



Published in final edited form as:

Soc Forces. 2005 December ; 84(2): 1273–1289.

Constraints and Opportunities with Interview Transcription: Towards Reflection in Qualitative Research

Daniel G. Oliver,
Ohio State University

Julianne M. Serovich, and
Ohio State University

Tina L. Mason
Ohio State University

Abstract

In this paper we discuss the complexities of interview transcription. While often seen as a behind-the-scenes task, we suggest that transcription is a powerful act of representation. Transcription is practiced in multiple ways, often using naturalism, in which every utterance is captured in as much detail as possible, and/or denaturalism, in which grammar is corrected, interview noise (e.g., stutters, pauses, etc.) is removed and nonstandard accents (i.e., non-majority) are standardized. In this article, we discuss the constraints and opportunities of our transcription decisions and point to an intermediate, reflective step. We suggest that researchers incorporate reflection into their research design by interrogating their transcription decisions and the possible impact these decisions may have on participants and research outcomes.

Introduction

Despite its centrality in qualitative data collection, transcription practices remain superficially examined. It is not uncommon for transcription to be presented as a behind-the-scenes aspect of data management rather than as an object of study in its own right. As Agar (1996:153) writes, “Transcription is a chore.” While certainly there are more stimulating aspects of research, in this paper we argue that transcription is a pivotal aspect of qualitative inquiry. Largely emerging in linguistics (Ochs 1979) and linguistic anthropology (Duranti 1997), scholars from diverse disciplines (Lapadat and Lindsay 1999; Mishler 1984; Sandelowski 1994; Tilley 1998) have begun to recognize the centrality of transcription in qualitative research (Poland 2002). From these scholars, we have learned how transcription can powerfully affect the way participants are understood, the information they share, and the conclusions drawn.

Transcription practices can be thought of in terms of a continuum with two dominant modes: naturalism, in which every utterance is transcribed in as much detail as possible, and denaturalism, in which idiosyncratic elements of speech (e.g., stutters, pauses, nonverbals, involuntary vocalizations) are removed. These two positions correspond to certain views about the representation of language. With a naturalized approach, language represents the *real* world. Therefore, the transcript reflects a verbatim depiction of speech (Schegloff 1997). Denaturalized transcripts, however, suggest that within speech are meanings and perceptions that construct our reality (Cameron 2001). These are but the bookends in a larger practice of transcription. Between these two methods are endless variations using elements of each to

achieve certain analytical objectives and research goals. In this paper, it is not our purpose to pass judgment on the relative superiority of naturalism or denaturalism or to suggest that researchers must choose between the two. Both methods, and the many permutations of each, can be relevant to specific research questions. Rather, in this paper we illustrate the constraints and opportunities that different transcription styles can have on research outcomes and research participants. What is in question is the researcher's decision-making. In the haste to begin data analysis, it can be easy to use a transcription style that fails to match one's research objectives or concerns over participant confidentiality. In this paper, we advocate an intermediate step: a period of reflection that allows researchers to contemplate transcription choices and assess how these choices affect both participants and the goals of research. We do this in reference to our own experiences in a highly-sensitive, public health project.

Setting the Transcription Context

In this article, the constraints and opportunities of transcription practices are discussed with reference to on-going research examining the disclosure decisions of HIV-positive men who have sex with men (MSM). The qualitative portion of this mixed-methods study consisted of semi-structured interviews and focus groups with 57 participants about their disclosure practices to casual sexual partners. Data collected as part of this study were highly sensitive given disclosure legislation in many states. For example, in 2000, the Ohio legislature passed legislation stipulating that no person with knowledge that he or she has tested positive for HIV/AIDS can engage in sexual conduct with another person without disclosing their status prior to sex nor engage in sexual conduct with a person the offender knows or believes lacks the mental capacity to understand the significance of the offender's infection (Ohio HB. 100 2000). Violators can be prosecuted on a second degree felony assault charge and imprisoned up to eight years.

Participant recruitment moved quickly due to tight deadlines. We hired transcribers who were trained to replicate the taped interview, noting pauses, overlapping talk, incomprehensible speech and response/non-response tokens (e.g., *Uh huh, Mm, Yeah*, etc.). A community-based research team, including research scientists, HIV educators, consumers and graduate students, was formed to provide the first level of coding. When the initial transcripts were given to the research team, we faced an unexpected dilemma. While requesting accurate transcripts, the result was data that often exposed participants' identities (i.e., as African-American, immigrant, Appalachian). This created two obvious problems. First it endangered participant confidentiality, particularly when combined with other sensitive information revealed in the interview. Second, knowledge about ethnic/class identity permitted the committee to make assumptions about the participant that were not conducive to collaborative data analysis. Given the legal environment that surrounded our participants, we became keenly aware of the need to take extraordinary measures to carefully represent participants' stories. One of the ways we worked to accomplish this was through reflection on our methods. Upon reflection, we soon came to see transcription as a diverse practice with often competing objectives.

Transcription in Practice

Qualitative research often includes some form of transcription. This is not a new trend. The early ethnographies that took place in the South Pacific and Western America attempted to represent human and natural environments in field notes. Yet, even these early ventures in sustained, academic observation were fraught with representational difficulties. In Duranti (1997: 122–123), Boas, whose participant observations occurred on American Indian reservations, wrote about these problems:

I am worrying now about the style of oratory because I do not yet know how to get it down. Anyways I have troubles with ordinary conversation. Narrative I can

understand quite well, if they talk distinctly, but many have the Indian habit of slurring over the ends of their words – whispering – and that makes it difficult.

These are questions of validity that continue to haunt qualitative researchers. While the social sciences frequently overlook transcription as an important methodological step, there has been a lively debate in linguistics to help fill this gap.

Naturalized Transcription

Naturalized transcription, where utterances are transcribed in as much detail as possible, is most often seen in conversation analysis studies. Conversation analysis can be defined as the study of “talk-in-interaction.” (Hutchby and Wooffitt 1998: 13) The analysis is of actual speech patterns *between* people, that is, conversations. Conversation analysts focus on the tools used to coordinate a conversation (e.g., turn-taking, repairs, overlapping talk, response tokens) (Edwards and Lampert 1993; Hutchby and Wooffitt 1998; Jefferson 1985; Ochs 1979). Attention is paid to describing the conversation and examining it for patterns. Researchers who are interested in the intricacies of spoken language often turn to naturalized transcription.

A naturalized view of conversation is captured both in the structure of the transcript and the representation of speech. This concerns the spatial organization of dialogue and the notation of speech, respectively (Edwards 2001; Jefferson 1985; Ochs 1979). Concerning structure, the more common, basic transcript is prepared as a dramatic script (Ochs 1979). Just as one reads a play or a page of prose, eyes move top to bottom and left to right. That which is written above and to the left occurs before that written below and to the right. Concerning data representation, an extensive, linguistic shorthand has arisen to suit the needs for verbatim depictions of speech data. (See Table 1 for a brief summary or Atkinson and Heritage (1999) for a more thorough description.)

The system provides the transcriber textual symbols to indicate, among other things, time gaps in tenths of a second (e.g., .1), drawn out syllables (e.g., jus:t) and emphasis (e.g., currently). An example of this, drawn from an interview with an African-American participant, would be depicted as:

1. Ok (.1) so you went to (.1) the (.1) Health Department =
2. Yeh =
3. = and got tested then? Are you currently in a relationship?
4. Um (.2) not so much (.3) an'thin' (.1) at all. I jus:t casu:al

With the transcript constructed this way, the belief is that misrepresentation is lessened, as one moves more closely to actually-existing speech. Ochs (1979) maintained, however, that these transcription conventions are preferences, but also biases about the representation of speech. In response, other transcript styles have been developed to reflect actually-existing speech.

Columnar and partiture formats are common alternatives used to more accurately present speech (Edwards 1993). The first separates speakers into columns. Using the same data as before:

Speaker A	Speaker B
Ok (.1) so you went to (.1) the (.1) Health [Department] and got tested then? Are you <u>currently</u> in a relationship?	[Yeh] Um (.2) not so much (.3) an'thin' (.1) at all. I jus:t casu:al

In this case, the speaker’s turns are noted vertically along the column. In addition to notation, overlapping speech is bracketed. Edwards (1993) noted that the advantage of columnar formats

over the more common dramatic type is that it shows how conversational asymmetries exist between speakers. That is, the timing of dialogue can be preserved in the transcript, with overlapping talk and turn-taking more graphically depicted than in the basic format. Partiture is another method of attending to the timing of conversations, where the speech is presented horizontally. Ehlich (1993) described data in the partiture method as “semiotic events arrayed horizontally on a line [that] follow each other in time, while events on the same vertical axis represent simultaneous acoustic events.” (p. 129) Like the columnar form, partiture is an attempt to more naturalistically represent dialogue.

Dialogue is rarely the simple exchange of ideas, however. Talk is peppered with verbal and non-verbal signals that can change the tenor of conversations and meaning. The more common signals include overlapping speech, laughter, stuttering and response/non-response tokens (e.g., *Yeah, Uh huh, Mm*, etc.). These can be difficult to interpret and present the transcriber with difficult, representational decisions (Bucholtz 2000). On one hand, such signals can set the tone of a conversation and/or offer insight into the participant’s affect (Schegloff 1997). On the other hand, signals could have no bearing on the content of the interview at all, and instead obfuscate the participants’ meanings, misleading the analyst. Fundamentally, this is a question of validity and representation. That is, how does the transcriber represent the non-verbal or non-intelligible? For example, in our work with HIV-positive men, the research team read a transcript where the participant’s statement was continually interrupted by his sniffing, indicated in the transcript by ((*sniff*)). When the team met to discuss this transcript, the sniffing became confusing and the subject of some debate. Some thought the participant was crying during the interview, whereas others made assumptions about drug use. The confusion was settled when the interviewer explained that the participant was sick and his nose was running. Confusion such as this occupies anxieties over proper representation and data validity (Borland 1991). For some, especially conversation analysts, this has meant increased attention to naturalized transcription, arguing that one can decrease this confusion by focusing on the details of the conversation (Billig 1999a). For others, it has meant a move towards denaturalism.

Denaturalized Transcription

Denaturalized transcription grows out of an interest in the informational content (MacLean *et al* 2004) of speech and dissatisfaction with the empiricism of naturalized work (Billig 1999a, 1999b). A denaturalized approach to transcription also attempts a verbatim depiction of speech. Yet while still working for a “full and faithful transcription” (Cameron 1996: 33), denaturalism has less to do with depicting accents or involuntary vocalization. Rather, accuracy concerns the substance of the interview, that is, the meanings and perceptions created and shared during a conversation. This approach has found particular relevance in variants of ethnography (Agar 1996; Carspecken 1996), grounded theory (Charmaz 2000) and critical discourse analysis (Fairclough 1993; van Dijk 1999). We address denaturalized transcription in reference to the latter two.

Critical discourse analysis is a mode of inquiry used to uncover the maneuverings of power. As the *critical* adjective suggests, its philosophical roots are in the Frankfurt School of critical sociology (Crotty 1998). Whereas the conversation analyst wants to learn about talk, the critical discourse analyst wants to learn what this talk says about other aspects of the participant’s life (Cameron 2001). The focus of critical discourse analysis is on the “ideological dimension” of speech, that is, the embodied discourses (Cameron 2001: 123). Interviews, and then the transcripts, are methodological tools used to capture these discourses.

Critical discourse analysts often turn to Foucault’s large body of work for theoretical support. Of particular interest to Foucault ([1972] 1982, [1979] 1995) was the extent to which discourses permeated society. Discourses were presented as ubiquitous, and they structured how we understood reality. For Foucault ([1972] 1982), the object of social analysis was to uncover

these powerful discourses. He described this approach as archeological, to note an uncovering of discourse in everyday practices (e.g., sexual practices, mental health care, schooling). However, as Fairclough (1993) noted, researchers cannot turn to Foucault for help with transcription. Foucault never addressed the point or employed transcription in any of his work. Accordingly, a Foucauldian approach to discourse analysis can be criticized for being abstracted from real contexts of practice (e.g., interviews, observation). Fairclough (1993), accordingly, has suggested that if researchers want to examine *real* practices (e.g., of power), they must analyze *real* texts. In that the understandings of power are often captured during interviews, one can collect this information in the transcript.

While it is difficult to find detailed guidance about the uses and misuses of transcription in critical discourse analysis, one can turn to actual critical discourse analyses to examine the method in use. For example, in looking at Fairclough's transcription style, we see a dramatic format, devoid of notation other than for overlapping speech. Describing his transcription approach, he said it is "a fairly minimal type of transcription, which is adequate for many purposes. No system could conceivably show everything, and it is always a matter of judgment, given the nature of research questions, what sort of features to show and in how much detail." (p. 229) Fairclough, therefore, emphasized that researchers reflect on the purposes of the research. For Fairclough, the purpose was an analysis power. In that the maneuverings of power are often captured in the content of the interview rather than in the mechanics of the conversation, denaturalized transcription is typically the chosen method.

This portrayal of denaturalized approaches is not meant to suggest that if one chooses a naturalized approach, critical analyses are not possible. Recently, feminist and critical conversation analysts have focused on how power is implicated in the mechanics of speech. For example, Kitzinger and Frith (1999) used a feminist approach to conversation analysis to uncover the manner in which women refused unwanted sexual overtures. However, the important distinction between this and critical discourse analysis is that the focus of conversation analysis is *how* these ideas are conveyed in dialogue rather than the ideas themselves. This is a difference in research objectives – an interest in meaning or mechanics. As expressed throughout this paper, methods should reflect research questions. Therefore, if a researcher is interested in how speech is used to negotiate rape prevention, then a critical conversation analysis would be useful in addressing this interest. If, on the other hand, a researcher is interested in the meanings and perceptions attached to rape or rape prevention, it is likely that grounded theory, critical discourse analysis or one of the many variants of ethnography would be more useful.

Similar to critical discourse analysts, grounded theorists also employ a more denaturalized transcription style. Charmaz (2000: 509) defined grounded theory methods as "systematic guidelines for collecting and analyzing data to build middle-range theoretical frameworks that explain the collected data." That is, the researcher constructs a theory of the phenomenon being studied that is rooted in the information shared during interviews, observations and focus groups (Glaser and Strauss 1967). The grounded theorist goes into data collection with an interest in meanings and perceptions.

As in critical discourse analysis, effort must be expended to find useful guidance about transcription in grounded theory. The researcher interested in approaches to transcription is less likely to find the extended discussions common to conversation analysis. In grounded theory research, discussions of transcription tend to occupy terse sections of manuscripts. Nevertheless, it is possible to piece together a sustained argument for denaturalized transcription by examining the actual practice of grounded theory. As in critical discourse analysis, the purpose of grounded theory is to get at *emic* points-of-view, or insider meanings, that are attached to social phenomena. The focus is less how one communicates perceptions

(although this can be useful in capturing meanings, cf. Mehan 1999), but the perceptions themselves. For example, in MacLeod's study, (1995), he worked to express the perceptions attached to the life aspirations of poor, urban youth. Using key quotes from his participants, MacLeod revealed to readers the complex meanings participants had about growing up in poverty: some feeling confident they would escape their housing project, others feeling resigned to a life of poverty. Throughout the text, key quotes were presented in a denaturalized style. While he did not explain this choice, he did explain his interest in the experience of poverty. About his methodological choice, he writes (1995: 8), "The field methods employed in this study are not unlike those. . . in which the researcher attempts to understand a culture from an insider's point of view." That is, what did it *mean* for these young men to live in poverty?

Bennstam *et al.* (2004) handled their data very similarly to MacLeod. Using a grounded theory design, they analyzed focus group data concerning perceptions of tuberculosis infection in the Congo. Despite the likelihood of very specific geo-ethnic accents (e.g., both indigenous and colonial), their data were presented in a denaturalized format. Within these data were rich details about what it meant to contract TB, particularly the stigmatization and isolation associated with the disease. For MacLeod (1995) and Bennstam *et al.* (2004), this had less to do with the mechanics of speech and more to do with the content of the interview. Therefore, the extensive detail of the naturalized transcript, replete with involuntary vocalizations and geo-ethnic accents, was missing from their account of these ethnically diverse participants.

Constraints and Opportunities in Transcription

Transcription choices reflect both explicit and implicit assumptions. In naturalized transcription, it can be argued that the analyst is presented with speech as it is spoken by the participant rather than overly-filtered through the transcriber. Schegloff (1997) states that when we attempt to stay true to the actual speech, we privilege participants' words and avoid *a priori* assumptions. This is done, he wrote, "because it is the orientations, meanings, interpretations, understandings, etc. of the *participants*... it is *those* characterizations which are privileged in the *constitution of social-interactive reality*, and therefore have a prima facie claim to being privileged." (Schegloff 1997: 166–167, emphasis in original) The focus is on presenting data in its natural environment, that is, objectively and precisely. Only after this, according to Schegloff (1997), was it appropriate to apply theoretical filters. To do this before valid data collection is to commit, according to Schegloff (1997), a kind of theoretical imperialism... a kind of hegemony of the intellectuals... whose theoretical apparatus gets to stipulate the terms by reference to which the world is to be understood – when there has already *been* a set of terms by reference to which the world was understood – by those... involved in its very coming to pass." (p. 167, emphasis in original) Schegloff (1997) suggested researchers ask "to whom do the words in a transcript belong?" By keeping the transcript in its natural state, he argued, the participants are allowed to speak for themselves.

Naturalism is not without its critics. Of the more vocal are those in critical discourse analysis (Billig 1999a, 1999b; Fairclough 1993), who question efforts to ensure an unbiased depiction of speech. Conversation analysis and naturalized transcription, it is argued, are rooted in a naive realism that accepts empirical realities unproblematically (Guba and Lincoln 1994). Critical discourse analysts suggest that this ignores the influence of society and ideology (Billig 1999a, 1999b; van Dijk 1999). Ignoring this, some argue, could work to misrepresent participants, their stories, and therefore the rigor of the interpretations made from the transcript (Jaffe and Walton 2000; Preston 1982).

The effect naturalism can have on our understanding of the social context of speech and the speaker can be problematic. By transcribing a taped interview naturalistically, assumptions can be made about what is standard and what is non-standard. Preston (1982) described the

tendency to represent non-standard English as “linguacentric” (p. 306), respelling the speech of African-Americans and southerners. Preston (1982) said this practice gave dialects a “shock folk status” (p. 306); accents were something exotic, if not collectible (Wolfram and Schilling-Estes 1997). A hierarchy is implied with standard, American English placed above those that deviate from this norm.

Jaffe and Walton (2000) further noted that when these non-standard orthographies are read, they often denote race and class that can then be attached to prejudiced assumptions and analyses. As we mentioned earlier, this happened in our own research when committee members began to associate ethnic and class identities with certain social characteristics (e.g., internalized homophobia and lack of HIV/AIDS awareness). For example, a 43-year-old African-American man spoke of how his disclosure decisions are often based on whether his partners believe he has sex with women as well as men. If his male partners do not believe he has sex with women, he will often not disclose. About this he said:

I didn't want to disclose [to] them because they didn't think I had been with women. I have a daughter and a son. But with those guys, I used that as a [reason] for me to not disclose because they didn't believe I was bisexual. Out of 35 [male partners] I told about 20, 25 of them. But the rest I didn't. And that was due to them not believing that I had ever had sex with a woman.

Later in this interview, he added that his disclosure decisions are also based on his partner's “character,” which turned out to mean his perception of their heterosexuality. About this he said:

It directs me to disclose and sometimes not to disclose. It's according to their character.... If they're not flamboyant.... I would say real flaming, real fagish, because I feel that's my part. And if you're the man... that turns me off. So I won't disclose.

This was problematic for one community member on our research team. He argued that this man represented the internalized homophobia endemic to African-American communities. After this, his reading of the interview reflected disdain for the participant and impeded his ability to code the remainder of the interview in a productive way. We began to understand that knowledge of a participant's ethnicity could compromise both the integrity of our analysis and confidentiality. This led us to pause and reflect on how to remediate this problem. Rather than removing valuable members of our research team, we began to think about removing certain indicators of ethnicity, including geo-ethnic accent and basic demographic data from all the transcripts community researchers would read.

In the end, transcription presents real challenges to qualitative researchers. Both naturalized and denaturalized approaches suit the purposes of certain research questions (e.g., dialogue patterns or meaning) or frameworks (e.g., conversation analysis or grounded theory) (Lapadat and Lindsay 1999; Ochs 1979). While many researchers may be less likely to practice either pure naturalism or denaturalism, opting for something that borrows liberally from each, there are, nevertheless, real concerns that must be addressed in these methodological choices. In our research, we came to realize that a period of reflection was invaluable to creating trustworthy qualitative data, largely by creating safe spaces where our participants would feel free to explain sensitive parts of their lives without fear of the repercussions their words might have.

Towards Reflection in Transcription

At the heart of the debate are questions of research objectives. Conversation analysts focus on the empirical description and analysis of speech. Grounded theorists and critical discourse analysts, on the other hand, are more focused on the meanings contained in a transcript. While

these sides are often placed at odds, very rarely have those embroiled in the debate discussed their transcription decisions in relation to their research questions. That is, what are we asking? And, how is what we ask addressed methodologically? Earlier in this paper we stated that one must pick the best method or set of methods that answer the question(s) being asked. This was the impetus for our reflective pause before transcription.

Reflection has gained increased popularity throughout the academy. It is, however, subject to various interpretations. Woolgar (1988) locates reflection within the wider reflexive turn in the social sciences (Bourdieu and Wacquant 1992). Reflexivity is, as Haggerty (2003: 158) writes “a performance that positions the author in relationship to the field, the act of research, writing and the production of knowledge more generally.” Woolgar (1988) goes on to suggest varieties of reflexivity that can be located on a continuum. At one end, is a radical reflexivity whereby knowledge creation is interdependent. That is, knowledge of an object becomes an act of representation filtered through an author’s or researcher’s preconceptions, experiences and bias. At the other end is a more introspective stance. Woolgar (1998: 22) writes that such introspection is “a kind of reflexivity – perhaps more accurately designated reflection – [that] entails loose injunctions to ‘think about what we are doing.’” Citing Dewey, Carter (1999: 28) further defined reflection as “an intentional endeavor to discover specific connections between something which we do and the consequences which result.” Schön (1983) provided yet another level of thinking in writing about reflection-in-action and reflection-on-action. The former referred to the ability to think-while-doing or ‘thinking-on-your-feet.’ The latter referred to the ability to think about one’s practice, after the fact, in an effort to improve, change or evaluate this practice. Schön emphasized that the two are not wholly distinct. That is, one reflects on his/her action which informs thinking-in-action. Taken together, the practitioner develops a repertoire of practices and frames of reference that help in making informed decisions. These discussions about reflection prove relevant in pointing to the processes of informed decision-making. This same impulse to pause and think about our practice emerged as we confronted obstacles to what we believed was useful data collection. Our reflection involved reconciling pitfalls of recorded speech data and its transcription with the objectives of our research.

Early into the project, several transcription-related problems were identified. Choosing a naturalized approach could provide detail that might obfuscate the substance of the interview. This could have an impact on the analysis (e.g., the sniffing participant may have been viewed differently due to a concern over illness or affectation). A denaturalized approach could result in *white-washed* data, which removed the fine-grained socio-cultural features of the data or even information that could improve the outcomes of the study (i.e., HIV disclosure intervention programs). Our team had reached a crossroad. Rather than choosing an approach and forging ahead, an intermediate step was added. We paused to reflect on our transcription methods. Although it delayed the project, this period of reflection was invaluable.

Language usage revealed itself as particularly problematic. This included (a) challenges with participant and interviewer pronunciation, (b) vocalizations and non-verbal communication, and (c) the use of irregular grammar. Each of these will be described, their challenges revealed, and options for remediation debated.

Pronunciation

How words are pronounced and then represented as text is complicated. Difficulties can occur due to participant’s and interviewer’s use of slang, language or diction. These transcription or interpretation errors can arise in different ways, both technological and human. The most obvious source of technological error emerges from hardware or software difficulties. For example, in our work we found numerous errors resulting from inadequate audio-taping hardware. Either the tape became difficult to hear or skipped during the interview¹. These

difficulties are troublesome but relatively easy to correct. Most other sources of error are human in nature. These can range from how the transcriber hears, interprets and records what he or she hears. This issue will be incorporated throughout the discussion of pronunciation issues.

Slang

All languages contain slang, lingo, idioms and euphemisms. In our work with HIV-positive MSM, slang tended to be sexual in nature. For example, a common reference to sexual positioning included referring to oneself as a “top” or “bottom,” “pitcher” or “catcher,” rather than the technical terminology of insertive or receptive partner. For the most part, common usages of slang are not problematic. However, more obscure terminology can be troublesome in that it is difficult to comprehend and may be rendered as something other than what was meant. Transcription errors can result from the transcriber’s naiveté regarding the meaning of the slang or intention of the participant for its usage. For example, when a participant said that his repertoire of sexual behaviors included “tossin’ the salad,” the transcriber was perplexed. Further along in the taped discussion, however, the interviewer probed about this term and learned this was slang used in prisons for oral-anal contact. While not intrinsically difficult to record, the meanings of such terminology frequently elude both the researcher and the transcriber. In these cases, the interviewer can request further description from the participant to ensure complete understanding and, therefore, proper transcription. In another interview, a participant used more personalized slang, referring to his anus as his “anie.” In this case, it was clear what the participant was referring to; however it does underscore that clarification is sometimes necessary.

Geo-ethnic Accent

There are three language issues that can arise when transcribing. These include navigating accents, English as a second language and Ebonics. Geo-ethnic accents can create misunderstanding and confusion in actual conversations, let alone in transcribing interviews. For example, Southern American and New England accents could require considerable effort to transcribe for those not indigenous to that region. Typically, the transcriber is left to decide whether to record the words exactly as they are pronounced or to “translate” what the participant says into standard (i.e., majority) American English (SAE). For example, a participant of Asian descent pronounced his Vs with a hard B sound (e.g., “however” became “howeber” and “river” became “riber”). Typically in these situations there are cultural or ethnic differences between the participant and transcriber. That is, a transcriber hears the interview through his/her own cultural-linguistic filters. The interviewer and the transcriber are often aware of participant intentions. However, this affords the transcriber significant interpretive and representational power that could affect analysis and results.

Ebonics, or African American Vernacular English (AAVE), is any of the nonstandard varieties of English spoken by some African-American people throughout the world (American Heritage Dictionary 2000). While the media, linguists and educators have debated whether AAVE is improper English or the markings of culture, for the qualitative researcher the issue is largely how to transcribe AAVE (Green 2002). In our study, the most common example was the use of “wif dat” rather than “with that.” Other examples included “ax” instead of “ask” and “bof uv em” instead of “both of them.”

In our study, transcription of AAVE was initially handled naturalistically, depicting it verbatim rather than in SAE. During our reflection sessions, however, the appropriateness of this strategy was debated. On the one hand, if a naturalized approach was adopted, during member check,

¹We did not use voice-activated recording during data collection. Tape skips that did occur during our recording of interviews were due to equipment failures (e.g., defective audiotapes and analog recorders).

participants could become offended that they were represented in an insensitive way. On the other hand, if we used a more denaturalized method and “cleaned up” the transcript of AAVE, valuable data might be lost. During reflection we asked ourselves if the transcript would look different if the participant was the transcriber. That is, would they write “wif dat” or “with that?” We wondered if our research was respectful. Equally, we wondered whose perspective was being honored. While Schegloff (1997) argues that naturalism always honors the participant, this assumes that the participant hears his/her voice just as the transcriber does or is comfortable when they do not. We also needed to consider the potential influence of naturalized transcription on research team members. As stated earlier, we found that when AAVE was handled naturalistically some team members made assumptions about education level and socioeconomic status of African-American participants, resulting in potentially biased data analysis. Therefore, we had twin concerns about representation and suitability to our research design.

Diction

The pronunciation and enunciation of words is typically described as diction. In qualitative research, diction concerns how interviewers and participants choose words, especially with regard to correctness, clearness or effectiveness. A common feature of diction is the dropping of the “g” behind words. Known in phonology as *consonant cluster reduction*, this describes the tendency in AAVE to drop the second consonant of a final consonant cluster (Smitherman 1977). Examples from our data included: “I don’t want to give you somethin’ I got” and “I don’t want nothin’ that you got” and “It’s like havin’ a friend... You know what I’m sayin’?”

Another complicated feature of diction is the mispronunciation of words. For example, during an interview, a participant was asked about which types of sexual activities he practiced. To this he responded, “annual” sex.

Speaker A: *So when the two of you entered the bathroom, what type of sexual activities did you engage in?*

Speaker B: *Annual.*

This type of scenario presented difficulties that we had not anticipated. We assumed that the participant meant to say “anal,” yet wondered whether to correct this on the transcript. If he was to read “annual” while member checking would he be offended or embarrassed? Neither are reactions that any researcher would be comfortable eliciting.

Involuntary Vocalizations, Response Tokens and Nonverbal Vocalizations

Vocalizations and nonverbal interactions that occur during an interview are other transcription issues to consider. Vocalizations other than speech (e.g., laughing, coughing, stuttering, etc.) and nonverbals (e.g., hand-waving, smiling, etc.) are common in most conversations. Transcribing these features of speech can add to the context of the conversation/interview, offer clarity or create erroneous assumptions. For the purpose of this paper, we have classified such vocalizations into three distinct categories: involuntary vocalizations response/non-response tokens and non-verbal interactions.

Involuntary Vocalizations

Sounds such as coughing, sneezing, burping, sniffing, laughing and crying are considered involuntary noises. Involuntary sounds that occur during an interview can be meaningful or meaningless to the analyst. In an example mentioned earlier, sometimes the inclusion of noise (e.g., sniffing) can be misleading. In that example, the belief that the participant was crying was confirmed for the transcriber when the interviewer asked, “Do you need a moment?” and then, “Would you like a tissue?” Yet, when the interviewer reviewed the transcript, he reported

that the participant actually had a cold and was not crying. In another example, a transcript captured a participant who laughed a great deal during the interview. Again, only the interviewer was able to explain to the transcriber that the participant was extremely nervous and that this was a nervous laugh. Training interviewers to give cues as to what is happening during the silence is helpful to the transcriber and analyst.

Response Tokens

Like involuntary vocalizations, there are other parts of speech that, while not quite words, are nevertheless language. Using certain mono- or bi-syllabic sounds, can relay both meaning and understanding to the interlocutors. Among the more common of these are *Hm*, *Ok*, *Ah*, *Yeah*, *Um*, *Uh*, and *Uh huh/Nuh uh*. Unlike involuntary noises, these vocalizations are intentional. There is meaning attached to them that can influence a conversation. Nevertheless, such vocalizations are often neglected as either inconsequential or extraneous. Research has shown, however, that such vocalizations can provide a great deal of insight into both the nature of conversation (i.e., how one converses), but also the informational content of the conversation (Gardner 2001). Among those working in the ethnomethodological tradition of conversation analysis (Heritage 1984; Jefferson 1984; Sacks 1992), these features of speech have been called response tokens. Gardner writes (2001: 3), “Response tokens are difficult to describe, as they lack meaning in the conventional dictionary sense of the word.”

Nevertheless, tokens can capture meaning and emotion. Gardner (2001) offers researchers a typology of response tokens and an indication of their use and intent. Among the most common are three. First, continuers such as *Mm hm*, or *Uh huh* are used to note agreement with the speaker and give them back the primary role in the conversation. Second, acknowledgements, such as *Mm* and *Yeah*, work to express agreement or understanding between a speaker and a listener. Third are repairs, such as *Huh*, that ask the speaker to rephrase or repeat an idea or question. In many cases, tokens serve to add more detail and/or emotion to what the speaker is trying to express. A participant used a response token (e.g., *Nuh uh* to express “no”) in explaining his preferred sexual activities.

Speaker A: *So you don't insert into anybody at all?*

Speaker B: *Nuh uh. No. I'm considered a typical bottom. Yeah, female bottom.*

Speaker A: *Okay.*

Speaker B: *Drag queen bottom, I mean.*

Speaker A: *Okay, alright. Well some people that are...*

Speaker B: *I know, I know, I mean, Nuh uh, I just, I have had guys, just this other, like this other guy that I just met, he was talking to me and he's positive, he was talking to me about having sex like that, and I was like, “No,” and he said he likes it too, and I'm like, “Nuh uh.” It was just, it's not creepy, it's just that it's almost, like, a turn off.*

As with involuntary vocalizations, these signals can be inessential, if not distracting, for the analyst. However, it is important to review the transcript to assess their importance before removing these potentially important data from subsequent analyses. Many researchers, particularly conversation analysts, have argued that by disregarding tokens one may fail to fully grasp the intricacies of dialogue. That is, tokens such as a thoughtful *Hm* or wistful *Mm* can serve as useful markers in speech, indicating participant discomfort or other affective states (e.g., distress, happiness, pride, etc.).

Non-verbal Vocalizations

Non-verbal communication includes actions, activities and interactions of both participant and interviewer. Gesticulations such as pointing, thought checking, fidgeting, head nodding and hand gestures are included as non-verbal interactions. As with the other forms of noise, non-verbal interactions can add context and explanation, or create misunderstandings for the analyst. For example, one participant likened his penis to a gun, intimating that HIV made it dangerous. In the following excerpt, he speaks about not disclosing his serostatus to a partner before sex, deciding to do so afterwards.

I remember the first time I [did not disclose] because it bugged the shit out of my conscience. I even went back to [him] afterwards and said, "Hey you know what? This is bugging me. I'm HIV positive. I should have told you up front. I totally apologize. I understand if you hate me. I understand if you want to beat the shit out of me. I'm really, really sorry." He asked why and I said, "I was afraid you wouldn't go home with me and I wanted you that bad that night." Which, of course, he took as a total compliment. He thought that was absolutely the sweetest thing in the world, which is fine. But why play around with, (making gun gesture with hand) oh this gun is pretty! Let me just stick it in my mouth! It's one of those things, like this is really pretty (making gun gesture with hand) it looks great next to your head ((laughing)). Here!

For the researcher, the decision can be to eliminate none, some or all non-verbals from the transcript. In some instances, these data may seem irrelevant and not worthy of including in the transcript. However, it could be argued that non-verbals, as with tokens, are as valuable as verbiage to achieving a deep understanding of the content of a conversation. One advantage of removing non-verbals and tokens is that transcripts become easier to read (MacLean *et al* 2004). These features of speech can be distracting and make reading and following conversation threads more difficult. However, if non-verbals are removed, there is a risk of missing important conversational cues. The reader or analyst risks missing interviewer's gestures of compassion, or participant's movements of discomfort. That is, the rich detail of qualitative data could be lost if the transcript is purged of all non-verbals or tokens. One solution is to have the interviewer function as the transcriber, verifier or analyst. This allows for the inclusion of relevant speech data or the clarification of confusing noises in the transcript which could reduce misinterpretation.

Grammar

A final language concern is the improper use of SAE. During interviews, it is likely that both interviewer and participant will make grammatical errors. The most common grammatical error we encountered was the use of "ain't." Transcribing grammatical errors verbatim is a likely protocol, however we found a more common problem with grammatical errors. For example, in the following excerpt, both the interviewer (Speaker A) and the participant (Speaker B) use incorrect grammar.

Speaker A: *Were both of you laying on the couch?*

Speaker B: *Yeah, we were both laying on the couch.*

When the interviewer is either the transcriber or the double checker, the tendency was to recognize the grammatical error and want to change "laying" to "lying" in *their* quote without disturbing the participant's quote. Thus, the problem arises when corrections of grammatical errors are made for the interviewer but not for the participant.

Recommendations

As argued, a period of reflection is useful in addressing important transcription issues. This time affords researchers the ability to deliberate over transcription practices and how it affects participants and the goals of research. In relating these issues to research outcomes, it may be necessary to assess the constraints and opportunities of naturalized or denaturalized transcription. This concerns the nature of the research question and what is being sought in the data. In our project, we were interested in both contributing to the knowledge about disclosure practices but also in developing an HIV disclosure intervention. Therefore, issues of the meanings and perceptions attached to disclosure were important to us, less so the mechanics of our interview. This distinction was central to our decision to transcribe more denaturalistically.

Sensitivity to participants and the nature of their involvement with the research is also important to consider. In our project, we were aware that participants might be involved in member checking and our transcription decisions would be quickly apparent. Knowing this, it is important that researchers make decisions in a manner that shows respect for participants' words and intentions (Tilley 1998). For participants engaging in member checking, naturalized transcription could be seen as disrespectful if the participant would have written the words differently or perceived their grammar more accurately than portrayed in naturalized text.

That being said, there are merits to retaining much of the conversational mechanics captured in a naturalistic transcript. As discussed, conversation analysis provides a wealth of information that could add rich detail to the data. In that a niche of qualitative inquiry is the depth of analysis that statistical indicators cannot provide, it seems counterintuitive to remove the very details that qualitative inquiry is known and appreciated for. The pronunciation, non-verbals and irregular grammar that are parts of everyday speech can offer important insights into a participant's life and meaning-making that could add richness that would otherwise be lost. For this reason, some qualitative researchers have advocated retaining two versions of the transcript.² The first of these would be a naturalized version, containing the many details common to conversation analysis. This copy could serve as a reference copy that the researcher could turn to if in-depth analysis of the conversation (i.e., accents, communication style and speech idiosyncrasies) needed to be examined. The second of these would be a denaturalized version. This transcript could be used both in member-checking (i.e., supplied to the participant) but also for different types of analyses. That is, if the researcher was not interested in the specifics of communication (e.g., repairs, response/non-response tokens, accent, etc.) but rather the informational content, then she/he could turn to this transcript.

Because large studies require numerous transcribers, transcription decisions have to be easily standardized. Transcribers can range from undergraduate volunteers to paid professionals, thus researchers are encouraged to consider a codebook which would aid in the equivalence of the transcription process (Tilley 1998). For example, the codebook section for non-verbals might include references to the omission of sneezing but the inclusion of non-verbals related to affect such as ((*crying*)) or ((*laughing*)). Codebooks should be reviewed and updated as necessary.

Finally, transcription decisions should filter back to interviewers who might have to be retrained. As noted earlier, sexual slang was commonplace in our interviews. During reflection we acknowledged that such terminology would be important for later incorporation into an intervention. Therefore, this information was both necessary and desired. Directing interviewers to have participants define their slang could help alleviate many interpretive

²We thank Rosalie Aroni and an anonymous reviewer for this suggestion.

problems. This would reduce misunderstandings and offer participants the opportunity to clarify and provide their own meanings.

Conclusion

Transcription is a powerful act of representation. This representation can affect how data are conceptualized. Instead of being viewed as a behind-the-scenes task, we argue that the transcription process be incorporated more intimately into qualitative research designs and methodologies. Periods of reflection at crucial design and implementation points may provide a valuable exercise in honoring both the research process and participant's voice.

References

- Agar, Michael. 1996. *The Professional Stranger: An Informal Introduction to Ethnography*. Academic Press. American Heritage Dictionary. 2000. Houghton Mifflin.
- Atkinson, J. Maxwell, and John Heritage. 1999. "Jefferson's transcript notation." Pp. 158–166. *The Discourse Reader*. Adam Jaworski and Nikolas Coupland, editors. Routledge.
- Bennstam, Agneta; Margaretha, Strandmark; Vinod, Diwan. "Perception of Tuberculosis in the Democratic Republic of Congo: *Wali Ya Nkumu* in the Mai Ndombe District." *Qualitative Health Research* 2004;14(3):299–312. [PubMed: 15011899]
- Billig, Michael. "Whose terms? Whose ordinariness? Rhetoric Ideology in Conversation Analysis." *Discourse & Society* 1999a;10:543–558.
- . 1999b "Conversation Analysis and the Claims of Naivety." *Discourse & Society* 10:572–576
- Borland, Katherine. 1991. "That's Not What I Said: Interpretive Conflict in Oral Narrative Research." Pp. 63–75. *Women's Words: The Feminist Practice of Oral History*. Sherna Gluck and Daphne Patai, editors. Routledge.
- Bourdieu, Pierre, and Loic Wacquant. 1992. *An Invitation to Reflexive Sociology*. University of Chicago Press.
- Bucholtz, Mary. "The Politics of Transcription." *Journal of Pragmatics* 2000;32:1439–1465.
- Cameron, Deborah. 2001. *Working With Spoken Discourse*. Sage.
- Carspecken, Paul. 1996. *Critical Ethnography in Educational Research: A Theoretical and Practical Guide*. Routledge.
- Carter, Mary. 1999. *A Profile of Service-Learning Programs in South Carolina and their Responsiveness to the National Priorities*. Bell & Howell Company.
- Charmaz, Kathy. 2000. "Grounded Theory: Objectivist and Constructivist Methods." Pp. 509–536. *Handbook of Qualitative Research*. Norman Denzin and Yvonna Lincoln, editors. Sage.
- Crotty, Michael. 1998. *The Foundations of Social Research: Meaning and Perspective in the Research Process*. Sage.
- Duranti, Alessandro. 1997. *Linguistic Anthropology*. Cambridge University Press.
- Edwards, Jane. 1993. "Survey of Electronic Corpora and Related Resources for Language Researchers." Pp. 263–300. *Talking Data: Transcription and Coding in Discourse Research*. Jane Edwards and Martin Lampert, editors. Lawrence Erlbaum Associates.
- . 2001. "The Transcription of Discourse." Pp. 321–348. *The Handbook of Discourse Analysis*. D. Schriften, D. Tannen and E. Hamilton, editors. Sage.
- Edwards, Jane, and Martin Lampert, editors. 1993. *Talking Data: Transcription and Coding in Discourse Research*. Lawrence Erlbaum Associates.
- Ehlich, Konrad. 1993. "HIAT: A Transcription System for Discourse Data." Pp. 123–148. *Talking Data: Transcription and Coding in Discourse Research*. Jane Edwards and Martin Lampert, editors. Lawrence Erlbaum Associates.
- Eraut, Micahel. 1994. *Developing Professional Knowledge and Competence*. Falmer.
- Fairclough, Norman. 1993. *Discourse and Social Change*. Polity Press.
- Foucault, Michel. [1972] 1982. *The Archaeology of Knowledge* (Alan Sheridan, Trans.). Pantheon Books.

- . [1979] 1995. *Discipline and Punish: The Birth of the Prison* (Alan Sheridan, Trans.). Vintage.
- Gardner, Rod. 2001. *When Listeners Talk: Response Tokens and Listener Stance*. John Benjamins Publishing Company.
- Glaser, Barney, and Anselm Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine.
- Green, Lisa. 2002. *African American English: A Linguistic Introduction*. Cambridge University Press.
- Guba, Egon, and Yvonna Lincoln. 1994. "Competing Paradigms in Qualitative Research." Pp. 105–117. *Handbook of Qualitative Research*. Norman Denzin and Yvonna Lincoln, editors. Sage.
- Haggerty, Kevin. "Review essay: Ruminations on reflexivity." *Current Sociology* 2003;51(2):153–162.
- Heritage, John. 1984. "A change-of-state token and aspects of its sequential placement." Pp. 299–347. *Structures of Social Action*. Maxwell Atkinson and John Heritage, editors. Cambridge University Press.
- Hutchby, Ian, and Robin Wooffitt. 1998. *Conversation Analysis: Principles, Practices and Applications*. Polity Press.
- Jaffe, Alexandra; Shana, Walton. "The Voices People Read: Orthography and the Representation of Non-standard Speech." *Journal of Sociolinguistics* 2000;4:561–587.
- Jefferson, Gail. "Notes on a Systematic Deployment of the Acknowledgement Tokens 'Yeah' and 'Mm hm'." *Papers in Linguistics* 1984;17(2):197–216.
- . 1985. "An Exercise in the Transcription and Analysis of Laughter." Pp. 25–34. *Handbook of Discourse Analysis: Volume 3 Discourse and Dialogue*. Teun van Dijk, editor. Academic Press.
- Kitzinger, Celia; Hannah, Frith. "Just Say No? The Use of Conversation Analysis in Developing a Feminist Perspective on Sexual Refusal." *Discourse & Society* 1999;10(3):293–316.
- Lapadat, Judith; Anne, Lindsay. "Transcription in Research and Practice: From Standardization of Technique to Interpretive Positioning." *Qualitative Inquiry* 1999;5:64–86.
- MacLean, Lynne; Mechthild, Meyer; Alma, Estable. "Improving Accuracy of Transcripts in Qualitative Research." *Qualitative Health Research* 2004;14(1):113–123. [PubMed: 14725179]
- MacLeod, Jay. 1995. *Ain't No Makin' It: Aspirations and Attainment in a Low-Income Neighborhood*. Westview.
- Mehan, Hugh. 1999. "Oracular Reasoning in a Psychiatric Exam." Pp. 559–575. *The Discourse Reader*. Adam Kaworski and Nikolas Coupland, editor. Routledge.
- Mishler, Elliot. 1984. *The Discourse of Medicine: Dialectics of Medical Interviews*. Ablex.
- Ochs, Elinor. 1979. "Transcription as Theory." Pp. 43–72. *Developmental Pragmatics*. Elinor Ochs and Bambi Schieffelin, editors. Academic Press.
- Ohio HB. 100, 123rd General Assembly. (2000).
- Preston, Dennis. "Ritin' Fowlower Daun 'Rong: Folklorists' Failures in Phonology." *Journal of American Folklore* 1982;95:304–326.
- Poland, Blake 2002. "Transcription quality." Pp. 629–650. *Handbook of interview research*. Jaber Gubrium and James Holstein, editors. Sage.
- Sacks, Harvey. 1992. *Lectures in conversation. Volume 1*. Blackwell.
- Sandelowski, Margarete. "Notes on transcription." *Research in Nursing and Health* 1994;7:311–34. [PubMed: 8036279]
- Schegloff, Emanuel. "Whose Text? Whose Context?." *Discourse & Society* 1997;8:165–187.
- Schön, Donald. 1983. *The Reflective Practitioner*. Basic Books.
- Smitherman, Geneva. 1977. *Talkin and Testifyin: The Language of Black America*. Houghton Mifflin.
- Tilley, Susan. "Conducting Respectful Research: A Critique of Practice." *Canadian Journal of Education* 1998;23:316–328.
- van Dijk, Teun. "Critical Discourse Analysis and Conversation Analysis." *Discourse & Society* 1999;10:459–460.
- Wolfram, Walt, and Natalie Schilling-Estes. 1997. *Hoi Toide on the Sound Soide: The Story of the Ocracoke Brogue*. University of North Carolina Press.
- Woolgar, Steve. 1988. "Reflexivity is the ethnographer of the text." Pp. 14–34. *Knowledge and Reflexivity: New Frontiers in the Sociology of Knowledge*. Steve Woolgar, editor. Sage.

Acknowledgements

This work was supported by a grant from the National Institute of Mental Health (R21 MH 067494) to the second author. The authors thank Sarah Smith, members of the Qualitative Research for the Human Sciences listserv and anonymous reviewers for their helpful suggestions. An earlier version of this paper was presented at the 10th annual Qualitative Health Research Conference, Banff, Alberta, Canada. The views expressed in this paper are solely those of the authors. Daniel G. Oliver is a PostDoctoral Research Fellow, Julianne M. Serovich is a Professor and Tina L. Mason is a Post-Doctoral Research Fellow in the Department of Human Development and Family Science at Ohio State University.

Table 1

Transcription Notation

(.)	Just noticeable pause
(.3)	Pause time in tenths of seconds
.hh	Speaker's in-breath
hh	Speaker's out-breath
:	Stretching of preceding sound or letter
ˆ	Speaker emphasis
.	Full stop or stopping fall in tone
((sniff))	Indicates a non-verbal activity
Wor-	Shows a sharp cut-off
