

Saldana, J. (2009). 'An Introduction to Codes and Coding,' in: *The Coding Manual for Qualitative Researchers*. London: SAGE. pp. 1-8.

# ONE

## An Introduction to Codes and Coding

### CHAPTER SUMMARY

This chapter first presents the purposes and goals of *The Coding Manual for Qualitative Researchers*. It then provides definitions and examples of codes and categories and their roles in qualitative data analysis. The procedures and mechanics of coding follow, along with discussions of analytic software and team collaboration. The chapter concludes with reflections on necessary researcher attributes and the role of method in coding.

### Purposes of the Manual

The three primary purposes of the manual are:

- to discuss the functions of codes, coding, and analytic memo writing during the qualitative data collection and analytic processes;
- to profile a selected yet diverse repertoire of coding methods generally applied in qualitative data analysis; and
- to provide readers with sources, descriptions, recommended applications, examples, and exercises for coding and further analyzing qualitative data.

This manual does not address such matters as qualitative research design or how to conduct interviews or participant observation fieldwork. Those topics are already masterfully discussed in other textbooks. The manual is intended as a reference to supplement those existing works. It focuses exclusively on codes and coding and how they play a role in the qualitative data analytic process. For newcomers to qualitative inquiry it presents a repertoire of coding methods in broad brushstrokes. Additional information and extended discussion of the methods can be found in most of the cited sources. Grounded theory (discussed in Chapter Two), for example, is clearly profiled, streamlined, and re-envisioned in Kathy Charmaz's (2006) *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*, while Graham R. Gibbs' (2007) *Analysing Qualitative Data* provides an elegant survey of basic analytic processes.

The manual does not maintain allegiance to any one specific research genre or methodology. Throughout this book you will read a breadth of perspectives on codes and coding, sometimes purposely juxtaposed to illustrate and highlight the diverse opinions among scholars in the field. The following are just two examples of such professional divergence:

Any researcher who wishes to become proficient at doing qualitative analysis must learn to code well and easily. The excellence of the research rests in large part on the excellence of the coding. (Strauss, 1987, p. 27)

But the strongest objection to coding as a way to analyze qualitative research interviews is not philosophical but the fact that it does not and cannot work. It is impossible in practice. (Packer, 2011, p. 80)

No one, including myself, can claim final authority on coding's utility or the "best" way to analyze qualitative data. In fact, there are a few instances where I take moderate liberty with adapting and even renaming prescribed coding methods for clarity or flexibility's sake. This is not intended to standardize terminology within the field, but simply to employ consistency throughout this particular manual.

I must also emphasize at the very beginning that *there are times when coding the data is absolutely necessary, and times when it is most inappropriate for the study at hand*. All research questions, methodologies, conceptual frameworks, and fieldwork parameters are context specific. Also, whether you choose to code or not depends on your individual value, attitude, and belief systems about qualitative inquiry. For the record, here are mine, from *Fundamentals of Qualitative Research*:

Qualitative research has evolved into a multidisciplinary enterprise, ranging from social science to art form. Yet many instructors of research methods vary in their allegiances, preferences, and prescriptions for how to conduct fieldwork and how to write about it. I myself take a pragmatic stance toward human inquiry and leave myself open to choosing the right tool for the right job. Sometimes a poem says it best; sometimes a data matrix does. Sometimes words say it best; sometimes numbers do. The more well versed you are in the field's eclectic methods of investigation, the better your ability to understand the diverse patterns and complex meanings of social life. (Saldaña, 2011b, pp. 177–8)

Coding is just *one* way of analyzing qualitative data, not *the* way. Be cautious of those who demonize the method outright. And be equally cautious of those who swear unyielding affinity to codes, or what has been colloquially labeled "coding fetishism." I prefer that you yourself, rather than some presumptive theorist or hardcore methodologist, determine whether coding is appropriate for your particular research project.

I also wrote this manual because I found it problematic (but not difficult) to teach coding in my own qualitative research methods course. I provided students with an array of readings about the process from multiple sources because I had yet to find that single satisfactory book (to me) that focused exclusively on the

topic. General introductory texts in qualitative inquiry are so numerous and well written that it becomes difficult not to find the best one to use, but which one of such quality works to select as the primary textbook. This manual supplements introductory works in the subject because most limit their discussions about coding to the writer's prescribed, preferred, or signature methods. I wanted to provide in a single resource a selected collection of various coding methods developed by other researchers (and myself) that provides students and colleagues with a useful reference for classroom exercises and assignments, and for their own independent research for thesis and dissertation fieldwork and future qualitative studies. But by no means is it an exhaustive resource. I deliberately exclude such discipline-specific methods as psychotherapy's Narrative Processes Coding System (Angus, Levitt, & Hardtke, 1999), and such signature methods as the Davis Observation Code system for medical interviews (Zoppi & Epstein, 2002, p. 375). If you need additional information and explanation about the coding methods, check the References.

This manual is intended primarily as a reference work. It is not necessarily meant to be read cover to cover, but it certainly can be if you wish to acquaint yourself with all 32 coding methods profiles and their analytic possibilities. There are, in fact, several principles related to coding matters not discussed in the first two chapters that are unique to some of the profiles. If you choose to review all the contents, read selected sections at a time, not all of them in one sitting, otherwise it can overwhelm you. If you are scanning the manual to see which coding method(s) might be appropriate for your particular study, read the profiles' Description and Applications sections to see if further reading of the profile is merited, or check the glossary in Appendix A. It is doubtful you will use every coding method included in this manual for your particular research endeavors throughout your career, but they are available here on an "as-needed" basis for your unique projects. Like an academic curriculum, the sequential order of the profiles has been carefully considered. They do not necessarily progress in a linear manner from simple to complex, but are clustered generally from the fundamental to the intermediate to the advanced.

## What Is a Code?

A code in qualitative inquiry is most often a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data. The data can consist of interview transcripts, participant observation field notes, journals, documents, drawings, artifacts, photographs, video, Internet sites, e-mail correspondence, literature, and so on. The portion of data to be coded during First Cycle coding processes can range in magnitude from a single word to a full paragraph to an entire page of text to a stream of moving images. In Second Cycle coding processes, the portions coded can be the exact same units, longer passages of text, analytic memos about the data, and even a reconfiguration of the codes themselves developed thus far. Charmaz (2001) describes coding as the "critical link" between data collection and their explanation of meaning.



Do not confuse the use of *code* in qualitative data analysis with the use of *code* in the field of semiotics, even though there are some slight parallels between the two applications. In semiotics, a code relates to the interpretation of symbols in their specific social and cultural contexts. In qualitative data analysis, a code is a researcher-generated construct that symbolizes and thus attributes interpreted meaning to each individual datum for later purposes of pattern detection, categorization, theory building, and other analytic processes. Just as a title represents and captures a book, film, or poem's primary content and essence, so does a code represent and capture a datum's primary content and essence.

### Coding examples

An example of a coded datum, as it is presented in this manual, looks like this when taken from a set of field notes about an inner city neighborhood. The one-word capitalized code in the right column is called a Descriptive Code, which summarizes the primary topic of the excerpt:

<p><sup>1</sup> I notice that the grand majority of homes have chain link fences in front of them. There are many dogs (mostly German shepherds) with signs on fences that say "Beware of the Dog."</p>	<p><sup>1</sup> SECURITY</p>
---	------------------------------

Here is an example of several codes applied to data from an interview transcript in which a high school senior describes his favorite teacher. The codes are based on what outcomes the student receives from his mentor. Note that one of the codes is taken directly from what the participant himself says and is placed in quotation marks – this is called an In Vivo Code:

<p><sup>1</sup> He cares about me. He has never told me but he does.</p>	<p><sup>1</sup> SENSE OF SELF-WORTH</p>
<p><sup>2</sup> He's always been there for me, even when my parents were not. He's one of the few things that I hold as a constant in my life. So it's nice.</p>	<p><sup>2</sup> STABILITY</p>
<p><sup>3</sup> I really feel comfortable around him.</p>	<p><sup>3</sup> "COMFORTABLE"</p>

Did you agree with the codes? Did other words or phrases run through your mind as you read the data? It is all right if your choices differed from mine. Coding is not a precise science; it is primarily an interpretive act. Also be aware that a code can sometimes *summarize*, *distill*, or *condense* data, not simply *reduce* them. Madden (2010), in fact, notes that such analytic work does not diminish but "value adds" to the research story (p. 10).

The introductory examples above were kept purposely simple and direct. But depending on the researcher's academic discipline, ontological and epistemological orientations, theoretical and conceptual frameworks, and even the choice of coding method itself, some codes can attribute more evocative meanings to data. In the





excerpt below, a mother describes her teenage son’s troubled school years. The codes emerge from the perspective of middle- and junior high school years as a difficult period for most youth. They are not specific types of codes; they are “first-impression” phrases derived from an open-ended process called Eclectic Coding:

<sup>1</sup> My son, Barry, went through a really tough time about, <sup>1</sup> MIDDLE-SCHOOL probably started the end of fifth grade and went into <sup>1</sup> HELL sixth grade. <sup>2</sup> When he was growing up young in school <sup>2</sup> TEACHER’S PET he was a people-pleaser and his teachers loved him to death. <sup>3</sup> Two boys in particular that he chose to try to <sup>3</sup> BAD INFLUENCES emulate, wouldn’t, were not very good for him. <sup>4</sup> They <sup>4</sup> TWEEN ANGST were very critical of him, they put him down all the time, and he kind of just took that and really kind of internalized it, I think, for a long time. <sup>5</sup> In that time <sup>5</sup> THE LOST BOY period, in the fifth grade, early sixth grade, they really just kind of shunned him all together, and so his network as he knew it was gone.

Note that when we reflect on a passage of data to decipher its core meaning, we are *decoding*; when we determine its appropriate code and label it, we are *encoding*. For ease of reference throughout this manual, *coding* will be the sole term used. Simply understand that coding is the transitional process between data collection and more extensive data analysis.

## Coding for patterns

In the examples presented thus far, each unit of data was assigned its own unique code. This is due primarily to the short length of the excerpts. In larger and complete data sets, you will find that several to many of the same codes will be used repeatedly throughout. This is both natural and deliberate – natural because there are mostly repetitive patterns of action and consistencies in human affairs, and deliberate because one of the coder’s primary goals is to find these repetitive patterns of action and consistencies in human affairs as documented in the data. In the example below, note how the same Process Code (a word or phrase which captures action) is used twice during this small unit of elementary school classroom activity:

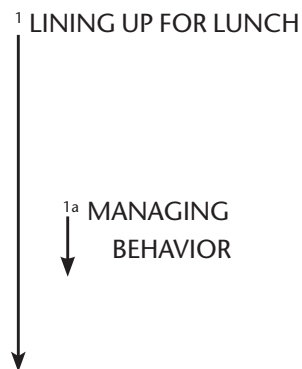
<sup>1</sup> Mrs. Jackson rises from her desk and announces, <sup>1</sup> LINING UP FOR LUNCH  
 “OK, you guys, let’s get lined up for lunch. Row One.” Five children seated in the first row of desks rise and walk to the classroom door.  
 Some of the seated children talk to each other.  
<sup>2</sup> Mrs. Jackson looks at them and says, <sup>2</sup> MANAGING BEHAVIOR  
 “No talking, save it for the cafeteria.  
<sup>3</sup> Row Two.” Five children seated in the <sup>3</sup> LINING UP FOR LUNCH



second row of desks rise and walk to the children already standing in line.

Another way the above passage could be coded is to acknowledge that MANAGING BEHAVIOR is not a separate action or an interruption of the routine that disrupts the flow of LINING UP FOR LUNCH, but to interpret that MANAGING BEHAVIOR is an embedded or interconnected part of the larger social scheme that composes LINING UP FOR LUNCH. The coding might appear thusly, using a method called Simultaneous Coding (which applies two or more codes within a single datum):

<sup>1</sup> Mrs. Jackson rises from her desk and announces, "OK, you guys, let's get lined up for lunch. Row One." Five children seated in the first row of desks rise and walk to the classroom door. Some of the seated children talk to each other.  
<sup>1a</sup> Mrs. Jackson looks at them and says, "No talking, save it for the cafeteria."  
<sup>1</sup> Row Two." Five children seated in the second row of desks rise and walk to the children already standing in line.



Take note of some important caveats when it comes to understanding patterns and regularity: idiosyncrasy is a pattern (Saldaña, 2003, pp. 118–22) and there can be patterned variation in data (Agar, 1996, p. 10). Sometimes we code and categorize data by what participants talk about. They may all share with you their personal perceptions of school experiences, for example, but their individual value, attitude, and belief systems about education may vary greatly from being bored and disengaged to being enthusiastic and intrinsically motivated. When you search for patterns in coded data to categorize them, understand that sometimes you may group things together not just because they are exactly alike or very much alike, but because they might also have something in common – even if, paradoxically, that commonality consists of differences.

For example, each one of us may have a strong opinion about who should be leading our country. The fact that we each have an individual opinion about that issue is what we have in common. As for *whom* we each believe should be leading the country, that is where the differences and variations occur. Acknowledge that a confounding property of category construction in qualitative inquiry is that data cannot always be precisely and discretely bounded; they are within “fuzzy” boundaries at best (Tesch, 1990, pp. 135–8). That is why a method called Simultaneous Coding is an option we have, when needed. Finally, Hatch (2002) offers that you think of patterns not just as stable regularities but as varying forms. A pattern can be characterized by:

- similarity (things happen the same way)
- difference (they happen in predictably different ways)

- frequency (they happen often or seldom)
- sequence (they happen in a certain order)
- correspondence (they happen in relation to other activities or events)
- causation (one appears to cause another) (p. 155)

## Coding filters

The act of coding requires that you wear your researcher's analytic lens. But how you perceive and interpret what is happening in the data depends on what type of filter covers that lens. For example, consider the following statement from an older male: "There's just no place in this country for illegal immigrants. Round them up and send those criminals back to where they came from." One researcher, a grounded theorist using In Vivo Coding to keep the data rooted in the participant's own language, might code the datum this way:

<sup>1</sup> There's just no place in this country for illegal immigrants. Round them up and send those criminals back to where they came from.	<sup>1</sup> "NO PLACE"
---	-------------------------

A second researcher, an urban ethnographer employing Descriptive Coding to document and categorize the breadth of opinions stated by multiple participants, might code the same datum this way:

<sup>1</sup> There's just no place in this country for illegal immigrants. Round them up and send those criminals back to where they came from.	<sup>1</sup> IMMIGRATION ISSUES
---	---------------------------------

And a third researcher, a critical race theorist employing Values Coding to capture and label subjective perspectives, may code the exact same datum this way:

<sup>1</sup> There's just no place in this country for illegal immigrants. Round them up and send those criminals back to where they came from.	<sup>1</sup> XENOPHOBIA
---	-------------------------

The collection of coding methods in this manual is a repertoire of possible filters to consider and apply to your approaches to qualitative inquiry. But even before that, your level of personal involvement as a participant observer – as a peripheral, active, or complete member during fieldwork – filters how you perceive, document, and thus code your data (Adler & Adler, 1987). So do the types of questions you ask and the types of responses you receive during interviews (Kvale & Brinkmann, 2009), the detail and structuring of your field notes (Emerson, Fretz, & Shaw, 2011), the gender, social class, and race/ethnicity of your participants – and yourself (Behar & Gordon, 1995; Stanfield & Dennis, 1993), and whether you collect data from adults or children (Greene & Hogan, 2005; Tisdall, Davis, & Gallagher, 2009; Zwiers & Morrisette, 1999).



Merriam (1998) states, “our analysis and interpretation – our study’s findings – will reflect the constructs, concepts, language, models, and theories that structured the study in the first place” (p. 48). And it is not only your approach to qualitative inquiry (e.g., case study, ethnography, phenomenology) and ontological, epistemological, and methodological issues that influence and affect your coding decisions (Creswell, 2013; Mason, 2002). Sipe and Ghiso (2004), in their revealing narrative about coding dilemmas for a children’s literacy study, note that “All coding is a judgment call” since we bring “our subjectivities, our personalities, our predispositions, [and] our quirks” to the process (pp. 482–3). Like the characters in director Akira Kurosawa’s classic film, *Rashômon*, multiple realities exist because we each perceive and interpret social life from different points of view.

### Coding as a heuristic

The majority of qualitative researchers will code their data both during and after collection as an analytic tactic, for coding *is* analysis (Miles & Huberman, 1994, p. 56). Differing perspectives, however, attest that “Coding and analysis are not synonymous, though coding is a crucial aspect of analysis” (Basil, 2003, p. 145). Coding is a heuristic (from the Greek, meaning “to discover”) – an exploratory problem-solving technique without specific formulas or algorithms to follow. Coding is only the initial step toward an even more rigorous and evocative analysis and interpretation for a report. Coding is not just labeling, it is *linking*: “It leads you from the data to the idea, and from the idea to all the data pertaining to that idea” (Richards & Morse, 2007, p. 137).

And, coding is a cyclical act. Rarely is the First Cycle of coding data perfectly attempted. The Second Cycle (and possibly the third and fourth cycles, and so on) of recoding further manages, filters, highlights, and focuses the salient features of the qualitative data record for generating categories, themes, and concepts, grasping meaning, and/or building theory. Coffey and Atkinson (1996) propose that “coding is usually a mixture of data [summation] and data complication ... breaking the data apart in analytically relevant ways in order to lead toward further questions about the data” (pp. 29–31).

Dey (1999) posits, though his original intent was to be critical, “With categories we impute meanings, with coding we compute them” (p. 95). To some, *code* is a “dirty four-letter word.” A few research methodologists perceive a code as mere shorthand or an abbreviation for the more important category yet to be discovered. Unfortunately, some use the terms *code* and *category* interchangeably and even in combination when they are, in fact, two separate components of data analysis. I advocate that qualitative codes are essence-capturing and essential elements of the research story that, when clustered together according to similarity and regularity (a pattern), they actively facilitate the development of categories and thus analysis of their connections. Ultimately, I like one of Charmaz’s (2006) metaphors for the process when she states that coding “generates the bones of your analysis. ... [I]ntegration will assemble those bones into a working skeleton” (p. 45).