



## Coding the Data

**Abstract** Next to gathering qualitative data, coding of data lies at the core of qualitative research. The chapter is the first of two that dig into the specificities of data analysis. Coding is one way of organising dense data and making sense of it, for instance by identifying patterns. Whilst there exist various approaches to coding, this chapter presents a set of strategies to code dense interview, ethnographic and document data. Additionally, we present the technicalities of coding data collaboratively, as part of teamwork. Importantly, Chapter 5 yields important tips and concrete examples regarding the use of software tools for qualitative analysis, such as Atlas.ti and the intricacies of using it as a team. Specifically, the chapter discusses the initial stages of developing code lists in inductive and deductive ways, the technicalities pertaining to actually coding the text of the data with Atlas.ti, and presents an overview of the advantages of some tools, such as creating code families, to make sense of the data. Throughout the chapter, we discuss the collaborative nature of our coding work by reviewing the pros and cons and by examining issues pertaining to intercoder reliability.

**Keywords** Data analysis · Coding practices · ATLAS.ti · Teamwork · Benefits · Challenges

## INTRODUCTION

Coding organises dense data into manageable amounts and helps to make sense of it by revealing trends and patterns. Whilst Chapters 3 and 4 introduced the data and explained how it was gathered, the next two chapters are attentive to how we moved towards data analysis. Here we are concerned with the processes and strategies for coding qualitative data, whilst amongst other things, Chapter 6 looks more closely at rendering it more accessible to interpretation.

There are various ways to code qualitative research (Saldaña, 2021), but most coding processes typically trigger the emergence of dominant themes which can then be analysed (Coffey & Atkinson, 1996; Strauss, 1987). Whilst some see coding as a way to merely organise the data (Coffey & Atkinson, 1996), we see the organisation of the data as more than a preliminary stage prior to analysis (Weston et al., 2001); for us, coding and interpretation are necessarily intertwined (Corbin & Strauss, 2008; Tesch, 1990). Coding is an important element because it helps to make sense of the data in relation to one's research questions and objectives (Elliott, 2018) and sets rigorous foundations for its interpretation.

In practical terms, the coding process can be thought of as circulating between reading, coding and thinking about the raw material in terms of overarching concepts and categories, whilst also comparing and contrasting the coded material. This corresponds to adopting a flexible and iterative approach to coding, in the sense that it is possible to revise the initial code list and to code the data multiple times in a back-and-forth process that allows for recalibration and refinement, and the investigation of new research questions that emerge from the data (Yin, 2011). In practice, this means being open to the kind of patterns and 'meta-narratives' that can arise in the data, and concomitantly being willing to go back to 'square one' in terms of research assumptions and expectations.

This approach can be especially successful when new research themes, questions and eventually findings emerge through immersing oneself in, coding, discussing and interpreting data as an ongoing interaction rather than as separate stages. From our own experience, prominent illustrations of what such a strategy can bring include important findings that went beyond the main themes of our research project—'gender equality'—and extended to findings on racism (Kantola et al., 2023), Brexit (Kantola & Miller, 2023), the impact of the Covid-19 pandemic (Elomäki & Kantola,

2022) and on the role and powers of national party delegations in the European Parliament (Elomäki et al., 2023). These unforeseen insights were revealed by a flexible coding scheme in which we could collaboratively discuss missing themes in the initial code list, and develop additional codes as we were coding the data (Deterding & Waters, 2021). We applied systematic and transparent principles by keeping track of our discussions and decision-making processes in a research diary (see Box 5.1). In doing so, we respond to the lack of transparency regarding decision-making in collaborative research processes (Reyes et al., 2021), which we seek to make visible to other researchers here. For practical purposes, this meant that we all needed to be fully acquainted with the material, even if we were not the ones who conducted particular interviews.

#### Box 5.1 Examples of Team Research Diary Entries

- ‘Coded my first two ethnographic fieldnotes. Feels quite different from interviews due to the structure. A lot more on affects and embodiment, very nice!’ (EUGenDem research diary 19 Feb 2021)
- ‘Was strange to code after such a long time! First I needed to re-read the code definitions to pick the right ones. And then it was a written one which has such a different flow... Was funny though that there were quite some affects in the responses, also regarding the researcher role!’ (EUGenDem research diary 17 Nov 2020)
- ‘A little difficult to follow—Interviewee sometimes talking about other things, such as her phone battery dying.’ (EUGenDem research diary 17 March 2020)
- ‘Surprising almost how little the interviewees talk about Covid, it is often present implicitly, in the goodbyes (stay safe, these strange times), or in references to having to postpone stuff or change plans’. (EUGenDem research diary 22 Feb 2021)
- ‘I remember I had a feeling in the interview that the guy really didn’t feel like talking to me and that everything I asked was somehow obvious to him. Reading the transcription through,

it's not too tragic (even though rather short)'. (EUGenDem research diary 9 Nov 2020)

Due to the research design of our project, interviews and ethnography notes were compiled together in ATLAS.ti and coded collaboratively, whereas document data were selected and coded separately for the purpose of research articles according to their own research design and questions. Therefore, we begin by discussing the process of developing a collaborative coding strategy using ATLAS.ti, before discussing the challenges this posed for a large team of researchers, and presenting how documents were finally coded.

### ATLAS.TI

ATLAS.ti was used to code and analyse our data. There are several scientific software packages for coding qualitative data but we opted for ATLAS.ti, not least because we had the expertise for this software within our team, and it was available freely (to us) through our institutions. The greatest volume of data we collected came in the form of interview data—the type of data for which ATLAS.ti is best suited. ATLAS.ti is best thought of as the vessel that houses the data, and along with researchers, facilitates a less complicated navigation through the coding process. In this sense, it stores and exchanges anonymised data; shares the workload of handling and coding dense data; keeps track of errors and inconsistencies amongst coders; and significantly highlights new ideas for codes. By using it collaboratively and extensively, that is by making full use of its functionality, helped respect the principles of transparency that are so important to the integrity of qualitative research (Reyes et al., 2021). The package offers export functions that enable the team to save and export entire ‘projects’—including raw data, code-books, coding links and research diaries (memos). The latter is described by some as ‘the substantive heart of qualitative data analysis’ (Reyes et al., 2021, 6) as they keep track of researchers’ reflections during the coding process, and help to make the decision-making process more transparent.

Whilst ATLAS.ti is self-explanatory and intuitive, some preparation from all the researchers is recommended, in particular if the team intends to code the data as it is gathered. Such preparation includes reading

selected academic texts on software-based data coding (Ahrens, 2018; Friese, 2012; Paulus & Lester, 2016), familiarisation with the ATLAS.ti handbook and watching tutorial videos.<sup>1</sup> The fact that we had in our team a member who was already familiar and experienced with the use of such software was beneficial, as she could explain and teach ATLAS.ti to the rest of us. However this is by no means a precondition and ATLAS.ti does provide detailed instructions.

## DEVELOPING A COLLABORATIVE CODING STRATEGY FOR INTERVIEWS AND ETHNOGRAPHIC DATA

Along with familiarity with ATLAS.ti, it was essential to develop a coding strategy. Both the design of the research project and the nature of the data gathered will influence the coding strategy. In our case, the coding strategy accommodated two levels of complexity: first, our data consisted of two different types of very dense data in the form of interview transcripts and ethnographic notes, and secondly, this had to be coded collaboratively. Our coding strategy combined the interviews and ethnographic data so they were treated as analogous and subjected to the same coding framework. This was our preference, although we acknowledge that scholars have debated this at great length, differing over how, and if at all, ethnographic data should be shared with other researchers (Contreras, 2019; Guenther, 2009; Jerolmack & Murphy, 2019; Reyes, 2018).

After gathering the first set of data, the team leapt straight into testing ATLAS.ti with everyone selecting one interview to code, applying any labels that emerged when reading. This process of inductive coding helped us to draft a list of initial codes (Chandra & Shang, 2019; Corbin & Strauss, 2008). Whilst much qualitative research adopts a deductive top-down approach by defining concepts first and coding second, we combined top-down and bottom-up approaches oriented towards grounded theory (Corbin & Strauss 2008; Creswell, 2013). A bottom-up approach means developing concepts and their dimensions inductively whilst coding. However, we contend it is not possible to analyse data without having pre-existing theoretical foundations in mind, as we are inevitably cognitive of (and arguably influenced by) them by

<sup>1</sup> See for instance <https://atlasti.com/video-tutorials>.

virtue of our knowledge of previous research. Thus, our initial list of inductive codes was completed by codes that were derived from the previous knowledge we gathered via reading groups and discussions of the literature as mentioned in Chapter 2.

The pilot study—or the first phase of data gathering (see Chapter 3)—offered a premium opportunity to develop and test a functioning collaborative coding process. Several focused team meetings served as the loci to collectively discuss, agree and develop a collaborative coding strategy. In these meetings, we agreed on broad definitions for each code and determined inclusive as well as exclusive criteria to explicitly clarify the situations in which each code would apply, using examples from our ‘trial coding interviews’. This helped to ensure that the meaning of codes was understood by all, even those coders who joined the team later to promote intercoder reliability. Such meetings were crucial in taking important decisions that would enable us to code the data systematically, even though the way we split the workload meant that not everyone read nor coded all the data. For example, this included agreements on coding large chunks of text, and on the inclusion of interview questions so that each quotation would be in context and remain intelligible to those who did not code it.

Based on this first practical exercise, we compared codes and identified similarities, but also revealed several differences in the way we used codes. Here again the research diary and notes proved invaluable in keeping track of our observations, and often our doubts, as we often left questions to each other in the memos. The team member that chaired our meetings collected questions from all the diary notes and we addressed them together. The exercise was designed to improve intercoder reliability by discussing in detail our different understandings. We did not calculate intercoder reliability scores, but rather followed a more inclusive and collaborative approach for developing a code list, by defining the codes and determining how to use them. In other words, we refined and debated our choices collaboratively, so that understanding was consistent as it could be across all team members (Reyes et al., 2021). We explored recurring contradictions and solved them as part of subsequent brainstorming sessions, in which we added further codes inductively and deductively after the first rounds of coding.

The first code list was further refined and extended in comprehensive team meetings where two practical activities were undertaken to inspire code development. First, team members coded five interviews from their

own ATLAS.ti project and were encouraged to develop new codes; this helped everyone to learn the technicalities of coding and to flag difficulties in doing so. Secondly, to counter the risk of early software-coding routines leading to narrow coding practice by using only certain codes, or using keyword searches instead of reading carefully, each team member hand-coded several interviews. Hand-coding is a technique that whilst considerably slowing down the coding process—it consists of flipping through printed material instead of scrolling on a screen—it tends to produce new codes.

We debated code names and definitions and jointly decided to ensure everyone used them correctly and systematically. This was one of the most important steps in the team coding process and re-occurred as a relevant practice throughout the coding stages. Clear definitions increase inter-coder reliability, ensure consistency in coding, train potential incoming members and make important interview segments visible to others. They must also specify what to include or exclude and when to use or not use the code. For instance, the code ‘Sexism’ was defined as instances where the interviewees ‘describe sexism, sexist experiences, language; practices discriminating directly’, whereas the code ‘Gendered practices’ was defined as ‘all instances where genders are treated differently; speaking time, divisions of posts, vertical stuff; *not sexism*’. Furthermore, three additional subcodes completed the code ‘Gendered practices’: ‘Gendered practices\_discrimination’, ‘Gendered practices\_division of labour’ and ‘Gendered practices\_hierarchies’. Each of these had a specific definition that differentiated them. For example, we defined ‘Gendered practices\_discrimination’ as instances where, ‘the word is used, also including mentions of bias, indirect discrimination etc.’; ‘Gendered practices\_division of labour’ as highlighting the ‘separation of women’s and men’s policy areas’ in interviews; and ‘Gendered practices\_hierarchies’ as all instances where the interviewees mentioned ‘women having difficulties getting reports, leadership positions etc. vertical segregation’.

Without clear definitions, such a collaborative, supportive and inclusive team coding process would not have been possible and would have obscured key elements of the data. Trusting each other to signpost important and relevant topics within the dense data was key to the success of the research project and considerably speeded up the process. Simply put, the first stage of coding allowed us to categorise the raw data under important and jointly developed topics, which then helped individual researchers to investigate them further. Without this, the screening and coding of

all the dense data for each individual study would have been far too time-consuming and overwhelming.

In our view, codes and definitions should not be ‘set in stone’ because the coding process requires their constant adjustment and extension. We followed this practice whilst the project was in full flow, with the consequence that we re-coded interviews depending on what revisions were implied. For example, if a code was split into two, we also re-coded the respective code segments and ensured that the new codes were applied to other quotations where appropriate. Likewise, when we merged codes, we re-coded the relevant data.

Our iterative approach to coding meant that we could complement our initial list of codes with new codes that emerged at later stages of the research process. We complemented and informed our list of codes deductively with ideas from literature, documentary research, and on the basis of pre-selected keywords relevant to the main objectives of the project. These included, for instance, ‘democratic practices’, ‘political groups’, ‘economy’, ‘gender-based violence’, ‘affects’ or ‘social policy’. Whilst inductively, we supplemented the list of codes with emerging themes, such as ‘political group meetings’, ‘resistance to gender equality’ or ‘sexual harassment’. In total, the first brainstorming sessions resulted in a list of 99 codes, including 55 main codes and 44 subcodes; where subcodes were the code families for main codes.

Codes that were added during the process of reading and re-coding of the material, were the starting point of many of our published findings; they were not initially planned, but made possible because we diligently travelled back-and-forth between coding and new analysis of the data. For instance, we extracted unexpected insights from our data on the power dynamics of national party delegations in the European Parliament, because we added the code ‘National party delegations’ (Elomäki et al., 2023), on normative whiteness and racism consequent upon the added code ‘intersectionality\_race’ (Kantola et al., 2023) and the role of gendered religious claims after the addition of the code ‘religion’ (Ahrens et al., 2022).



## MANAGING THE TECHNICALITIES OF COLLABORATIVE CODING

In addition to the intellectual work of designing a code strategy that was applicable to all researchers, team coding required finding solutions for technical and organisational issues intrinsic to the project (see Box 5.4). The shifting geographical locations of team members had an additional impact on the process, and not least, the Covid-19 pandemic had a profound impact on planned in-person meetings which became impossible for an extended period.

The team being split between places became problematic when we worked on a joint ATLAS.ti ‘project’—as recommended by ATLAS.ti. We initially planned to place the so-called ‘copy bundle’—that is the exported ATLAS.ti ‘project’, including all coded and raw data, list of codes, information on codes and memos—in our joint drive at Tampere University. However, this turned out to be impossible as running ATLAS.ti via VPN on personal laptops often failed. Instead, for the pilot study, we decided to follow the second option recommended by ATLAS.ti, whereby every team member sets up their own ATLAS.ti ‘project’ with the interviews assigned to them, and then one team member would merge all the projects. Nevertheless, with the high number of team members coding simultaneously this proved to be impractical. Whilst the coded interviews were rather unproblematic when compiled, shared or revised, memos were not easily merged and had to be put together manually—a step which would have been far too time-consuming.

As a result, we agreed on a different sharing process for the main study. One team member was ‘in charge’ of the copy bundle, of which they supervised, managed and controlled the whole coding process. That person assigned interviews to each team member who was then responsible for coding. This meant that each team member coded certain interviews, including the ones not available in English but in their mother tongue. We established a rotation system with one coder coding at a time in the same ATLAS.ti ‘project’, before exporting it as a copy bundle and sending it via university email to the next coder in line. The rotation system respected a clear order of names (e.g., coder A before coder B; and coder E after coder D) with each coder knowing who coded before them and who would receive the copy bundle after them.

Similarly, the person in charge assigned interviews on a rolling basis, typically two interviews per round, and coders knew approximately which

day of the week the copy bundle would come to them. They could thus reserve time for their allocated day to code the two interviews and send it to the next coder in time. We built in a degree of flexibility to allow for the possibility that days might need to be switched with another coder, but this needed to be clear to the whole team so that the entire process kept rolling smoothly, and did not affect the allocation of interviews which remained the same.

To ensure that nothing got lost in the process of exchanging copy bundles, we created one single email thread for sending them and all the information related to coding. Whenever someone finished their coding, they would send the new copy bundle in this thread along with important features of the interviews that needed to be flagged up. Every Thursday, the copy bundle would return to the team member in charge, who would add new interview data as they came back from transcription and allocate them to coders for the upcoming coding weeks. This ‘rolling coding’ strategy meant that with six coders, each coder had a coding-free day every other week due to the maximum of four coding days per week. During the project, we sometimes had to code with fewer people due to long fieldwork periods, illness, or care responsibilities during lockdowns.

As well as coding, the team member in charge was responsible for all other technical issues. Each Friday, they checked the latest copy bundle, saved it and resolved errors, a task made easier by the research diary—or ‘memo’—as we kept all the entries made by the coders in one place. In fact, each coder had to report on the research diary after each coding session. The research diary turned out to be a central element to our coding strategy, as it made the collective coding process transparent to other coders, kept track of ideas or thoughts whilst coding, highlighted errors or inconsistencies, and significantly, stressed any doubts that needed to be discussed in upcoming meetings. All the minutes from coding meetings were also stored as a memo directly on the ATLAS.ti ‘project’, as a means of increasing access and transparency during coding. In fact, these very lines are written on the basis of the notes we kept in the research diary throughout the coding process. Each coder recorded aspects like new ideas for codes, problems with coding or specific interviews, questions to discuss in meetings and also comments on funny quotes, oddities or levity in the interviews (see Boxes 5.2 and 5.3). After the basic check-up, the person in charge uploaded new anonymised transcripts, allocated them to coders, and documented everything in the research diary.

**Box 5.2 Examples from Our Research Diary of Exchange of Thoughts Whilst Coding**

- ‘I coded the interview ‘Renew MEP M 081,119 Brexit‘. I understand that it was particularly meant for the Brexit paper, but was still a bit surprised that other main questions from our interview list were not addressed. Would have been interesting to gather more information on gender aspects, too.’ (EUGenDem research diary 27 Jan 2020)
- ‘What struck me most when comparing these two interviews is the stark contrast of how the two describe their start and how gendered it is: the female says it is hard to get into positions because there are of course many returning MEPs who can choose first and she’ll have to wait; the male says he was surprised how easy it was to get the position he wants and how many requests he got.’ (EUGenDem research diary 22 Apr 2020)
- ‘I feel that in Zoom meetings it is hard to build rapport and it is easier for participants to say: ‘I’ve got to be somewhere in half an hour’ because you have spent less effort going to meet them in person in the parliament.’ (EUGenDem research diary 13 Nov 2020)
- ‘Since coding the notes, I would really see the need for the ‘power relations’ code—political influence just doesn’t cover what is in the notes.’ (EUGenDem research diary 25 Feb 2021)
- ‘I like that when coding ethnographic notes I get to use the codes that I felt I was often underusing when coding the interviews: ‘embodiment’, ‘EP spaces’, ‘researcher role’, etc.’ (EUGenDem research diary 13 Apr 2021)
- ‘A usual issue came up with codes, which is whether to include a code when it is referred to in the negative e.g. the EP as a unique parliament. I coded the section as this, even though the respondent says that it isn’t a unique parliament.’ (EUGenDem research diary 28 Jan 2020)

### Box 5.3 Moments of Levity in Our Interviews that We Highlighted to Each Other in the Research Diary

- ‘I would also say that men are often the ones who have been in parliament for a longer time. So, those [persons/men] who have divided things here already for three or four terms; the others are still searching for the toilets, and while they still don’t know where the toilets are in the house, the men have the jobs already divided’ (Female S&D MEP March 2020)
- During an interview in the midst of the Covid-19 restrictions: ‘Thank you. I think I’m a little... I haven’t been speaking to anyone this week and so maybe that’s why I don’t find the words.’ (Female Left assistant March 2020)
- ‘X comes back to the office. X says ‘I must go, I am late for my life’ and their assistant notes that this is a pithy saying and outlook.’ (Ethnographic field note shadowing a female EPP MEP November 2018)
- ‘R: Talking about, you had some extraordinary word in there I had never heard of before, ethno-something-or...

I: Ethnography...

R: Never heard of it.’ (Male ENF MEP February 2019)

- ‘We don’t believe in the European Union, so we’re just here because we want to destroy it.’ (Male EFDD MEP January 2019)
- ‘They say politics is rock music for ugly people.’ (Male ENF MEP February 2019)

We put each newly uploaded interview transcript into document groups, covering categories like the political group, female/male, MP/Staff/ and nationality. As will be explained below, this simplified both code outputs and the analysis. Then, if applicable, the team member in charge would create, revise or merge<sup>2</sup> codes as agreed in team coding meetings. When

<sup>2</sup> Merging codes means combining two or more previous codes into one new code, for instance, when previous codes were too detailed and resulted in very low numbers of coded material that could be equally well represented by one overarching code.

all these steps were completed, the new copy bundle was sent in the email thread with an overview of who was next in the rotation system (with dates), and the process would begin again. Finally, our person in charge collected questions, code proposals and any other business to be presented and discussed during the next coding meeting.

Once we finished coding most of the interview data, we moved on to the ethnographic data using the list of codes we had already developed. The ethnographic notes were an excellent way to contextualise interview data and elicit new perspectives. In the interest of simplicity and clarity, we used the existing list of codes rather than developing new ones, which was justified by the already extensive (and saturated) list of codes (in total 112 codes) we had already generated (see Box 5.5).

Overall, the strategy of collective ‘rolling coding’ ensured that the data was very quickly available after transcription for further analysis and interpretation. Furthermore, such a closely intertwined and intrinsically collective process kept all coders in the loop, encouraged constant cross-comparison between coders, and resulted in fruitful in-depth discussions of potential research questions that emerged from the data (see Box 5.4). The flexibility we built into a process that utilised so many coders, allowed us to be reflexive in response to people’s changing circumstances and unforeseen complications. Nonetheless, our periodic coding meetings revealed only minimal dissimilarities in the ways in which coders understood some codes, which was a testament to the constant and transparent communication required amongst all coders. Given the differences in research foci, and the very high number of codes, coders did not always, systematically attend to all of the codes, which at times left parts of the data less visible. To address this issue, we organised additional rounds of coding where all coders rechecked their assigned interviews for occurrences of specific codes, and invested additional time on an ongoing basis to assess where the coding process stands and to engage in discussions.

**Box 5.4 Core Points for Successful Collective Coding**

- Develop the code list with code names and definitions collectively (mention situational inclusion and exclusion, if necessary).
- Allow for constant adjustment of codes, their definitions, adding new ones and for splitting/merging existing ones.
- Trust your choice of codes, whether they are deductively or inductively developed, and encourage using codes as often as possible.
- Ensure communication transparency has the highest priority: establish a research diary for the whole team to collect notes and ideas on the process for extending and adjusting codes; share all information related to coding in one email thread or drive folder.
- Appoint one person in charge: a team member, researcher, and coder that supervises and manages the collective coding process. This person's tasks need to include collecting remarks from the research diary and raising them for discussion in team coding meetings.
- Make a clear plan including responsibilities for everyone but allow for flexibility and be prepared for interruptions.
- Try to keep the coding and discussion process continuously rolling to make the most of memorising content and technicalities.

## ORGANISING AND SORTING THE CODED INTERVIEW AND ETHNOGRAPHIC DATA

Once the raw interview and ethnographic notes were coded, we applied two main sorting mechanisms to make sense of our coded data. We used 'code groups' and 'document groups'—functions defined by ATLAS.ti. 'Code groups' offered the opportunity to select specific topics, such as specific policy fields, parliamentary bodies, actors, relationships or affects, whereas 'document groups' categorised interviews along specific descriptive categories, such as male/female or MEPs/staff and per political

groups or nationalities. This allowed us to work with coded data quickly for such descriptive groups, making our dense data more manageable and analyzable, by extracting quotations that intersect with one code and one political group (eg., ‘sexual harassment’ and ‘EPP’).

In total, our ATLAS.ti project included 112 different codes, made up of 69 main codes and 43 subcodes grouped into 16 code families, which helped bridge codes that complemented each other on similar themes (see Table 5.1 for examples). Other functions of ATLAS.ti, such as ‘output tables’ or ‘reports of co-occurring codes’ and ‘reports of neighbouring codes’ provided an easy way to extract data for analysis.

### Box 5.5 Different Types of Codes

The codes comprised different kinds with the following constituting the main aspects:

- Codes relating to process and sequencing: when and how political groups were formed? how specific policy proposals moved through different stages? (e.g., ‘political group formation’, ‘democratic practices’, ‘EP elections 2019’; ‘political group internal policy formation’, ‘political influence’);
- Codes related to policy fields (e.g., economic policy, gender-based violence and social policy);
- Codes on specific topics (e.g., ‘leadership’, ‘civil society’, ‘opposition to gender equality’, ‘gender mainstreaming’, ‘reproductive rights’, ‘Covid-19’ and ‘Brexit’);
- Codes for internal communication, either ethical aspects or reminders to ourselves for further steps, such as new names for interviews or requests for the confidentiality of single comments (e.g., ‘researcher role’, ‘to follow-up’, ‘confidential text in interview’);
- Codes on relationships between actors inside and outside the European Parliament (e.g. ‘Europarties’, ‘political groups about other political groups’, ‘MEPs vis-a-vis political groups’, ‘negotiations and compromise between political groups’ and ‘interinstitutional relationships’);

- Codes on certain bodies and functions (e.g., ‘political groups identity’, ‘political groups organisation’, ‘National party delegations’, ‘EP administration’, ‘Secretary General’, ‘rapporteurs’, ‘coordinators’);
- Codes on internal practices (e.g., ‘political groups as workplace’, ‘political groups conflicts\_internal’, ‘MEPs daily work’);
- Other relevant codes (e.g., ‘racism’, ‘Spitzenkandidatur’, ‘religion’, ‘sexism’, ‘feminism’, ‘populism’, ‘Euroscepticism’).

By creating ‘document groups’, researchers were able to extract all data relevant to their research questions at once. For instance, if analysing the gendered aspects of leadership in political groups, researchers could quickly extract the relevant data by exporting quotes that intersect with the document groups ‘Greens/EFA political group’ and ‘female MEPs’, and with the code group ‘leadership’. Working with such combinations extracts data in ways that make further analysis manageable by restricting the searched volume of data to the most relevant part. This was particularly helpful for codes that we applied often. The code ‘National party delegations’, for example, generated roughly 500 quotations or over 200 pages of coded data. Intersecting that code with other ‘document groups’ or ‘code groups’ helped simplify the process of data analysis. Thus, when combining the code ‘National party delegations’ with the document group of all political groups, we generated 42 pages of coded data—making the data analysis considerably more manageable.

ATLAS.ti includes various tools that help to organise, sort and make sense of the coded data in view of interpreting it. Some tools allowed for the tracking of connected codes, concepts and theoretical thoughts that emerged whilst coding. For instance, coders were able to link codes and quotations with relations such as ‘contradicts’, ‘is associated with’ or ‘is part of’. ATLAS.ti easily allowed the application, merging, or splitting of codes, and writing and attaching memos to any part of the data deemed relevant. More advanced features included sorting the coded data into networks of codes, which helped to quickly visualise the co-occurrences of codes and the relations between quotations. As a result, and by playing with and visualising the coded data differently, researchers can become more familiar with its material and develop a ‘professional vision’ towards it (Goodwin, 1994, in Elliott, 2018).



Table 5.1 Examples of code families

<i>Code family</i>	<i>Main code</i>	<i>Subcodes</i>	<i>Definition</i>	<i>Occurrences</i>
Affects	Affects		Descriptions of emotions and affects, strong language and figures in the interviews, expressions of emotions in the interviews	349
		Affective atmospheres	More collective, more mood, e.g. in locale/e.g. group meeting 'tense' or 'business like'	211
Committee	Committee	Expression of emotions	Cognitive, more isolated/interviewed expresses an emotion or uses emotional language	253
			Including: all committees, interactions and division of competences between them, etc.	318
		ECON	Mentions of ECON	177
		EMPL	Mentions of EMPL	89
		FEMM LIBE	Mentions of FEMM Mentions of LIBE	178 89

(continued)

Table 5.1 (continued)

<i>Code family</i>	<i>Main code</i>	<i>Subcodes</i>	<i>Definition</i>	<i>Occurrences</i>
Gender equality	Gender equality		How gender equality is referred to and used by the interviewee	292
		Policies	References to different gender quality policies, such as equal pay, reconciling work and family, violence, gender mainstreaming	162
		Practices	Practices to promote gender equality, quotas, harassment policies, women's networks, training, and groups. Not FEMM	75
		EP	European Parliament specific practices for promoting gender equality	73
		PG	Political groups specific practices for promoting gender equality	186
		Descriptive representation	Gender balance and mentions thereof	

<i>Code family</i>	<i>Main code</i>	<i>Subcodes</i>	<i>Definition</i>	<i>Occurrences</i>
PG organisation	PG organisation		Description of bodies, structures, functions, funding, budgets	223
		Group meeting	Mentions of Group meeting	219
		Press office	Activities of PG press office	74
		Statues	Mentions of Statues	40
Sexism	Sexism		They describe sexism, sexist experiences, language; practices discriminating directly	74
		Good expertise	Stereotyped expertise (masculine/feminine)	45
		Sexiest language	Description of interviewees when sexiest language was used	20

In sum, most of ATLAS.ti's tools and functions help to zoom in on specific narratives, rhetoric and frames under one or more code(s), allowing a closer reading of our data in relation to our research questions.

## CODING DOCUMENTS FOR RESEARCH ARTICLE

Whilst all the researchers coded our interview and ethnographic data, document data was coded separately for each individual research article. Nonetheless, this followed a similar pattern of organising dense data into manageable chunks, and preparing the text for interpretation. In this respect, we followed the approach to coding that envisaged an evolving, rather than fixed strategy, to be used throughout the project-related publications (Elliott, 2018). In that sense, the coding strategies developed for research articles largely depended on their research questions.

As Chapter 3 showed, we collected a wealth of internal documents from the EP including practice (e.g., rules of procedure and codes of conduct) and policy-related documents (e.g., reports, amendments, position papers and press releases). We gathered such documents on a case-by-case basis, dependent on the research design and research question germane to a specific research article. As a result, we coded documents according to frameworks developed by the researcher(s) in charge of the article: either single or group coding approaches (see Box 5.6). On occasion, this framework was used to code both the document data, and to re-code chunks of coded interview and ethnographic data. In these instances, the interview and ethnographic data were coded in a first stage of collective coding, as explained above, and then re-coded along with additional material such as documents with a coding strategy (i.e., a new code list, code definition, etc....) developed for the research article. For example, in one research article we focused on the policy-related issue of the ratification of the Istanbul Convention on violence against women and domestic violence by the EU. At the first stage of collective coding, the code 'Istanbul Convention' was applied to any mention of it in the interview and ethnographic data. At the second stage of coding for individual study, the research design and research question required the expansion of research material to include specific documents, such as transcripts of debates, and the re-coding of the pre-coded data under 'Istanbul Convention' in a separate ATLAS.ti 'project' with a new list of tailored codes (Berthet, 2022a). Both the document data and the pre-coded interview

and ethnographic data were re-coded with the same code list developed at the second stage of coding to ensure a systematic process.

Coding documents is demanding. Written records of amendments and debates, for example, can be lengthy, consisting of large amounts of text that need to be closely analysed. By way of illustration, we analysed over 1090 amendments for the pay transparency draft directive and 750 for the work-life balance draft directive (Copeland et al., 2023). The more salient the topic, the greater the amount of amendments there were to analyse. For instance, the non-legislative draft report on sexual and reproductive health and rights in the EU, which included provisions on abortion rights, generated over 500 amendments at the committee level (Berthet, 2022b). Similarly, a plenary debate on a salient topic could include over 500 oral and written interventions. In this sense, coding document data was useful for reducing the data into ‘manageable proportions’ (Coffey & Atkinson, 1996, 28).

A recurrent form of document data was amendments made by MEPs and political groups to committee reports. Whilst these are important when analysing policy processes, because they allow for the identification of group positions that are taken, we adopted different approaches to coding amendments for different research articles. When a study was attentive to different discursive constructions, we selected and coded those amendments that were relevant to the discursive analysis of one specific issue (e.g., abortion rights). Equally, when a study was interested in identifying group positions and comparing them in a quantifiable way, we coded all amendments made to a specific report (e.g., how often groups weakened the proposals and which groups) (Copeland et al., 2023).

#### **Box 5.6 Coding Document Data for Research Articles**

We coded documents based on our interest in specific policy issues, for instance, abortion rights (Berthet, 2022b), economic policy (Elomäki, 2021), economic and social rights (Elomäki & Gaweda, 2022), austerity politics (Elomäki, forthcoming) and strategies of opposition to gender equality (Berthet, 2022a; Kantola &

Lombardo, 2021a, 2021b). We also coded documents to investigate the influence of groups on Commission Directive proposals (Copeland et al., 2023) and the power dynamics and modes of decision-making in groups (Elomäki et al., 2022). In these cases, coding had an analytical function; it helped with ‘(a) noticing relevant phenomena, (b) collecting examples of those phenomena, and (c) analysing those phenomena in order to find commonalities, differences, patterns, and structures’ (Seidel & Kelle, 1995, 55–56).

Because we coded documents based on the research question(s) specifically developed for each research article (see Box 5.7), the coding lists were developed deductively and inductively according to the specific theoretical and epistemological approaches taken (Coffey & Atkinson, 1996, 32). In some cases, the development of codes took place in multiple steps, starting from the descriptive level and moving towards analytical typologies. This confirms that our coding was dynamic, and influenced by theory-driven interpretation(s), as well as analytical.

#### **Box 5.7 Example of a Coding Strategy for a Research Article**

In one research article on economic ideas about austerity and its alternatives in the European Parliament, (Elomäki, forthcoming), amendments and plenary interventions related to ten of the EP’s own initiative reports on EU economic governance, were initially coded through a code list that involved four categories: (i) approach to austerity (opposing/supporting), (ii) rationales for supporting austerity, (iii) rationales for opposing austerity and (iv) alternatives to austerity. The categories ii–iv involved several options each, deducted from a combination of existing scholarship and matters that emerged from the data. In the analysis and writing process, the emphasis moved from rationales to paradigms. The final coding of the data consisted of classifying the amendments and plenary interventions into three main pro-austerity paradigms and three main paradigms providing alternatives.

Some of the ATLAS.Ti tools, for instance, the co-occurrence and query functions, eventually helped us identify patterns of meanings (Bazeley, 2009a). In this respect, during the adoption of a report, we were able to observe how some political groups in the European Parliament discussed public services more often as a cost or as an investment. Likewise, we could assess if and how they discussed gender equality via economic rationales or as a value in itself.

Although our analysis was always qualitative, some of us used the possibilities provided by ATLAS.ti to generate quantified comparisons (Bazeley, 2009b). Although quantifying was for us often a preliminary step, with our emphasis being discursive-interpretive analysis, journal reviewers often asked specifically for quantified data—which we were happy and able to provide. Our process meant that we could compare the distribution of amendments from political groups that either strengthened, weakened, or verified/clarified a specific draft directive, allowing us to see at a glance where political groups stood relative to each other, or how the patterns of strengthening and weakening differed by directive. Quantification was also useful to understand patterns of change over time in those cases where longitudinal analysis of recurring EP reports was conducted. Through the code/document function of ATLAS.ti, we could observe shifts in the positions adopted by EP as well as political groups—for instance, a shift from austerity to investment in at least some EPP MEPs discourse, or how the initial acceptance of austerity by some S&D MEPs in the early 2010s turned into an outright rejection (Elomäki, forthcoming).

Since our research objectives concentrated mostly on the lines of convergence and conflict between the groups, it was important to identify them correctly. Consequently, when documents covering amendments and debates were coded, for example, we paid particular attention to the political affiliation of the speakers. Ensuring the integrity of our identification often required extra work, which was particularly the case with amendments, as they were not always attributable to a particular group. It also became important to note the nationalities of MEPs since fault lines in groups and cross-group alliances tend to form on the basis of shared nationalities. Such codes are descriptive but necessary for later analysis (Elliott, 2018; see Box 5.5).

Similar to the challenges of coding interview data collectively, when problems relating to different interpretations of codes and content emerged during co-authorship, we strove for intercoder reliability via

discussion and interaction during the coding process. This included several discussions about our coding framework and definitions, testing the coding framework before starting the actual coding work and comparing it. For example, co-authors exchanged parts of the coded data to see if others would have coded it differently, and any emergent differences were addressed and settled in our resolve to ensure transparency in coding-related decision-making.

## CONCLUSION

Like any other aspect of the research process, data coding using analytical software provides benefits and challenges. As we have demonstrated, coding is one way of organising dense data and of making sense of it by identifying overarching patterns. Whilst various approaches to coding exist, this chapter was attentive to the strategies we implemented to code dense interview, ethnographic and document data collaboratively. We have provided important tips and concrete examples of using software tools for qualitative analysis, such as ATLAS.ti, and the intricacies of using it as a team. Specifically, we addressed the initial stages of developing code lists in inductive and deductive ways, the technicalities intrinsic to coding the text of the data with ATLAS.ti, and presented an overview of how we took advantage of some tools, such as creating code families, to make sense of the data. Our main focal point throughout the chapter was to highlight the collaborative nature of our coding work by reviewing the pros and cons, and by examining how issues of intercoder reliability were resolved. Our approach to data analysis was firmly rooted in collaborative work and provided the basis for all further interpretative analysis through individual or co-authored research articles, as well as collaboration with external scholars. Having a more nuanced understanding of the processes we followed, and the techniques we established throughout the coding process, provides a firm foundation to better understand our interpretation of the results, which the discussion of we turn to next.

## REFERENCES

- Ahrens, P. (2018). Qualitative network analysis: A useful tool for investigating policy networks in transnational settings? *Methodological Innovations*, 11(1), 1–9.



- Ahrens, P., Gaweda, B., & Kantola, J. (2022). Reframing the language of human rights? Political group contestations on women's and LGBTQI rights in European Parliament debates. *Journal of European Integration*, 44(6), 803–819.
- Bazeley, P. (2009a). Analysing qualitative data: More than 'identifying themes.' *Malaysian Journal of Qualitative Research*, 2(2), 6–22.
- Bazeley, P. (2009b). Integrating data analyses in mixed methods research. *Journal of Mixed Methods Research*, 3(3), 203–207.
- Berthet, V. (2022a). Norm under fire: Support for and opposition to the European Union's ratification of the Istanbul Convention in the European Parliament. *International Feminist Journal of Politics*, 24(5), 675–698.
- Berthet, V. (2022b). United in Crisis: Abortion Politics in the European Parliament and Political Groups' Disputes over EU Values. *Journal of Common Market Studies*, 60(6), 1797–1814.
- Chandra, Y., & Shang, L. (2019). Inductive coding. In Y. Chandra & L. Shang (Eds.), *Qualitative research using R: A systematic approach* (pp. 91–106). Springer Singapore.
- Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: Complementary research strategies*. Sage Publications.
- Corbin, J., & Strauss, A. L. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage.
- Contreras, R. (2019). The broken ethnography: Lessons from an almost hero. *Qualitative Sociology*, 42, 161–179.
- Copeland, P., Elomäki, A., & Gaweda, B. (2023). Filtering politicization towards a more social Europe? The European Parliament's role in EU legislation. Paper presented at the 29th International Conference of Europeanists, 27–29 June 2023, University of Reykjavik, Iceland.
- Creswell, J. (2013). *Qualitative inquiry and research design*. Sage.
- Deterding, N. M., & Waters, M. C. (2021). Flexible coding of in-depth interviews: A twenty-first-century approach. *Sociological Methods & Research*, 50(2), 708–739.
- Elliott, V. (2018). Thinking about the coding process in qualitative data analysis. *The Qualitative Report*, 23(11), 2850–2861.
- Elomäki, A. (2021). "It's a total no-no": The strategic silence about gender in the European Parliament's economic governance policies. *International Political Science Review*. <https://doi.org/10.1177/019251212097832>
- Elomäki, A. (Forthcoming). Austerity and its alternatives in the European Parliament: From the Eurozone crisis to the Covid-19. *Comparative European Politics*.
- Elomäki, A., & Gaweda, B. (2022). Looking for the 'Social' in the European Semester: The ambiguous 'Socialisation' of EU Economic Governance in the

- European Parliament. *Journal of Contemporary European Research*, 18(1), 166–183.
- Elomäki, A., Gaweda, B., & Berthet, V. (2022). Democratic Practices and Political Dynamics of Intra-Group Policy Formation in the European Parliament. In P. Ahrens, A. Elomäki, & J. Kantola (Eds.), *European Parliament's Political Groups in Turbulent Times* (pp. 73–96). Springer International Publishing.
- Elomäki, A., & Kantola, J. (2022). Feminist governance in the European Parliament: The political struggle over the inclusion of gender in the EU's COVID-19 response. *Politics & Gender*, 1-22, <https://doi.org/10.1017/S1743923X21000544>
- Elomäki, A., Kantola, J., Ahrens, P., Berthet, V., Gaweda, B., & Miller, C. (2023). The politics of national delegations in the European Parliament. Paper presented at the Center for European Studies (CES) Seminar on Researching the European Union, 1 February 2023. University of Helsinki, Finland.
- Friese, S. (2012). *Qualitative Data Analysis with ATLAS.ti*. Sage.
- Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96(3), 606–633.
- Guenther, K. M. (2009). The politics of names: Rethinking the methodological and ethical significance of naming people, organizations, and places. *Qualitative Research*, 9(4), 411–421.
- Jerolmack, C., & Murphy, A. K. (2019). The ethical dilemmas and social scientific trade-offs of masking in ethnography. *Sociological Methods & Research*, 48(4), 801–827.
- Kantola, J., Elomäki, A., Gaweda, B., Miller, C., Ahrens, P., & Berthet, V. (2023). “It’s Like Shouting to a Brick Wall”: Normative Whiteness and Racism in the European Parliament. *American Political Science Review*, 117(1), 184–199.
- Kantola, J., & Miller, C. (2023). Eternal Friends or Jubilant Brexiteers? The Impact of Brexit on UK MEPs’ Parliamentary Work in the European Parliament. *Journal of Common Market Studies*, 61, 712–729.
- Kantola, J., & Lombardo, E. (2021a). Strategies of right populists in opposing gender equality in a polarized European Parliament. *International Political Science Review*, 42(5), 565–579.
- Kantola, J., & Lombardo, E. (2021b). Challenges to democratic practices and discourses in the European Parliament: Feminist perspectives on the politics of political groups. *Social Politics*, 28(3), 579–602.
- Paulus, T. M., & Lester, J. N. (2016). ATLAS.ti for conversation and discourse analysis studies. *International Journal of Social Research Methodology*, 19(4), 405–428.
- Reyes, V., Bogumil, E., & Welch, L. E. (2021). The living codebook: Documenting the process of qualitative data analysis. *Sociological Methods & Research*. <https://doi.org/10.1177/0049124120986185>

- Reyes, V. (2018). Three models of transparency in ethnographic research: Naming places, naming people, and sharing data. *Ethnography, 19*(2), 204–226.
- Saldaña, J. (2021). *The coding manual for qualitative researchers*. Sage Publications.
- Seidel, J., & Kelle, U. (1995). Different functions of coding in the analysis of textual data. In U. Kelle (Ed.), *Computer-aided qualitative data analysis: Theory, methods and practice* (pp. 52–61). Sage.
- Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge University Press.
- Tesch, R. (1990). *Qualitative research : analysis types and software tools*. Falmer Press.
- Weston, C., Gandell, T., Beauchamp, J., McAlpine, L., Wiseman, C., & Beauchamp, C. (2001). Analyzing interview data: The development and evolution of a coding system. *Qualitative Sociology, 24*, 381–400.
- Yin, R. (2011). *Qualitative Research from Start to Finish*. The Guilford Press.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

