involves the creation of new codes and with that, new analytic and theoretical understanding of your data. They suggest trying, as far as possible, to avoid applying existing frameworks to your data. Others, believing that a complete elimination of presuppositions is impossible, suggest starting with a framework or template of existing codes that reflect current analytic thinking.
- Grounded theory is an important example of a coding approach. The approach has some good suggestions about how to look for passages to code and how to identify the ideas they represent. This amounts to the recommendation to undertake a constant comparison: comparing similarly coded passages with each other, different codes with each other and coding in one case with other cases. A particular technique that grounded theorists suggest, which helps the creation of new codes, is line-by-line coding. Though this approach can be creative, there is still a need to ensure that the coding you come up with does not simply accept the participants' views of the world.

---

## Further reading

These three sources elaborate on the issues discussed in this chapter.

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

**Coffey, A. Atkinson, P.**(1996) *Making Sense of Qualitative Data Analysis: Complementary Research Strategies*. London: Sage.

**Mason, J.**(2002) *Qualitative Researching*. London: Sage.

<http://dx.doi.org/10.4135/9781849208574.n4>