

Chapter 10

Transcription

To describe every Particular, and to relate the whole Conversation of the ensuing Scene, is not within my Power, unless I had forty Pens, and could, at once, write with them all together, as the Company now spoke (Fielding, 1749/1974, p. 377; cited in Meise, 1996, p. 28).

Chapter Prospectus

Chapter 10, *Transcription*, is dedicated to the production and use of written transcripts for the research analysis of spontaneous and reproductive (i.e., reading aloud) spoken discourse. More specifically, we wish herein to consider the preparation, the use, and the reproduction of transcripts, all as types of language use in their own right. Our own research regarding these various forms of language use, particularly with regard to specific problems and biases of transcribers, the question of standardization of notation systems for transcription, and the subsequent reproduction of transcripts in research publications is reviewed. The need for tailoring notation systems to specific research goals is emphasized once again.

The Transcriber as Language User

In Chapter 4, *The Written*, we have outlined a set of principles that we consider fundamental to the design of notation systems for the transcription of spoken discourse for research purposes. We now turn to the transcriber as the change agent involved in this process of transforming spoken discourse into written text. Transcribing is thus to be considered a type of language use on the part of the transcriber.

The production of a transcript from recorded speech depends upon the intentions, abilities, and attention of the transcriber. He or she can produce a transcript that is in accord with the utterances spoken in a given corpus or a transcript in which – deliberately or involuntarily – utterances or parts of

utterances are deleted, added, substituted, and/or relocated. Since these decisions are not always a matter of error, we have chosen to speak of changes (specifically, *deletions*, *additions*, *substitutions*, and *relocations*) rather than errors on the part of the transcriber. It should be noted that, in our own research, we have concentrated on the transcription of verbal and temporal components of utterances; other prosodic and nonverbal components have not been taken into account. Changes are frequently incorporated into transcripts deliberately or at least out of some specific, though often implicit bias. As noted above, the influence of such biases on the part of the transcriber has led Ochs (1979, p. 71) to the assertion that “transcription is theory.”

Some Transcriber Difficulties and Biases

O’Connell and Kowal (1994) have analyzed six heterogeneous corpora of spoken discourse in the German language by comparing the original audio recordings with their respective transcripts. In other words, we did not request the production of a transcript by experimental subjects as part of the research in this instance, but rather analyzed transcripts made for other purposes, on other occasions, and by other researchers. These transcripts were compared with a set of master transcripts prepared by ourselves from the original audio recordings. And since we are subject to the same limitations and biases as are all transcribers, the master transcripts were prepared as follows: Both authors listened to the spoken discourse separately. The procedure was off-line in the sense that we listened to a passage again and again until both of us were certain as to how to transcribe it. Sometimes this required that both of us eventually had to listen together to a passage before a final decision was made. In indecipherable cases, the doubtful syllables were entered into the transcript only as a parenthesis marked with a number of syllables, e.g., (4 syl).

The first challenge to be met by the transcriber is the type of spoken corpus to be transcribed. The simplest task is the preparation of a transcript of reproductive spoken discourse, i.e., of the reading aloud of a text. The baseline is obviously the text that is read aloud. O’Connell and Kowal (1994) did not include such an extreme case, in which the number of expected changes is always relatively small (although a third-grade youngster might have a huge number for a simple text). We began instead with parliamentary transcription in which a perfectly well-formed, archival transcript was the desired product. It should be added that, in this instance, we did not make the master transcripts, but used the published ones. Such spoken corpora are easy to transcribe insofar as they are typically produced by very articulate speakers in a setting in which rhetoric is very important. However, the process of transcribing can still be made quite difficult by an – either antagonistically or approvingly – intrusive audience with their interruptions and brief commentaries. At the other extreme of our heterogeneous corpora was a rapid-fire conversation (i.e., an articulation rate of 6.16

syl/s) in colloquial German engaged by two college students. This corpus also included overlapping passages, laughter, and extraneous noise. Needless to say, the former corpus should be much easier to transcribe than the latter. One very useful index for these corpora turns out to be mean number of syllables per change (syl/change). For example, if a speaker actually said, “In the *uh* four years before the *uh* reunification, several things happened,” and the transcript read “In the four years before the reunification, several things happened,” then $\text{syl/change} = 20/2 = 10$. In other words, a change was made on the average every 10 syllables. For these two corpora, respectively, the mean number of syllables per change in the original transcripts in comparison with the master transcripts was $13 < 17$. It should be noted that the lower index of syl/change actually reflects more changes than a higher index. In the present instance, the 13 syl/change reflects a higher number of changes due to the transcribers’ goal of obtaining well-formed sentences for the publication of the parliamentary record; the 17 syl/change reflects the fact that the students who transcribed their own audio recorded conversation were intent upon transcribing as accurately as possible. In this instance, our finding pinpoints the salient importance of the transcribers’ motivation and specific purpose in comparison with the complexity of the audio source to be transcribed.

The broadest general conclusion to be derived from our research on the six German corpora is that “*transcribers introduce verbal changes in corpora of spoken discourse*” (p. 139). Across the board, the numerousness of the various types of changes in these corpora involving a total of 1558 changes overall was as follows: deletions > additions > substitutions > relocations (655/1558 [42%] > 534/1558 [34%] > 282/1558 [18%] > 87/1558 [6%]). The percentage of originally spoken syllables actually transcribed varied from 82% to 100% ($M = 93\%$); the lowest percentage of transcribed syllables was also that of the transcriber with the largest percentage of deletions (71%), who indulged in the self-instruction to correct the spoken corpus by omitting erroneous German expressions and hesitations. The most common deletions across the board were *und* and *auch* > *äh* > *also* (161 > 144 > 36); the most common additions were *is(t)* > *nich(t)* > *(ei)n(e-)* > *und* and *auch* (88 > 53 > 46 > 20). All six spoken corpora included fillers (*äh*), but the transcribers whose goal was a transcript of well-formed sentences transcribed none of them. The elision *is’* was nearly always transcribed as *ist*, but *ist* was never transcribed in the elided form. Only the college students, who had been specifically instructed to produce an exact transcript, transcribed *is’* in all cases as it had been spoken. Only one corpus was transcribed in accordance with a formal system of notation; this transcript and that of the college students were the only ones without relocations. In summary, even this first project has eloquently manifested that transcription is an extraordinarily complex instance of language use that depends on many different factors, including the intention and ability of the transcribers, the speech genre, and the quality of the audio source.

The reader may note that individual changes involved for the most part short function words. The danger exists that the numerousness of these hide the sometimes quite substantive changes made in content words as well. The latter

changes were usually occasioned by characteristics of the audio source: the presence of extraneous noise, unclear pronunciation on the part of the speaker, or poor acoustic quality of the recordings. In medical, legal, and emergency settings, such changes can alter the meaning of a transcript so as to do great harm. Walker (1986, p. 209) has mentioned such a case from a court transcript in which the spoken designation “male in extremis” was changed in the transcript to “male, an extremist.” Suffice it to say that the legal consequences for a gentleman at the point of death are most likely nonexistent, but those to be exacted by the court against an extremist might well involve years of imprisonment.

Slips of the Ear

Ferber (1991) has argued that there is no way of validating most of the collections in the archival literature of slips of the tongue, insofar as they have been collected mostly from memory, without the assistance of audio recordings. Accordingly, she set out to ascertain empirically whether slips of the tongue are not really slips of the ear, i.e., “incorrect transcription” (p. 106). For example, students who hear an isolated “oth” as in “other,” nearly always transcribe “of”. Ferber found that “no slip was recorded by all four [of her] listeners” (p. 119), and she concluded that “the only way of collecting spontaneous slips would seem to be by means of tape recordings, which should be listened to repeatedly, preferably by more than one person” (p. 120): The on-line listeners “*recorded only about one-third as many slips as were detected by repeated listening, and, even so, about half the items noted as slips proved erroneous*” (p. 105). In this context, on-line refers to an uninterrupted playing of the recorded speech, whereas off-line refers to the opportunity to playback any portion of the recorded speech at will.

Taking their cue from Ferber and from Lindsay (1988), Lindsay and O’Connell (1995, p. 101) had four undergraduate volunteers transcribe an audio-taped interview of former president Ronald Reagan with Dan Rather. Their instructions were simply to transcribe the tape-recorded interview from a single playing; stopping the taped recording was allowed, but no repetition or replay. Thereafter, two of the experimental subjects repeated the transcription on-line, and the two others repeated it off-line. Lindsay and O’Connell have summarized their results as follows:

None produced a verbatim transcription, but all preserved semantic content quite well. Still, deletions were numerous, particularly of discourse markers and hesitation phenomena, both of which characterize spoken, not written discourse. Significantly more deletions in the on-line than in the off-line condition indicated the difficulty of audiotape processing without off-line replay.

The differences occasioned by an on-line vs. off-line method of transcription are clearly of considerable magnitude: The on-line method cannot be recommended as an appropriate research methodology for transcribing audio recordings.

The cumulative evidence does indeed appear to indicate that much of what had been presented as slips of the tongue really constitute slips of the ear, i.e., errors made by transcribers. Hence, Bock (1996, p. 405) has referred to the identification of slips of the tongue in the literature as “abysmal,” even though they have largely been detected by “trained listeners.”

Some Limitations of Transcripts

Brown (1995, p. 39 f.) has pinpointed the considerable loss of information about the behavior of listeners when a conversation is transcribed, because the transcript does not contain the interlocutors’ reading of the face and movements of listeners:

The very nature of transcription conventions concentrates on the speaker and what the speaker is doing while uttering, leading us readily to a view of the active speaker and a listener who is quite passive during the speaker’s turn. But collaborative conversation does not consist of a series of discrete stages, as the physical nature of the transcription suggests, with a participant either being actively on-stage or passively off-stage. From each participant’s point of view, that participant is constantly on-stage but playing different roles, which overlap and merge into each other.

Our present discussion, then, goes beyond the limitations of abilities and purposes on the part of the transcriber, and even beyond the complexity of the acoustic signal and its setting. *Transcription itself is a limited and defective device.* Even the simplest of spoken discourse involves an unlimited richness of analyzable facets. There is no notation system that is in principle capable of embracing altogether this virtually infinite richness. Abercrombie (1967, p. 114) has expressed this virtual infinity quite bluntly: “It is impossible to give a truly complete description of a segment.” Furthermore, the rote addition of elements in a transcript simply leads to a cumbersome transcript that is itself not analyzable or even legible in any practical way: The seen/read simply cannot adequately depict the spoken/heard. An extreme example of this outcome is Pike’s (1943, p. 155) 88-character description of [o].

In transcribing, more is not necessarily better. One can pick up at random a current issue of a journal in the language sciences and find there transcripts bristling with various notations: idiosyncratic orthography, diacritical marks, conventional punctuation marks used in some idiosyncratic way, multiplication of graphemes to indicate a variety of phenomena, along with a multitude of other symbols. Many such notations neither serve the user-friendly function of allowing the journal reader to process the passage intelligibly nor do they enter into any kind of analysis of the passage. In other words, they seem to be made for show; they make the presentation appear more technical, more scientific. This is not science. The most extreme example of this sort of over-transcription that we have found to date is a 356-page book by Dorval (1990) of which 40% is dedicated to transcripts and transcript notations – without any inferential

argumentation whatsoever. His appendix of 75 pages consists entirely of transcripts, with the instruction to the reader that “they should be used for illustrative purposes only” (p. 276).

Since transcripts are tools for analysis and intelligibility, not cosmetic devices, they should include only what is relevant for a given research project. Hence, the call for a standardized notation system for transcribing (e.g., Edwards’s, 1989, 1993, p. 141 ff., “field-wide standard” and MacWhinney’s, 1995, p. 1, “sharing of data”; see also Selting, Auer, Barden, Bergmann, Couper-Kuhlen, Günthner, Meier, Quasthoff, Schlobinski, & Uhmann, 1998, p. 91) must be challenged. MacWhinney (1995) has been the most explicit regarding the necessity for “a standardized system for data transcription and analysis” (p. 2). Indeed, we do need guidelines to maximize compatibility and comparability from one project to another. But for this purpose, a single, standardized notation system is neither practical nor scientifically heuristic. Sinclair (1995) has put it nicely. We do not need “parading in front of us these incomprehensible stretches of mumbo jumbo” (p. 107), but some common sense: “Avoid interfering with the plain text” (p. 109).

In summary, one might readily agree that simple phoneme/grapheme correspondence is an acceptable form of standardization in transcript notation, but the effort to standardize the entire notation system is ultimately inappropriate, even impossible. Transcribing the virtually infinite richness of even a simple spoken corpus is pie-in-the-sky science.

Reproduction of Transcripts for Research Purposes

One application of transcription research that exemplifies yet another form of language use – reproduction of transcripts for research purposes – manifests very clearly many of the problematic aspects of this domain. Specifically, excerpts of transcripts are frequently reproduced in publications subsequent to the original publication, both to contribute to further research endeavors and to instruct colleagues in the research applications of such transcripts. In both cases, the importance of accuracy is paramount. An indication of how frequently this sort of reproduction occurs can be found in Levinson (1983, pp. 284–370), where, in a single chapter on “Conversational Structure,” 124 such excerpts have been reproduced.

Discrepancies between the original and the reproduced transcript – in terms of our standard set of changes, including deletions, additions, substitutions, and relocations – are indicative that something is amiss in this application of a notation system. It was precisely the discovery of these discrepancies that led O’Connell and Kowal (2000) to a more systematic investigation of such reproductions in order to discover empirically whether the incidence of discrepancies was inordinately high. In order to assemble not only a representative corpus but one that exemplified the highest quality, we chose 10 excerpts from prominent textbooks (Duranti, 1997; Garman, 1990; Whitney, 1998), 10 excerpts from Levinson (1983), and six versions of a single German transcript from Keppler (1987, p. 291).

No reproduced excerpt that we examined was found to be without at least one change – in a feature relevant to the notation system – by comparison to the originally published excerpt. In terms of numerosness of changes across the board, the 308 changes were distributed according to the following frequencies: format > prosodic > verbal > extralinguistic > paralinguistic: (131/308 [42%] > 91/308 [30%] > 77/308 [25%] > 9/308 [3%] > 0/308 [0%]). And in terms of types of change, frequencies were distributed similarly to the distributions found for original transcription in O’Connell and Kowal (1994), except that substitutions were proportionately more frequent than additions: deletions > substitutions > additions > relocations (again, of the 308 changes, 113/308 [37%] > 111/308 [36%] > 72/308 [23%] > 12/308 [4%]). In summary terms, “the overall rate of change is 6.6 syllables per change (2032/308) across 41 comparisons” (O’Connell & Kowal, 2000, p. 247) of originally published excerpts with reproduced excerpts, i.e., some change was made roughly every seven syllables in this corpus.

At the risk of presenting even more errors of transcript reproduction through the process of printing this book, we offer the following comparison of an original excerpt of a transcript from Schegloff (1979, p. 52) and the reproduced excerpt as it appeared in Levinson (1983, p. 344):

Example 10.1

The Original Transcript

I: Hello:,
 →B: H’llo Ilse ?
 →I: Yes. Be:tty.

The Reproduced Transcript

R: Hello:,
 →C: Hello *Ilse*?
 R: Yes. *Be* :tty.

Without the inclusion of changes involving the name initials, the arrows, and the underlining in the original, there are still five changes from the original 8-syllable excerpt to the reproduced excerpt: (1) H’llo → Hello; (2) Ilse → *Ilse*; (3) Ilse ? → *Ilse*?; (4) Be → *Be*; and (5) Be:tty. → *Be* :tty. Changes (1), (2), and (4) introduce prosodic changes in the reproduced excerpt; changes (3) and (5) introduce changes in spacing in the reproduced excerpt. Change (4) requires some further explanation. The change from B to *B* is not considered a change insofar as underlining was previously the common notation for italics; however, the change from e to *e* involves a prosodically meaningful shift in the conversation-analytic transcript notation system.

Levinson has provided no explanation or justification for any of these changes. In addition, the changes in spacing have not been explained in the appendix to his Chapter 6 (p. 369 f.) where the details of the notation system are listed. The lack of commentary seems paradoxical in view of Levinson’s own claim that in conversation-analytic research “heavy reliance inevitably comes to be placed on transcriptions” (p. 295).

Clayman and Heritage’s (2002) reproductions of excerpts from transcripts constitute a special case: In this instance, the authors have reproduced their own original excerpts in the same volume. O’Connell and Kowal (2006b, p. 160) have summarized the evidence regarding Clayman and Heritage’s reproductions:

Overall, of the 55 identical or partially overlapping excerpts used by Clayman and Heritage for empirical argumentation, 31 (56.4%) involved one or more (sometimes numerous) erroneous changes.

O’Connell and Kowal have suggested that the same excerpt had perhaps been transcribed in these instances by different assistants without any effort to compare the variant versions.

Conversation-analytic researchers have insisted that “the transcript plays a central role in research on spoken discourse” (Edwards, 1993, p. 3; see also Psathas & Anderson, 1990, p. 76 f.), but our empirical analyses have indicated that both the validity and the reliability of reproduced transcripts may be quite low.

The Diagnosis

It is hardly overdrawn to refer to the high rate of changes in reproduced excerpts of transcripts as disconcerting. The usefulness of such defective reproductions is thereby considerably reduced. Kitzinger (1998), who carried out similar research, has ascribed the phenomenon to simple carelessness on the part of the scholars in question. We find this diagnosis too harsh for a number of reasons. The materials themselves constitute a formidable challenge. They are dense, unfamiliar, and remote; their reproduction is a task that violates many of the language habits and expectations of a native speaker who is dealing with the written reproduction of an already published transcript. For example, the presence of the item *stors* in a transcript excerpt could be a misspelling of *stores* or a correctly transcribed mispronunciation (characteristic of the St. Louis, MO region) of *stars*. Many such minute instances add up to a complexity that overloads the human processor. But the identification of the specific human processor responsible for defective reproductions of transcripts is extraordinarily difficult because any given instance of a reproduced transcript goes through a very complex series of stages: A scholar prepares a manuscript, which is then typeset, edited, proofread, and finally printed. Where in this sequence the changes are inserted is itself an empirical question. We ourselves have found in the process of publishing journal articles that page proofs not infrequently contain many changes in excerpts from transcripts. And in those instances in which authors do not receive page proofs, there is no recourse short of the subsequent publication of an erratum in a later issue of the journal in question. In other words, the problem should be acknowledged as an important and real one that is not entirely traceable. Accordingly, extreme caution is needed in the use of reproduced transcripts.

Currently, there are no notation systems for the transcription of spoken discourse that are truly user-friendly and efficient. Schenkein’s (1978, p. xi) goal of producing “a reader’s transcript – one that will look to the eye how it sounds to the ear,” has not been realized in the intervening three decades, simply because it is not possible. The habits associated with the learning of our native

language do not include the reading of complexly notated transcripts. The evidence presented by O'Connell and Kowal (2000, p. 266) seriously challenges "the practical usability of current notation systems" in research publications. Their suggestion warrants the reader's attention. It is that:

Henceforth researchers transcribe spoken discourse with *only* those notations which are to be used for analyses in keeping with the purposes of the research. The resulting transcripts will be less dense and hence easier to reproduce – and an appropriate level of parsimony will be preserved.

In summary, our research on the preparation of transcripts of spontaneous spoken discourse has shown that it is a very complex type of language use: The transcriber's own rhetorical habits, his or her intentions as regard the specific task of transcribing, immersion in the dialogical, and ability to listen carefully all influence the product in important ways. The same complexity applies as well to the reproducer and the reader of transcript excerpts.