

Coding and thematic analysis explained in 5 minutes

If I assigned you a task of watching 10 episodes of your favorite TV show and then trying to replicate the content as closely as possible. For example, describe it to me what would you do? Most likely you would want to start by breaking it down into smaller pieces to reduce the volume of this content and wrap your head around it. You would, for example, want to describe the different separate scenes and detail across each episode and across the whole show. Here they are talking about this problem they have, and here they are driving to the airport. Here they're fighting over a girl, here someone is having a dinner with his or her partner. If then I asked you a question about some occurring trend or behavior across these various episodes, you would probably try to somehow make sense of the list of various scenes and behaviors that you compiled by looking at them, grouping them when possible to reduce the volume of this data further and then thinking of similarities or trends to eventually come up with the answer to my questions.

If I asked you, for example, what is the most often used strategy to achieve goals in Game of Thrones, you would eventually conclude that things like negotiation treaties or battle are what the characters do. So in the above example what you did is in essence first code the data which is the part where you describe it in detail and developed themes to answer my research question and this is exactly what happens in thematic analysis. Thematic analysis involves coding, then revising and organizing these codes and concludes with using the codes to come up with themes. Which are essentially topics and patterns that demonstrate how the data answers the research questions. But it always starts with coding.

Although the name coding may be a little bit overwhelming at first, in practice is actually a relatively straightforward process. It involves labeling the data with specific names to reduce the amount of information and break it down into more digestible chunks in order to eventually understand the data so well that we can describe it in terms of themes or topics that help us answer the research questions. When you think about the example that I started with coding is describing the different scenes, events, and sequences. In the case of interview data it would involve describing what is being said in a given line, sentence or paragraph. Approaches to coding may differ in how detailed they are but I generally recommend being quite detailed.

As you build up this list of codes, it helps to think of them as a table of contents of your data. You want to make sure that this list accurately describes what was said in the data, so that you can rely on this list without having to read the full transcripts again after a while and this may either happen after coding several files or when you're done coding all of them you may start to notice some patterns either in the same codes being used again and again or maybe in seeing codes that share some characteristics or are related to some broader topic. At the same time, it is likely that your coding framework is a bit messy which is understandable and a normal part of the coding process. It is a sign, however, that it may be the time to review and organize these codes. This does not have to be anything complex and you just want to think of any common-sense way to organize them that will help you make sense of what you have in your data. It's just as if you were cleaning a room and decided to organize things into several piles, clothes here, pencils there, etc.

Same thing happens with the codes at this stage, in a study of a teacher's experiences of teaching during the pandemic, for example, you may notice that you have several codes referring to struggles and challenges and decide to group these codes together. You may create a separate group that involves codes about perceived benefits or codes about coping strategies, if you feel that this helps you organize and make sense of the data. The point here is just that, to make sense of the data and understand even better what it is telling us. After this stage you're moving towards developing your themes which by the way do have to be developed by you, they don't just emerge by themselves. We often talk about emerging themes, but as much as I like this to be the case they just don't emerge by themselves, so you do have to develop your themes. And what this means is that you will be looking at various codes categorized into several groups and then looking at your research questions and aims trying to establish what you really know about the data now and how these codes help you answer the research question. Once you have the answer to this you will start building your thematic framework using your available tools which are your codes.

The aim here is to make sure that this framework includes everything needed to tell the reader the story of your data, that it tells the reader the answers to your research questions. You want to make sure that the reader knows what you know and what you want him or her to know about your data and this is the final aim of thematic analysis. If you'd like to know more about the details of these processes described in this video, feel free to look at the list of resources below the video. Meanwhile, if you enjoyed this video and learned something new, please leave a like to help others find it.