


Coding Metaphors in Interaction: A Study Protocol and Reflection on Validity and Reliability Challenges

International Journal of Qualitative Methods
Volume 22: 1–9
© The Author(s) 2023
DOI: 10.1177/16094069231164608
journals.sagepub.com/home/ijq


Jana Declercq¹ , and Lotte van Poppel¹ 

Abstract

With this study protocol, we present our approach to collaboratively coding metaphors in medical consultations, using the qualitative analysis software package ATLAS.ti. This project came with a number of challenges but also yielded a coded data set that was rich and versatile, and allowed for a wide array of analyses. We therefore believe it can be of interest to share the details of the process, to (1) provide an hands-on approach to qualitative, collaborative (metaphor) coding, and (2) contribute to our understanding and the advancement of qualitative methods in relation to coding metaphors, and coding more generally. The protocol can be of interest for researchers interested in collaborative qualitative coding, of metaphors or other (linguistic) units, using software. We specifically describe how we set up our initial coding system, and how we further developed it inductively; how we dealt with the fuzziness of metaphors; how we approach cleaning between coding rounds; how we divided work over and collaborated as a team of three coders with different but related cultural and linguistic backgrounds; how we structured our documentation and logbook. We then discuss how this set of guidelines complements the theoretical literature and more general guidelines on (metaphor) coding, and on reliability and validity in qualitative research. We will reflect on both a number of logistic and practical issues, as well as methodological and empirical challenges, and advantages and disadvantages of our approach.

Keywords

Discourse analysis, ethnography, methods in qualitative inquiry, grounded theory, constructivist GT

Introduction

This study protocol details how a team of three researchers has collaboratively coded metaphors in health professional-patient interactions, and provides reflections on this process in relation to qualitative methods using coding approaches more generally. By providing this protocol, we aim to present a detailed, practical hands-on tool for the analysis of metaphor, which is likely also (partly) suitable for the coding of other (linguistic) units, as well as to further develop our understanding and advancement of (challenges in) qualitative methods more generally.

Metaphor analysis provides a good case for developing such hands-on tools and for such reflection, for several reasons. First, metaphors are a much-studied phenomenon in the humanities and social sciences, as they can give us insight into how people reason, construct experiences and (social)

phenomena around them (Armstrong et al., 2011; Gibbs & Franks, 2009; Semino et al., 2017). They are a research topic in a range of disciplines, often studied using qualitative approaches (although computer-assisted corpus approaches are possible too – e.g., see Demmen et al. (2015)). Second, these are usually based on some form of coding or labelling. However, consistently coding metaphors is challenging, as it requires interpretation, which always carries a degree of subjectivity (Armstrong et al., 2011; Cornelissen et al., 2008). Interpreting metaphor in conversations is particularly

¹University of Groningen, Groningen, The Netherlands

Corresponding Author:

Jana Declercq, University of Groningen, Oude Kijk in't Jatstraat 26, Groningen 9712EK, The Netherlands.
Email: j.h.m.declercq@rug.nl



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

complex, due to ellipsis, overlap, repair and other particularities of oral interaction. This complexity plays an even bigger role in research focusing on the development of metaphor use throughout the conversation, such as co-creation and resistance to metaphor.

Third, different traditions have different conceptualisations of what counts as a metaphor, and what should be included in analysis (Steen, 2016), and also differ in terms of the level of detail regarding practical/empirical implementation when identifying and analysing metaphors. For instance, in the tradition of Conceptual Metaphor Theory (Lakoff & Johnson, 1980), researchers have distinguished metaphors as conceptual mappings from their linguistic representation in discourse, and developed and formulated precise criteria to identify these linguistic metaphors – resulting in the Metaphor Identification Procedure (MIP) and the further developed MIPVU (Steen, 2010). However, this approach too is still quite open in terms of implementation, and lacks clear instructions on labelling source and target domains. This more practical approach of organising data analysis through coding, in particular the coding of source and target domains, and potential implications for issues around reliability and validity, is often discussed in less depth in handbooks, theoretical papers and empirical work.

It is this gap this protocol aims to fill, by providing the procedure in the section *Coding Metaphors*. However, we will embed the development of this procedure in reflection on the process and more general challenges we encountered, which we discuss in the section *Discussion and Conclusion*. This section also elaborates on how these reflections are more broadly relevant for researchers doing some form of qualitative coding (mainly of smaller linguistic units). First, however, we will discuss some key literature on metaphor analysis and qualitative coding in the section *Literature Review*, and data and background in the section *Research Questions and Sampling*.

Literature Review

A number of scholars have reflected on how to approach coding of metaphors, and how to increase the strength and reliability of metaphor analysis. Many of these strategies to increase validity and reliability are as such not unique to metaphor analysis, but deemed good practice in qualitative research more broadly (e.g., Creswell & Creswell, 2018; Patton, 1999; Saldana, 2009). However, each research design, data set and analysis is different, and identifying metaphors is often different from for instance themes in thematic analysis or Grounded Theory. The identification of linguistic metaphors usually involves a smaller unit of analysis, such as words or word groups (e.g., Steen, 2010). In addition, metaphor studies aim for uncovering the conceptual metaphors that underlie their linguistic representation (Deignan, 2016). It therefore remains valuable to zoom in on specific validity and reliability strategies, or how to adapt them to the context of metaphor

coding, before relating them to these issues of reliability and validity around qualitative methods more generally.

Regardless of the scale or implementation, metaphor analysis usually involves two phases: first, the identification of metaphorical expressions, and second, the categorisation of the source domain (or vehicle) that is used by interlocutors to communicate about a particular target domain (or topic). For the first phase, a well-documented procedure to identify metaphors is MIPVU (Steen, 2010), involving a set of steps to identify all so-called ‘metaphor-related words’ in discourse. These are lexical units involving an incongruity between their contextual meaning and a more basic meaning, which can be related to each other with a comparison. To establish the ‘basic meaning’ of lexical units, MIPVU prescribes consulting a corpus-based dictionary, which is not available for all languages, so the procedure has been adapted to research on these languages (such as Dutch, see Pasma, 2019). MIPVU has nonetheless been applied to various corpora and languages, usually involving multiple coders, allowing reliable analysis of metaphor frequency. However, although following a clear procedure, some interpretation by the researcher can still be necessary when using resources such as dictionaries.

For the second phase of metaphor analysis, categorising source and target domains, no such clear procedure exists. Grouping of articulated source domains may occur either inductively, i.e., by categorising semantically similar terms in the data, or deductively, i.e. by drawing upon the source domains listed in the metaphor literature, or a combination of both (e.g., Cameron et al., 2009). Especially in this phase, interpretation is inevitable, and the researcher’s (cultural and other) background will play a role: individuals are likely to utilise different conceptual metaphors for the same topic and these may be reflected in their analyses. Scholars have reflected on ways of dealing with the complexity of this part of the coding process, such as Maslen (2016), who uses the systematic metaphor approach that categorises metaphors into vehicle groups:

Selecting vehicle groups is an iterative process. The group you begin with is not necessarily the one you end up with. To guarantee trustworthiness as a balance to imagination, coding should ideally be collaborative, with a portion of one person’s work being checked by another, and reflective, in that one should be open to applying changes back through the data if it appears an earlier decision was not the best one. Keeping notes during coding is invaluable when it comes to checking how you got to where you are, especially when a lot of data are involved. (p. 94)

However, the coding of metaphors, and dealing with bias in interpretation, remains a challenge (Armstrong et al., 2011; Davis et al., 2015; Schmitt, 2005). Colleagues have therefore proposed to not analyse metaphors in a vacuum, but to always triangulate and use other forms of data analysis. Armstrong et al. (2011), for instance, discuss two forms of triangulation to deal with the culturally-dependent, subjective nature of

recognising and categorising metaphors. The first one is metaphor checking or crowd checking by participants: showing the identified metaphors to participants and including their input in the analysis, as also done by Davis et al. (2015). The second one is doing metaphor analysis as part of ethnographic research efforts, and thus triangulating it with other data sources such as interviews, observations, and field notes. By combining metaphor analysis with these forms of other contextual, immersed understanding, the cultural understanding of the participants is taken into account, as also suggested by others (e.g., Patton, 1999; Seung et al., 2015). Another strategy to become aware of one's own cultural background in interpretation is to self-interview, and analyse one's own use of metaphors (Schmitt, 2005).

Furthermore, Schmitt (2005) suggests to broadly document the research process, to interpret in groups, and to use standardised procedures. Another common suggestion, like in the quote by Maslen (2016) above, is collaborative coding and analysing with multiple coders. This can help identify and overcome personal biases and personal interpretation, but can come with its own challenges, both in metaphor coding and beyond (Beresford et al., 2022). In most cases, it requires a strong and supportive management structure (Giesen & Roeser, 2020), and extensive documentation (Giesen & Roeser, 2020; Hemmler et al., 2022). To make sure all team members are capable of doing the coding properly, time and resources for training need to be provided. When all coders are ready for coding, organising sessions for simultaneous test coding and troubleshooting are often helpful (Giesen & Roeser, 2020).

In sum, the available literature addresses a number of practical challenges and how to deal with them, both in coding metaphors specifically and coding qualitative data in general. These suggestions – inductive, iterative and continuous development of codes, thorough documentation, group interpretation and checking, and reflection – also resonate with more general qualitative coding guidelines, for instance in thematic coding or Grounded Theory (e.g., Boyatzis, 1998; Braun & Clarke, 2006; Charmaz, 2008; Evans, 2013; Saldana, 2009). However, as these remain general concepts and approaches, each study requires a tailored implementation of these. The *Coding Metaphors* section in this note therefore provides a hands-on protocol. First, we will discuss the background of our data set and our research topic.

Research Questions and Sampling

Our metaphor project was part of a larger project on discourses on the body and pain in medical consultations between chronic pain patients and anaesthesiologists, psychologists and physiotherapists. For this, 37 consultations and 12 interviews with patients were collected at a Belgian pain clinic (March-May 2019). The data were

Table 1. Data Set.

Patient	Doctor	Psychologist	Physiotherapist	#/patient
P10	X	X	X	3
P11		X	X	2
P24	X	X	X	3
P25		X	X	2
P26	X	X	X	3
P27	X	X	X	3
TOTAL				16

collected by the PI (first author). Patients were informed about the study and provided written consent.

For the empirical studies on metaphor, we were interested in the following research questions:

- Which metaphors on medicine, health, illness, the body and pain can be found in pain clinic consultations, and what do they tell us about illness/pain experiences and communicating about this?
- How do these metaphors recur across one speaker's discourse and across all speakers' discourse, and (how) are they taken up, accepted and resisted?

For the analyses, a subset of 16 consultations were used, each lasting between 12 to 60 minutes. We decided on these 16 because these consultations were part of the intake trajectory of the pain clinic, which also included an intake with the in-house physiotherapist and psychologist. This resulted in the data subset represented in Table 1.

We chose this subset as it represents different health professionals and different medical disciplines, as well as different patients, who featured at least 2 times in the data set. At the same time, the set-up of these consultations, as they were intakes, is fairly standardised and similar across patients.

All of the coding was done by the principal investigator (PI, first author), the main collaborator (MC, second author) and a student assistant. The empirical analyses were developed by the PI and MC.

Coding Metaphors

Setting up Coding

The coding process consisted of five stages from setting up coding till the second cleaning round (see Figure 1).

After hiring a student assistant to help in the coding process, the PI wrote a 'getting started' document to inform the student and collaborator on the metaphor work package of this project. This included notes on practical issues, e.g., access to the transcripts and confidentiality, some details on the data set and context, and some preliminary ideas on the focus of the study, based on the larger project this study was a

part of. Both the collaborator and student assistant familiarised themselves with the larger context of this study and the type of data, by reading this document and the research proposal, and having a first look at the transcripts. The student assistant had some experience with metaphor analysis, but was given some time to do more reading, filling in their knowledge where necessary.

Once the goal of this project was clear, we had project meetings to develop a first set-up of the coding approach. We decided to use ATLAS.ti and develop our codes inductively. However, we still needed a strategy for (1) identifying what counted as a metaphor/metaphorical expression, (2) determining whether the identified metaphor addressed the target domains of interest (see below) and (3) labelling each metaphor.

For step 1, we applied an adjusted version of the MIPVU. We only took into account metaphorically used content words (i.e., verbs, nouns, adjectives, adverbs) and then coded the entire clause or sentence in which they occurred (including both indirect and direct metaphors, see Steen, 2010). This was because we were not necessarily interested in the level of metaphoricality of the interactions, but rather in the kind of source domains that were mapped upon the target domains we selected.

We developed step 2 as we did not want to include every metaphor, but only those (broadly) relevant to living with illness and pain (also including social and psychological dimensions). In the getting started document and in the coding protocol (see below), we created a more detailed list of target domains we wanted to include, which we also further developed inductively in the test coding. At this stage we also returned to the literature and existing inventories of metaphors/SDs/TDs. So, to be able to group more metaphors together, we shared and listed overviews of source domains and semantic fields found in the literature (e.g., Mohler et al., 2016), using an iterative inductive strategy to develop our labels.

For step 3, we wanted to inductively determine the source and target domain for each linguistic metaphor. After some test coding and brainstorming, we decided to give each linguistic metaphor multiple labels, following a fixed procedure:

- Speaker
- Source domain
- Target domain

A (translated) example of this is:

Het gaat in de goede richting <i>It is going in the right direction</i>	# P (for patient) SD: Journey TD: Treatment
--	---

In the analysis stage, this approach allowed for discussing metaphorical mappings between particular source and target domains, but also separate analyses for source and target domains. By coding the speaker, we created the option to analyse speaker-specific metaphor use and include this in future interactional analysis.

After a test coding round by the student assistant using this system and inclusion criteria, the student coder fully coded one consultation. The PI and MC then reviewed this coding effort individually, which we then discussed together during a project meeting. We focused on checking whether we included the same extracts as metaphors, and how we labelled source and target domains. We further streamlined our approach, e.g., we standardised labels for SDs and TDs by removing articles (to avoid having two identical but separate codes like *SD: the journey* and *SD: journey*), and we refined inclusion criteria further, for instance regarding simile, metonymy, expressions and idioms, and English words and phrases in the otherwise Dutch corpus.

We then documented our coding system in a coding protocol. This document contained inclusion criteria (when to include something as metaphorical, which TDs to include, instructions on English, metonymy and idioms, etc.); instructions on how to work with ATLAS.TI; and a small selection of examples of codes for SDs and TDs that we had already encountered multiple times at this point, for which we had already decided what terms to use. This was the basis for the next step, the first full coding round.

First Coding Round

In the first round, all three coders coded 5–6 consultations individually, but in a phased way: we had regular project meetings every time we coded one or a few consultations. During these, we discussed difficult cases that we documented in an online Excel file. The file had three tabs, one per coder, which were constructed as follows:

- Consultation number
- Time stamp of extract

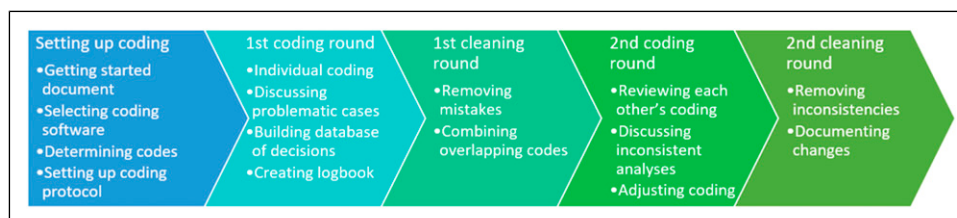


Figure 1. Coding process.

- Extract/text from transcript
- Question/issue for discussion
- Opinion other coder 1
- Opinion other coder 2
- Final decision
- Closed? Yes/no
- Implemented? Yes/no

We first provided input on each difficult case in the online Excel file. If this yielded a clear consensus among all three coders, the initial coder took a final decision, and documented this in the file. If not, we discussed it orally during the (online) project meetings and came to a consensus then. After the final decision, either reached in the Excel or orally, the initial coder also added/deleted/adapted codes based on the feedback.

This approach meant that while coding, we further inductively and collaboratively developed our understanding of metaphors, our source and target domains, and thus our code book. By documenting it extensively, per difficult case, on (virtual) paper, we also built a database or inventory of our decisions as well as our reasoning and the ideas and literature supporting it. Consequently, it increasingly became a document we could independently and individually consult when encountering a new difficult case, to see whether an identical or similar one could inform this new decision. Although project meetings remained important, coding and making decisions did become progressively more efficient like this.

To keep track of oral decisions in project meetings, to do lists, motivations for decisions and other thoughts and ideas, but also preliminary insights for analysis and connections we saw to the literature, we also kept a detailed logbook, with entries ordered by date.

First Cleaning Round. After the first coding round, it became clear that coding was not finished: a number of codes were messy, and some issues were unresolved. We carried out a cleaning round by correcting spelling errors and merging a number of similar or identical source and target domain codes, such as ‘bomb’ and ‘explosion’, or ‘nerves’ and ‘nervous system’. As part of this, we particularly checked all codes with only 1 or 2 data points in them. Although in an analysis like this it is not a problem to have unique or low-frequent codes, we wanted to make sure they did not actually overlap with other codes.

However, further refinement was required to address a number of substantial issues. Our prior experience and feedback from colleagues indicated the need for more sophisticated codes for metaphors we then assigned ‘location’ or ‘object’ as source domain, in utterances like: ‘the pain was following me everywhere’, and ‘the pain is always with me’. To address this, we consulted the literature and held additional project meetings to improve our codes. After reaching consensus on a new set of codes, the first author performed a full re-coding of all items in the relevant categories.

All cleaning was documented in an Excel file, and all motivations for decisions were written down in the logbook.

Finally, after this cleaning, we formulated a list of issues to improve in the second coding round, such as including the new codes and selecting a longer text segment in the transcripts to attach the codes, to provide more context.

Second Coding Round

For the second coding round, we decided not to do a blind coding round, as we had already looked at so much data together, and had done so much joint development of codes. We decided it was more useful to review each other’s coding work, and give feedback on that. So, each coder coded 5–6 consultations they had not coded before, looking at 2–3 consultations for each other coder in round 1.

So, the second coding round consisted of the following steps: Second coder individual work, part 1:

- Reread the full transcript, this time fully coded with the codes of round 1 in the ATLAS.TI file
- Determine whether any utterances that are metaphorical were not coded in the first round;
 - if so, add new codes
 - report all additions in the excel file for round 2 (see below) for review
 - decide whether additions need reviewing by the team
- Review the existing coding. Check whether you agree if
 - the SD being relevant/eligible for inclusion? If not, delete all codes
 - the TD, SD and speaker are appropriate/correct? If not, change codes
 - report all deletions in the excel file for round 2 for review
 - decide whether changes need reviewing by the team

Team work:

- Check the Excel file with the list of all the changes made by other coders, and decide whether you agree. If not, add a note to the Excel file this needs to be discussed in a project meeting
- Discuss difficult cases/cases marked for discussion; take a decision

Second coder individual work, part 2:

- Make final adaptations based on team decisions on difficult/marked cases

As these steps already mention, we developed another online Excel file to document the second coding round, in this case to document all changes and additions, to guarantee full transparency and make sure we could reach consensus on all changes. We tested the Excel setup with one consultation, and found it appropriate, and used it for the rest of the coding round. The structure of the Excel file was as follows:

- Consultation number
- Time stamp of extract
- Extract from transcript
- Type of change in coding: drop down menu with following options:
 - Adaptation of existing codes
 - Addition new codes
 - Deletion existing codes
 - Other
- Old codes SD/TD
 - (not relevant for newly coded extracts)
- New codes SD/TD
 - (not relevant for deleted codes)
- Have we done it like this before with similar extracts? Yes/no
- Difficult case/doubt/team discussion needed? Yes/no
- Explanation/remarks (optional)
- Opinion other coder 1
- Opinion other coder 2
- Final decision
- Implemented in ATLAS.ti? Yes/no

Again, documentation was extensive, but proved a useful resource (in combination with the excel file with the overview of difficult cases from coding round 1) to consult throughout the coding process, as it became an archive and a written version of our collective knowledge of the codes. This often helped reducing meeting time, although having oral meetings remained indispensable.

Second Cleaning Round

After the second coding round, a small number of issues remained, and some codes needed a final check for consistency. We made lists of to do items for cleaning. This included doing an exhaustive check for consistency of the location/object codes we redeveloped after coding round 1, a number of other specific items that we wanted to check for consistency, and some merging of codes. The PI did the final check, implemented changes where needed, and all of these changes again were documented in an online excel file. The excel file had the following columns:

- Item/issue
- What do we want to change in the coding/what kind of inconsistency do we expect/know is still there
- Implemented yes/no
- Did we make a lot of changes? Yes/no, and how many instances

The final category allowed for tracking how intensive this final cleaning round was, and thus to check whether two rounds of coding was enough. Fortunately, for most items, changes were minimal or not necessary at all, which made

clear the coding was now consistent enough to finalise the coding stage and to start analysis.

As this was the final coding effort, for some items/issues, the PI and MC jointly reviewed the data to determine whether and how coding needed adapting. So here too, two project members jointly took decisions in order to decrease inconsistency.

Options for Analysis

Although this protocol does not focus on the actual analysis and outcomes, we do want to give a brief overview of the possible analyses that can be run with this, or a similar, coded data set. Some options for analysing the whole data set are:

- Examine all source domains present in the data set, including their frequency of occurrence. This can be done with all codes separately, but we also did some analyses in which we first aggregated some codes. E.g., we had different specific codes for the metaphor pain as object (e.g., object that can be moved vs. object that cannot be moved). Although we wanted to capture that nuance in coding, it was less relevant for some of our analyses, so we made a code group for the 5 different object codes we had, and used that in our analysis rather than the separate codes.
- Examine all target domains, including their frequency of occurrence. Similar to the source domains, aggregation of codes before analysis is also possible here.
- Examine all/the most frequent metaphors, by crossing all source and target domains using the co-occurrence function of ATLAS.ti.
- Examine which target domains co-occurred most frequently with which source domains, and vice versa.
- Examine which source domains/target domains co-occurred (most) with either the patient or the health professional as speakers.
- Examine which metaphors were taken up, extended or resisted by consecutive speakers.

Because ATLAS.ti allows you to work with subsets of data, all these analyses could also be run for 1 specific consultation, or subsets of consultations.

These are just a few options which can be explored using ATLAS.ti; more of course is possible, and this list also does not include any form of further manual qualitative analysis and qualitative interpretation using examples, which we deem indispensable to contextualise the understanding of the main trends. In sum, the files our coding approach yielded allows for many different types of analysis, and for exploring many dimensions of the use of metaphors in our and similar data sets.

Discussion and Conclusion

For this project in which we identified and coded metaphors, the many layers of interpretation required a clear collaborative and transparent approach to coding. We needed clear inclusion criteria for what we considered metaphorical, which requires determining cut-off points on the continuum from very clear novel metaphors to very conventional ones that are almost no longer identified as metaphorical. We also needed inclusion criteria for whether a metaphor was relevant for the focus we identified for analysis, and finally, for which SD and TD a metaphor could be related to. Moreover, classifying SDs and TDs can happen on different levels: for instance, a research team can decide to capture different expressions of and nuances in violence and war metaphors, or machine metaphors, and thus develop subcategories. Our project therefore needed time and multiple coding rounds, to develop criteria and shared understandings of SDs and TDs, and extensive documentation, to keep track of previous decisions and changes for the coding to become consistent.

Although the approach detailed above has helped reduce subjectivity and variability among coders, interpretation remains key in this type of coding and qualitative research more broadly. In metaphor analysis, a degree of subjectivity cannot be fully overcome (Koro-Ljungberg, 2001). In our case too, when browsing the data for examples or for further manual analysis, we sometimes inevitably still encounter cases that raise doubt, or are not fully consistent. However, we built in a number of validity and reliability measures. In this section, we discuss how we implemented them, and how they relate to the existing literature on this topic.

First, coming back to Armstrong et al.'s (2011) suggestions, member checking was not possible in our case, due to the highly confidential nature of the data (the PI no longer has access to the patients' full names; the pseudonimisation key resides in the hospital). However, the project has an ethnographic dimension like Armstrong et al.'s (2011): besides the medical consultations, the PI also interviewed patients. Moreover, during the negotiation of access and the data collection itself, the PI and the health professionals had many more informal discussions. This concerned medical information (e.g., why pain patients often get prescribed certain types of medication), but also context (e.g., the health care system, and how the pain clinic's work relates to the work of other medical practitioners). These discussions provided invaluable background to what happened in the consultations, which, when possible in terms of anonymity, were written down. This often aided interpretation; for instance, to determine whether references to electricity were a form of machine metaphors, or literal references to electricity, as used in electrostimulation, a particular form of medical therapy. Second, the results are and will be triangulated as this metaphor analysis is part of a larger project. The data have already previously been analysed for a different paper (Declercq, 2021), and will be further analysed in future work.

Following up on Schmitt's (2005) suggestion, we did not self-interview to identify our own metaphor use. However, there was another way of, to a certain extent, becoming aware of our own metaphor use and how it influenced interpretation: the team of coders and this project was multicultural - both in the classic sense of national borders, but also in the sense that a specific context with a specific community (chronically ill patients and their highly specialised health professionals) was the object of study. First, for this project, the data was collected in Flanders, the northern part of Belgium. In this region, a variant of Dutch is spoken, also called Flemish. Although mutually intelligible, this form of Dutch is different from Flemish (Vismans, 2017). The first author, who collected the data, grew up in Flanders, and speaks Flemish as her native tongue. However, the coding took place in the Netherlands, with two coders that grew up in the Netherlands and have Dutch as their native tongue. Moreover, the consultations often contain highly specialised and technical language (which may differ between Flemish and Dutch).

Both the differences between Flemish and Dutch and the highly specialised medical jargon, sometimes required more background, either provided by Flemish dictionaries, the broader, ethnographic understanding of the PI, or an Internet search. In sum, collaborating with coders with different backgrounds and data in a specific, specialised context thus at times complicated the process, as it generated more discussion. However, this approach also allowed for a critical examination of the data and brought new perspectives, ultimately leading to improved coding and analysis. This made a lot of our interpretative work explicit in the group meetings and in the excel inventories, which allowed us to uncover biases, for instance in how we experienced the conventionality of metaphorical language. We believe that this can also be a useful takeaway for metaphor analysis and qualitative coding: it may be beneficial to consider a multicultural coding team (encompassing not just different nationalities, but also in terms of insiders and outsiders of specific cultural communities), or at least reflect on the complementarity of the coders' cultural backgrounds. We believe this makes visible the unconscious layers of interpretation that occur in any qualitative analytical process. This ties in with the distinction between etic and emic perspectives in (ethnographic) research (Lillis, 2008): having a mix of both types of coders on a team might be advantageous. In our case, the PI had a more emic perspective on the data, while the MC and student assistant had etic perspectives.

Finally, besides the ethnographic approach and working with multiple coders with different backgrounds, we took other measures to increase reliability, most of which are not new in qualitative research:

- Extensive documentation (Maslen, 2016; Schmitt, 2005)
- Iteration: literature – data – literature – data (Hemmler et al., 2022)
- Contextualising results and transparent reporting

Regarding the first measure, a specific trait of the project was the large amount of documentation. In metaphor analysis, it is easy to lose track of similar but slightly different instances of a metaphor or source/target domain, especially in our data set with often similar conversations in the consultations. Moreover, many joint decisions needed to be made, and this often also required documenting our reasoning or ideas behind them, in order to remain consistent, and explicate interpretative work. The team work is particularly needed in metaphor research aiming for discovering all used source domains instead of starting with a predetermined set of concepts, because of the wide variety of metaphor-related words that need to be categorised consistently. As discussed above, we documented many small steps in our analysis, enabling us to track in detail how the coding developed, what we did, when and why. In addition to the already existing research proposal, transcripts and ATLAS.ti files, this approach resulted in: a getting started document, a test coding document, a coding protocol, a logbook, and excel files per coding round and cleaning round.

Second, when in doubt, we also consulted existing literature, inventories and directories on metaphors. Although this is likely often done by metaphor researchers, it seems to be less highlighted in methodological literature on metaphor analysis. However, numerous empirical studies have already established effective, well-tested categorisations of metaphors. These may be language- and culture-specific, of course, but we found that work on English data, often relates well to our (Flemish) Dutch data. This not only reinforced the reliability of our coding by avoiding overly subjective interpretations, but also allowed for engagement with the literature and prior research in our empirical papers. We therefore want to highlight the importance and relevance of iteration (Hemmler et al., 2022) in metaphor coding, which, to our knowledge, has not been mentioned in the specific literature on metaphor coding/analysis. This form of iteration is also more generally relevant for any study that makes use of categories that are documented in some kind of directory or database, or previously in studies that make use of very similar data sets and research questions.

Finally, we believe that validity also comes with a transparent presentation of results. We abundantly use examples to show the data behind our (interpretative) codes, which ultimately leaves it up to the reviewers and readers to judge whether they agree with our categorisation and interpretation.

To conclude, many of the strategies used to increase rigour and consistency in our coding are standard qualitative reliability and validity measures, that thus are not limited to metaphor analysis. Some of the choices made and steps in the protocol did turn out to be quite specific to the nature of our analysis: they for instance related to the fact that we had a three-step interpretative procedure (one: does an extract contain metaphorical language? Two: does the metaphor relate to health, illness or another target domain under scrutiny? Three: how do we categorise the source and target domain?),

and worked with a large set of codes for source and target domain that required much collaborative finetuning across coding rounds. However, we believe that many parts of the protocol and the reflection in this section apply to qualitative research and coding (mostly of smaller units such as linguistic units), and therefore may be useful for a wider range of qualitative researchers.

Acknowledgements

We want to thank the anonymous reviewers, Thomas Velvis, and our colleagues at the Discourse and Communication department for the useful feedback.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is funded by the Dutch Research Council (Nederlands Organisatie voor Wetenschappelijk Onderzoek): grant number VI.Veni.201T.003.

Ethical Approval

The study discussed in this paper was approved by the Committee for Medical Ethics of the Ghent University Hospital, Belgium.

ORCID iDs

Jana Declercq  <https://orcid.org/0000-0003-2281-3398>

Lotte van Poppel  <https://orcid.org/0000-0002-7951-5482>

References

- Armstrong, S. L., Davis, H. S., & Paulson, E. J. (2011). The subjectivity problem: Improving triangulation approaches in metaphor analysis studies. *International Journal of Qualitative Methods, 10*(2), 151–163. <https://doi.org/10.1177/160940691101000204>
- Beresford, M., Wutich, A., du Bray, M. V., Ruth, A., Stotts, R., Sturtz Sreetharan, C., & Brewis, A. (2022). Coding qualitative data at scale: Guidance for large coder teams based on 18 studies. *International Journal of Qualitative Methods, 21*, 1–15. <https://doi.org/10.1177/16094069221075860>
- Boyatzis, R. (1998). *Transforming qualitative information: Thematic analysis and code development*.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Cameron, L., Maslen, R., Todd, Z., Maule, J., Stratton, P., & Stanley, N. (2009). The discourse dynamics approach to metaphor and metaphor-led discourse analysis. *Metaphor and Symbol 24*(2), 63–89. DOI: <https://doi.org/10.1080/10926480902830821>

- Charmaz, K. (2008). Constructionism and the grounded theory method. In J. A. Holstein, & J. F. Gubrium (Red.), *Handbook of constructionist research* (pp. 397–412). The Guilford Press.
- Cornelissen, J. P., Oswick, C., Thøger Christensen, L., & Phillips, N. (2008). Metaphor in organizational research: Context, modalities and implications for research — Introduction. *Organization Studies*, 29(1), 7–22. <https://doi.org/10.1177/0170840607086634>
- Creswell, & Creswell, D. (2018). *Research design* (5th ed.). SAGE.
- Davis, H. S., Watson, A. B., & Bakerson, M. (2015). Crowdchecking conceptual metaphors: How principals and teachers frame the principal's role in academics through metaphor. In W. Wan, & G. Low (Red.), *Metaphor in language, cognition, and communication* (pp. 139–165). John Benjamins Publishing Company. <https://doi.org/10.1075/milcc.3.06dav>
- Declercq, J. (2021). Talking about chronic pain: Misalignment in discussions of the body, mind and social aspects in pain clinic consultations. *Health*. <https://doi.org/10.1177/13634593211032875>
- Deignan, A. (2016). From linguistic to conceptual metaphors. In E. Semino, & Z. Demjén (Red.), *The Routledge handbook of metaphor and language* (pp. 102–116). Routledge.
- Demmen, J., Semino, E., Demjén, Z., Koller, V., Hardie, A., Rayson, P., & Payne, S. (2015). A computer-assisted study of the use of Violence metaphors for cancer and end of life by patients, family carers and health professionals. *International Journal of Corpus Linguistics*, 2, 205–231. <https://doi.org/10.1075/ijcl.20.2.03dem>
- Evans, G. (2013). A novice researcher's first walk through the maze of grounded theory: Rationalization for classical grounded theory. *The Grounded Theory Review*, 12(1), 37–55.
- Gibbs, R. J., & Franks, H. (2009). Embodied metaphor in women's narratives about their experiences with cancer. *Health Communication* 14(2), 37–41. <https://doi.org/10.1207/S15327027HC1402>
- Giesen, L., & Roeser, A. (2020). Structuring a team-based approach to coding qualitative data. *International Journal of Qualitative Methods*, 19, 1–7. <https://doi.org/10.1177/1609406920968700>
- Hemmler, V. L., Kenney, A. W., Langley, S. D., Callahan, C. M., Gubbins, E. J., & Holder, S. (2022). Beyond a coefficient: An interactive process for achieving inter-rater consistency in qualitative coding. *Qualitative Research*, 22(2), 194–219. <https://doi.org/10.1177/1468794120976072>
- Koro-Ljungberg, M. (2001). Metaphors as a way to explore qualitative data. *International Journal of Qualitative Studies in Education*, 14(3), 367–379. <https://doi.org/10.1080/09518390110029102>
- Lakoff, G., & Johnson, M. (1980). Conceptual metaphor in everyday language. *The Journal of Philosophy*, 77(8), 453–486, <https://doi.org/10.2307/2025464>
- Lillis, T. (2008). Ethnography as method, methodology, and 'deep theorizing': Closing the gap between text and context in academic writing research. *Written Communication*, 25(3), 353–388. <https://doi.org/10.1177/0741088308319229>
- Maslen, R. (2016). Finding systematic metaphors. In E. Semino, & Z. Demjén (Red.), *The Routledge handbook of metaphor and language* (pp. 88–101). Routledge.
- Mohler, M., Brunson, M., Rink, B., & Tomlinson, M. (2016). Introducing the LCC metaphor datasets. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pp. 4221–4227. European Language Resources Association (ELRA).
- Pasma, T. (2019). Linguistic metaphor identification in Dutch. In: S. Nacey, A. G. Dorst, T. Krennmayr, & W. G. Reijnierse (Red.). *Metaphor identification in multiple languages: MIPVU around the world* (pp. 91–112). John Benjamins.
- Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, 34(5 Pt 2), 1189–1208. <http://dx.doi.org/10.4135/9781412985727>
- Saldana, J. (2009). *The coding manual for qualitative researchers*. SAGE.
- Schmitt, R. (2005). Systematic metaphor analysis as a method of qualitative research. *The Qualitative Report* 10(2), 358–394. <https://doi.org/10.46743/2160-3715/2005.1854>
- Semino, E., Demjén, Z., Demmen, J., Koller, V., Payne, S., Hardie, A., & Rayson, P. (2017). The online use of violence and journey metaphors by patients with cancer, as compared with health professionals: A mixed methods study. *BMJ Supportive & Palliative Care*, 7(1), 60–66. <https://doi.org/10.1136/bmjspcare-2014-000785>
- Seung, E., Park, S., & Jung, J. (2015). Methodological approaches and strategies for elicited metaphor-based research: A critical review. In W. Wan, & G. Low (Red.), *Metaphor in language, cognition, and communication* (pp. 39–64). John Benjamins Publishing Company. <https://doi.org/10.1075/milcc.3.02seu>
- Steen, G. (Red.). (2010). *A method for linguistic metaphor identification: From MIP to MIPVU*. John Benjamins Pub. Co.
- Steen, G. (2016). Identifying metaphors in language. In E. Semino, & Z. Demjén (Red.), *The Routledge handbook of metaphor and language* (pp. 91–105). Routledge.
- Vismans, R. (2017). Negotiating address in a pluricentric language: Dutch/Flemish. In C. Norrby, & C. Wide (Red.), *Address practice as social action: European perspectives* (pp. 91–94). Palgrave Macmillan.