Transcription as a Form of Qualitative Inquiry

Charles Vanover

Abstract

The choice of a transcription strategy reverberates across the qualitative research life cycle to influence the meanings people derive from transcribed interviews and other parts of the research project. This chapter describes how and why transcription matters and discusses important decision points and strategies for transforming recorded speech. It publishes multiple transcriptions of the same interview segments to convey meanings generated by different forms of transcription. Atheoretical transcription practices that render participants’ discourse with shallow streams of meaning are contrasted to intentional transcription strategies that encourage understanding and theory building. These choices make a difference. The author discusses how his decision to use a simple set of Microsoft Word formatting commands to construct intentional transcripts that narrativized participants’ speech altered the course of his research project. The chapter concludes by discussing the costs of engaging in intentional transcription practice and shares strategies for minimizing laborious aspects of the work.

Keywords: Transcription strategies; narrative interviews; reflexive qualitative research practice

Transcription is a method of transforming what people say and do into forms of text. Audio and video recordings are changed into words and images that allow researchers to analyze and interpret aspects of those research events. Transcription
selects and condenses the complex series of interactions produced by interviews; it renders the big booming buzzing confusion of raw conversation into forms that support inquiry.

Transcription is more than a means to an end. The work of transcription provides occasions for researchers to immerse themselves in their data and develop insight. By seeing and hearing recordings of their interviews and building text from that content, researchers deepen their comprehension of the materials they study. The work of building these meanings and insights lies at the heart of the analytic work for qualitative inquiry (Charmaz, 2014, 2017; Locke, Golden-Biddle, & Feldman, 2004).

In this chapter, I review the research on transcription and discuss the symbols and Microsoft Word formatting commands I used to transcribe a set of narrative interviews for my dissertation research. I share the results of two different transcription strategies for two interview segments from this narrative dataset. The first examples provide atheoretical transcriptions of the segments—what the verbal data might look if the transcript was produced by a paid transcriber. The same interview segments are then transformed by a set of naturalized transcription practices intended to enhance readability and comprehension. My dissertation research evolved into arts-based inquiry. I conclude the chapter by briefly discussing how my transcription strategies supported this aspect of my work.

**Literature Review: Transcription as Method**

Using transcription as an intentional practice requires careful planning and decision-making. Different research designs and study goals necessitate different transcription strategies (Greenwood, Kendrick, Davies, & Gill, 2017; Lapadat & Lindsay, 1999). Conversational analysis does not use the same transcription practices as grounded theory. Transcripts of interviews recorded on video may demand different strategies than interviews recorded on audiotape (Heath, Hindmarsh, & Luff, 2010; Mondada, 2018).

The first step in transcribing all forms of data is to assess how the raw data look and sound. It may be worthwhile to review the recordings and test out possible forms of transcription. Researchers must evaluate whether a transcription strategy provides rich material for the deployment of the analytic practices specified by the research design. They may also investigate whether the transcribed text supports the future coding system and/or decide whether the transcripts complement their chosen arts-based practices.

An important step is to figure out how much of the data should be transcribed. It is always possible, and sometimes even desirable, to transcribe every interview in the same standardized fashion—see McLellan, MacQueen, and Neidig (2003) for rules and procedures to achieve such a goal. However, such standardization is not necessary for all research questions and study designs (Eaton, Stritzke, & Ohan,
Conversation analysts tend to transcribe small portions of their transcriptions in order to investigate the speech acts that are the focus of their inquiry (Bartesagh, Chapter 5; ten Have, 2007). Grounded theory and narrative researchers might create verbatim transcriptions of the material related to their research questions while taking notes that describe less relevant parts of the interview. These notes might have time stamps or links to the recorded data that guide future users back to specific passages if they wish to investigate in more detail.

A wide array of published systems, practices, and strategies are available to guide researchers’ efforts to transcribe their chosen interview segments (Bucholtz, 2007; Davidson, 2009; Lapadat & Lindsay, 1999). Dressler and Kreuz (2000) provide a model system of transcription based on a five-year study of articles published in the journal, *Discourse Processes*. The article provides what Dressler and Kreuz hope to be a standardized and universal set of symbols for 21 dimensions of oral discourse ranging from intonation to pauses, speech speeds, and breath. Crow (1988) uses ethnomethodology to create a framework of 17 performance acts such as playful deceits and code switching. Rather than sharing recommendations for a standardized system, Davidson (2009) provides a guide to choosing from different approaches to transcription. Davidson discusses the conduct of transcription during qualitative research and describes important decision points within the transcription process.

The use of specific transcription strategies may have important ethical and political implications. Oliver, Serovich, and Mason (2005) discuss the challenges confronting their research team as they transcribed and analyzed a set of interviews designed to study the disclosure decisions of HIV-positive men who have sex with men. The research team interviewed men from a wide range of ethnic backgrounds and were able to collect rich accounts of the men’s decisions to disclose their HIV status to their sexual partners and, a felony offense in many states, the men’s decisions not to disclose. Oliver et al.’s article shares examples of the complex decision-making the research team engaged in to produce renderings of the interviews for their analytic and interpretive work while protecting confidentiality.

A major focus of recent research on transcription is the challenge of rendering gestures, speech, and other aspects of discourse in interviews and observations recorded on video. Flewitt (2011) provides a discussion of key issues in the development of transcripts from video recordings and the use of these dynamic texts during inquiry. Cowan (2014) applies different sets of transcription conventions to the same video segment to show how these conventions draw attention to alternate dimensions of the recorded discourse. Markle, West, and Rich (2011) describe the many weaknesses of transcription as a research method. Markel et al. argue that coding video data directly from the digital recording provides more meaningful material for analysis than expending the time and resources required to develop potentially inaccurate and untrustworthy transcripts. Bernauer (Chapter 10) discusses the benefits and challenges of engaging in this form of transcriptionless coding.
Transcribing by hand and cleaning raw transcripts are ways of getting to know the data. Regardless of the specific analytic choices made to develop the text, time spent doing transcription may lay the grounds for abduction, insight, and theory-building (Alvesson & Karreman, 2011; Charmaz, 2011; Saldaña & Omasta, 2018). Careful transcription is also an ethical imperative. A core expectation of qualitative research is that interview participants’ discourse is rendered with appropriate levels of accuracy.

**Positioning Statement and Study Description**

I am a White man who grew up in the Chicago suburbs and worked in the Chicago Public Schools (CPS) for eight years before I won a fellowship to study educational policy at an out-of-state research university. When I returned to Chicago to do the fieldwork for my dissertation, I conducted four rounds of semi-structured narrative interviews with five first-year teachers and seven veteran and accomplished educators working in the CPS (see Vanover, 2014 for interview guides and other study materials). At the time, my practice as an interviewer and qualitative analyst was immersed in the work of two writers. The first was Patricia Benner, whose focus groups and narrative interview methods generated rich streams of stories about nurses’ courage and agency (Benner, Hooper-Kyriakidis, & Stannard 1999; Benner, Tanner, & Chelsea, 1996). The second writer was Robert Weiss, whose work provided the primary guidance for my research design. Weiss (1995) developed tools and strategies for semi-structured interview sessions that give participants control over what they say during their interviews and help participants share what they learn from experience.

While it may be a wonderful feeling to engage in a data collection effort that produced 48, ninety-minute interviews, I experienced something close to panic once I came back to my university. I carried my bags to my apartment, checked the messages on my answering machine, and then sat at my desk and tried to figure out what to do with the hundreds of photographs and 80 hours of audio recorded stories I had collected in Chicago.

During the time when I developed my research design, I was part of an interdisciplinary seminar of narrative researchers. Our discussions inspired me to transcribe the first round of interviews by hand. The graduate students and faculty in my research seminar viewed hand transcription as a best practice: transcription slowed the process down and created time for reflexivity and theory-building. Instead of listening to the stories the teachers shared, I decided to embody them by physically reproducing the teachers’ words.

I will never be able to calculate the relationship between the time I spent transcribing the data and the quality of the final products I produced. However, one thing is certain: The work took a great deal of time. Transcribing 12 ninety-minute interviews is a labor-intensive task. The effort grows large if the goal is to reflexively
engage with participants’ storytelling, rather than to type out the words. I loved listening to the interviews, but many weeks did pass as I transcribed by hand from the original cassette tapes using a pedal transcription player.

Narrativizing the Interview Text and Other Transcription Strategies

The first problem I confronted when I transcribed the teachers’ interviews was to create a meaningful document out of the thousands of words and other discourse produced by my interview methods. I conducted most of the first round of interviews in June, within two weeks of the end of the school year. During those sessions, there were times when the teachers fell so deep into their storytelling, five minutes or more might pass without any need for me to ask a follow-up question. The teachers shared so many words during their interviews, I found it challenging to transcribe what was spoken. Stories that fascinated me during the interview session became flat rows of words on a computer screen.

Table 4.1 illustrates the stakes of the work. The table’s first row publishes an atheoretical transcription from my interview with Halsted Hoyne (pseudonym), the first recorded interview I transcribed for my dissertation. This row shares the type of text that might be produced if my intention had been to transcribe Halsted’s interview quickly and record the spoken words without reference to overriding theories or goals—i.e., it renders the teacher’s speech as if I was paid to transcribe the interview by someone I did not know for a project I did not understand. The transcript also has some inaccuracies; it does not transcribe every single word. A quick glance at Table 4.1 shows that such atheoretical transcriptions are not fertile ground for pondering, musing, and engaging in other analytic and interpretive practices. The text is exhausting and difficult to read.

Given Halsted spoke more than 26,000 words during her first interview, one can imagine the importance of developing a worthwhile transcription strategy. To organize these verbal data, I decided to engage in what I now describe as a form of naturalized transcription practice (see Bucholtz, 2000; Davidson, 2009; Oliver, Serovich, & Mason, 2005). I created a set of strategies to narrativize the text: I transcribed the teacher’s stories using transcription techniques that enhanced readability and helped the text read like a story. I believe the decision to develop these strategies and narrativize the interviews changed the course of my research project.

The text in the second row of Table 4.1 is an example of my transcription strategy. Every time Halsted and the other teachers revoiced their speech and spoke in the voices of students, parents, colleagues, principals, or themselves, I hit the return key and used different Microsoft Word fonts to differentiate such speech from the teacher’s more general narrative voice. After the usual time spent in trial and error learning, I decided what made the most sense was to use one font for
student voices, another font for adult voices, and then a different font for teachers’ revoicing of their own speech. I did not need to mark the speech of each type of school professional with a different font or designate one font for Kid A and another font for Kid B. I could use NVivo software coding commands to make such comparisons if they were important later in the analysis.

The results of my transcription strategy can be seen by glancing between the two renderings of Halsted’s story in Table 4.1. The narrativized text in Row 2 is much easier to read and more meaningful. I was able to achieve this goal not by altering or rearranging what the teacher said, but by changing how Halsted’s words appear on the page and then adding some rough punctuation to highlight her speech. I also attempted to write down every word she uttered, without attempting to fix her grammar.

I do not have records of the decision-making process that produced this transcription strategy. I do know Halsted’s interview was the first session I transcribed.
for my dissertation and that I developed most of my transcription strategy doing
the work for that interview. If I had saved a Microsoft Word copy of the transcript
each day after I transcribed, it might be possible—if perhaps not desirable—for me
to publish my initial renderings of her stories. I could then analyze the changes to
particular interview segments as my transcription strategy developed by hand and
foot pedal.

**Marking the Interviewer’s Affirmations**

In the next three sections, I use the term “interviewer” and the pronoun “he”
for clarity in order to emphasize that I am no longer the person who sat in the
offices of the Chicago Teacher’s Union and the University of Chicago Alumni
Association and conducted the interviews. Instead, I am the person who is
analyzing the interviewer’s work.

One of the goals for my research was to compare the stories of the first-year teachers,
such as Halsted, with those of the more experienced veteran educators. In order for this
comparison to make sense, however, it was necessary to show that the interviewer was
not attempting to lead the interviews. Thus, I thought it important to transcribe the
interviewer’s exact words in the same verbatim stream I used to record the teachers’
speech. I believed it was important for readers to see the questions the teachers
responded to during the flow of the interview. I also decided to mark the moments
when the interviewer responded to the teachers’ stories and, as Weiss (1995) recom-
mends, affirmed their discourse, mostly by muttering in agreement.

I eventually realized that I did not need to record every “‘uh-uh,” “nnhuh,” and
“yeah nnnmmhmmmm” the interviewer muttered. For my study goals, what
mattered was the interviewer had affirmed the teacher’s story. I decided to mark all
of the interviewer’s many grunts of agreement with two asterisks. This strategy
broke the transcript into more easily readable units, while keeping the focus of the
text on the teachers, rather than the interviewer. These marks can be seen in the
text in Table 4.2, Row 2.

**Marking the Teachers’ Pauses**

Halsted Hoyne told me she spoke about her teaching every morning of her first year
in the Chicago Public Schools when she and one of her friends drove to work, and
every afternoon when the two teachers drove home. Halsted spent her evenings
talking about her classroom to her husband and to friends from her teacher edu-
cation program. When I interviewed Halsted in the offices of the Chicago Teacher’s
Union in June 2004, she had a lot to say. Halsted leaped from story to story and she
shared more words than any other teacher I interviewed. My initial transcription
strategy was developed to make such material meaningful.

In contrast, Indiana Ingleside (pseudonym), the teacher whose interview
segment is transcribed in Table 4.2, told me she spoke to almost no one about her
teaching and rarely talked to anyone in her school besides her students. During her interview, Indiana paused frequently, mostly I believe, to search for words. As I transcribed Indiana’s interview, I decided to document those significant pauses. I felt they added meaning to her transcript.

Some conversation analysts decide to mark pauses down to a 10th of a second, but after the usual amount of trial and error learning, I realized I did not need to

---

**TABLE 4.2  Two Versions of the Same Interview Segment From Pseudonymous Teacher, Indiana Ingleside**

<table>
<thead>
<tr>
<th>Version 1: An Atheoretical Transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interviewer</strong></td>
</tr>
<tr>
<td>Okay great. Could you talk about moments during the year when you felt that you had really learned something new about your teaching or your students?</td>
</tr>
<tr>
<td><strong>Indiana</strong></td>
</tr>
<tr>
<td>Everyday. I learned so much from the kids this year. It’s hard to pick one, but</td>
</tr>
<tr>
<td><strong>Interviewer</strong></td>
</tr>
<tr>
<td>Just pick a bunch, that’s fine.</td>
</tr>
<tr>
<td><strong>Indiana</strong></td>
</tr>
<tr>
<td>A lot of what I learned is just how kids react to their environment and how they react to how people treat them. Before this year I did a year on the East Side as part of my Master’s program and then this was my first year with my own classroom. I had never worked with kids this closely as this year. So it was a really a learning experience on a lot of levels. Just the psychology behind interacting with them and their interacting with each other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Version 2: A Narrativized Transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interviewer</strong></td>
</tr>
<tr>
<td>Okay great. Could you talk about moments during the year when you felt that you had really learned something new about your teaching or your students?</td>
</tr>
<tr>
<td><strong>Indiana</strong></td>
</tr>
<tr>
<td>Everyday. (LAUGHS). Um (10 SECOND PAUSE) I learned so much from the kids this year. I—it’s hard to pick one, but</td>
</tr>
<tr>
<td><strong>Interviewer</strong></td>
</tr>
<tr>
<td>Just pick a bunch, that’s fine.</td>
</tr>
<tr>
<td><strong>Indiana</strong></td>
</tr>
<tr>
<td>A lot of what I learned is just how (5 SECOND PAUSE) like how kids react to their environment and how they react to how people treat them and (6 SECOND PAUSE) before this year I just—I had ** I did a year on the East Side as part of my Master’s program and then this was my first year with my own classroom. I had never worked with kids this closely (LAUGHS) as this year. So it was a really a learning experience on a lot of levels. Just ** (4 SECOND PAUSE) Uhh Just the psychology behind interacting with them and and and their interacting with each other</td>
</tr>
</tbody>
</table>
transcribe in such fine detail. What worked best was to mark all pauses that seemed to be meaningful interruptions in the flow of speech, and I decided to measure, by the second, all pauses that lasted 3 seconds or more. The two versions of Indiana’s interview in Table 4.2 show the changes produced by this strategy—I strongly recommend the reader time each pause to hear the full meanings of those speech acts. The transcripts also show the benefits of my decision to transcribe the teachers’ exact words and mark important forms of discourse such as laughter.

Quantifying Transcription Symbols

I applied my chosen transcription strategies in a standardized fashion across each of the 12 interviews in the first round. I wrote each full word that was spoken in each interview and I formatted every piece of revoiced dialogue. I decided that if I was going to compare the interviews of the beginning and veteran teachers, their transcripts should all read in the same way. During the analysis, I often went back to the original interview tape; this labor helped me fix inaccuracies in the transcription and ensured my initial transcription choices made sense. Because I did not transcribe the pauses in the first two interviews, for instance, I had to relisten to those cassette tapes to add that notation.

This standardization provided an unexpected analytic benefit. When I was writing my methods chapter, I discovered I could use the formatted text and symbols I had created to quantify the transcriptions. I learned I could use commands in Microsoft Word to count the number of revoicings, pauses, and interviewer affirmations in each interview. I also summed up the number of words shared by the teacher and the interviewer in each session and calculated the ratio between the interviewer’s and the teacher’s recorded words. Table 4.3 shows the results for Halsted and Indiana’s interviews. (The full set of results for the first round of interviews is discussed on pp. 106–116 of Vanover, 2009.)

I used this evidence to support my claim that while the conduct of each interview varied, there was no intentional variation in the interviewer’s work. The interviewer did not use one style of interviewing for the beginning teachers and engage in a different style for the accomplished, veteran educators. Such findings, of course, cannot be used to claim that the interviews were conducted without bias. My status as a White, male, former CPS teacher, and out-of-state PhD student mattered. I am certain a different interviewer with other personal characteristics would inspire a different set of stories, even if they used the same interview guide.

Transcripts as Ethnodramatic Scripts

There were many strengths in the work I engaged in transcribing the teachers’ stories. I was committed to doing the work and I felt it mattered that I took the time
to listen to what the teachers said. There were also many problems with my approach. Time is not cheap. By spending time, I was also falling behind. I believed I benefited by transcribing the first few interviews by hand, but once I figured out my transcription strategy, it would have been reasonable to pay a transcriber to produce the raw transcript—as I did for the second, third, and fourth round of interviews. I could have then cleaned that material and used my transcription strategies to prepare the data with little loss of engagement.

It was also the case I was working alone through the entire process and, as the previous section implies, I did not produce a regular set of memos to record my progress. My lack of memo writing hid a bigger problem: I was a good writer, but I had not developed a voice to communicate what I had learned from the interviews. Likely as a result, when I finished my dissertation and it came time to develop papers from my research in Chicago, I had trouble pulling the material together.

As I have written elsewhere (Vanover, 2016a, 2017), what worked for me was to transform the transcripts into what are now six ethnodramatic scripts. It was challenging to figure out an ethnodramatic approach for the teachers’ interviews (see Saldaña, Chapter 19), but once I figured one out, script writing came easily. Because I had narrativized the data, I saw the voices on the page when I reread the transcripts and began to develop the scripts. I was able to edit the boring parts out of the teachers’ transcripts and condense the interviews for performance because I already knew what the exciting parts were. The scripted versions of the interview segments published in this chapter can be found in Vanover (2019, p. 26) and Vanover (2016b, p. 182). Saldaña (2003, 2011) provides suggestions for developing ethnodrama from data and examples of theatrical script writing conventions.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Student voice—speaking as a student</th>
<th>Adult voice—speaking as another adult</th>
<th>Classroom voice—speaking as herself, teaching</th>
<th># of Interviewer Affirmations per Interview</th>
<th>Ratio of Interviewer Words to Respondent words</th>
<th>Percent Teacher Words</th>
<th># of Respondent Generated Pauses per Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halsted</td>
<td>64</td>
<td>79</td>
<td>288</td>
<td>384</td>
<td>755/26,715</td>
<td>97%</td>
<td>4</td>
</tr>
<tr>
<td>Hoyne</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indiana</td>
<td>18</td>
<td>1</td>
<td>48</td>
<td>100</td>
<td>644/11,315</td>
<td>94%</td>
<td>50</td>
</tr>
<tr>
<td>Ingleside</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

Decisions made during the process of transcription shape the meanings researchers develop as they engage with the text during the subsequent stages of the qualitative research life cycle. All transcription is political (Bucholtz, 2000; Ochs, 1979). Words, gestures, and other aspects of discourse are always rendered from a particular set of biases and beliefs. Transcribers’ articulated and inarticulated theories shape the words and symbols they use to render recorded speech acts. Such theories also shape machine learning. Automatic coding systems vary in their ability to render types of speech (di Gregorio, Chapter 6). Rigorous qualitative data analysis requires intentional and careful transcription.

A major theme of the chapter is that research questions and study goals shape researchers’ choice of transcription strategies and practices. An intentional transcription strategy takes into account the characteristics of the data and the analytic framework the research team applies to this material. Time spent early in the analysis engaging with the recorded data and drafts of the transcribed text may

Copyright SAGE Publications Inc. 2022. Do not distribute without permission.
build the foundation for unexpected insights and deepen researchers’ understanding of the content of the interview.

The simple Microsoft Word formatting commands used in this chapter illustrate how different transcription practices might guide the research team’s attention. As I discussed, my decision as a dissertation researcher to narrativize my transcripts enriched the possibilities for analysis and interpretation available during later stages of the project life cycle. This transcription strategy shaped the flow of meanings and images I experienced as I read the teachers’ stories.

Time is a precious resource. Qualitative researchers must find a balance between time spent engaging with the recorded interviews and time spent doing all the other work required to construct a finished study. In my view, rigorous qualitative work is not built from the unreflexive markings of paid transcribers and automatic coding systems. In my view, it is not necessary for researchers to transcribe each and every recording they make by hand. As di Gregorio (Chapter 6) discusses, it is worthwhile for researchers to outsource the more laborious aspects of the transcription process to paid transcribers and automatic coding systems if researchers then listen to the recording and engage with the unfinished transcript and organize the data to support their study goals.
The future of qualitative research is video (see Harris, 2016; Pink, 2013). As I discussed in the literature review for this chapter, almost all current, cutting-edge work on transcription focuses on this media. In my view, what holds the field back is the same thing that advances it forward: technology. Technology has yet to democratize video production in the way technology has democratized print, but the day is coming.

I no longer record my interviews onto cassette tape, but there are occasions when I take out my digital transcription pedal and transcribe by hand. I find that these transcription sessions bring me back to the joys of storytelling that first brought me to qualitative inquiry. These transcriptions are emerging as an important area of research (Langtiw & Vanover, under submission; Vanover, Knobloch, Salaam, & Agosto, under submission), and I now find it easier to justify the time to engage in this work. My work as a transcriber has been labor-intensive, but I treasure the memories.
Supplemental Readings


Reflection and Activities

1. Compare and contrast conceptions and practices related to transcription as a method of inquiry, with practices that unreflectively use transcription as an atheoretical medium for analysis.

2. Discuss how alternative sets of research questions and study goals might require different transcription strategies.

3. The ethnodrama built from the transcript in Table 4.1, Halsted’s interview, was performed as part of a series of talks on educational disparities at the Gerald R. Ford School of Public Policy: http://fordschool.umich.edu/video/2014/what-does-it-mean-work-system-fails-you-and-your-kids-beginning-teachers-journey-through-. Go to the start of the scripted performance in that video—41:30 minutes—and ask the class to develop a transcription strategy to render the first five minutes of the play. Students might have the option of building this text from the Ford’s School’s machine transcription. What aspects of the video should be selected and what aspects ignored?

References

Alvesson, M., & Karreman, D. (2011). *Qualitative research and theory development: Mystery as method*. SAGE.


Charmaz, K. (2014). *Constructing grounded theory*. SAGE.


Copyright SAGE Publications Inc. 2022. Do not distribute without permission.


