9

From Speech to Text

Before turning to the analysis of the knowledge constructed in the interview interaction, I will address the transcription of interviews. Rather than being a simple clerical task, transcription is itself an interpretative process. Whereas the interaction of the interview situation has been extensively treated in the literature on method, the translation from oral conversations to written texts has received less attention. This chapter addresses the procedures for making interview conversations accessible to analysis—taping the oral interview interaction, transcribing the tapes into written texts, and the use of computer programs to assist the analysis of the interviews. The practical problems of transcription raise theoretical issues about the differences between oral and written language, which leads to the rather neglected position of language in interview research.

Recording Interviews

Methods of recording interviews for documentation and later analysis include audiotape recording, videotape recording, note taking, and remembering. The usual way of recording interviews today is with a *tape recorder*. The interviewer can then concentrate on the topic and the dynamics of the interview. The words and their tone, pauses, and the like, are recorded in a permanent form that can be returned to again and again for relistening. The audiotape gives a decontextualized version of the interview, however: It does not in-

clude the visual aspects of the situation, neither the setting nor the facial and bodily expressions of the participants.

A *videotape recorder* will encompass the visual aspects of the interview. With the inclusion of facial expressions and bodily posture, a videotape provides richer contexts for interpretations than does audiotape. Video recordings offer a unique opportunity for analyzing the interpersonal interaction in an interview, an aspect that has led to extensive use of videos in research on, and training for, therapy.

The wealth of information makes videotape analysis a time-consuming process. For most interview projects, particularly those with many interviews and where the main interest is the content of what is said, video recordings may be too cumbersome for analysis. A video is useful for the training of interviewers, making them aware of their facial and bodily expressions during an interview that could either inhibit or promote communication. The same is true of subtle ways of reinforcing specific types of answers by nods, smiles, and bodily postures that the interviewer may not be aware of and that are not recorded on the audiotape.

It should be noted that the inclusion of the visual setting does not solve the issue of an objective representation of the interview situation. Researchers who use videotape recordings are today rather sensitive to the constructive natures of their documentation, which are products of the researcher's many choices of angles and framing, as well as the sequence of shots (see, e.g., Harel & Papert, 1991).

An interview may also be recorded through a reflected use of the researcher's subjectivity and *remembering,* relying on his or her empathy and memory and then writing down the main aspects of the interview after the session, sometimes assisted by notes taken during the interview. There are obvious limitations to a reliance on memory for interview analysis, such as the rapid forgetting of details and the influence of a selective memory. The interviewer's immediate memory will, however, include the visual information of the situation as well as the social atmosphere and personal interaction, which to a large extent is lost in the audiotape recording. The interviewer's active listening and remembering may ideally also work as a selective filter, retaining those very meanings that are essential for the topic and purpose of the study.

While remembering is today often decried as a subjective method replete with biases, it should not be overlooked that the main empirical basis of psychoanalytic theory came from the therapist's empathic listening to and remembering of therapeutic interviews. Freud developed his psychoanalytical theory at a time when tape recorders did not exist. He refrained from taking notes during the therapeutic hours and listened with an even-hovering attention, attended to the meaning of what was said, and first made notes after the therapeutic session (Freud, 1963). This form of recollection is based on active listening during the situation; it requires sensitivity and training, which interview researchers today may forgo, treating the tapes and transcripts as their real data. One might speculate that if tape recorders had existed in Freud's time, psychoanalytical theory might not have developed beyond infinite series of verbatim quotes from the patients, and psychoanalysis might today have remained confined to a small Viennese sect of psychoanalysts lost in a chaos of tapes and transcriptions from their therapies.

Taping. In the present context, the most common method of recording interviews today—audiotape recording and subsequent transcription—will be treated more extensively. The first requirement for transcribing a recorded interview is that it was in fact recorded. Some interviewers have painful memories of an exceptional interview where nothing got on the tape due to technical faults or, most often, human error. The interviewer may have been so caught by the newness and complexities of the interview situation that he or she simply forgot to turn the recorder on, or a special interview may have been so engaging that any thought of technicalities was lost.

A second requirement for transcription is that the conversation on the tape is audible. A good tape recorder and microphone are basic requirements. So is finding a room without background noise such as voices in neighboring rooms and heavy outside traffic. To secure good recording quality it is necessary that the microphone is close enough to both participants; that the interviewer is not afraid to ask a mumbling interviewee to speak up; and that the transcriber's coming work is kept in mind, for example by avoiding coffee cups and the like hitting the table, sending bolts of thunder into the transcriber's ears (see Yow [1994] and Poland [1995] for more extensive treatments of the recording quality of interviews).

Transcription Reliability and Validity

Interviews are today seldom analyzed directly from tape recordings. The usual procedure for analyzing is to have the taped interviews transcribed into written texts. Although this seems like an apparently simple and reasonable procedure, transcriptions involve a series of methodical and theoretical problems. For example, once the interview transcriptions are made, they tend to be regarded as *the* solid empirical data in the interview project. The transcripts are, however, not the rock-bottom data of interview research, they are artificial constructions from an oral to a written mode of communication. Every transcription from one context to another involves a series of judgments and decisions. I will introduce the constructive nature of transcripts by taking a closer look at their reliability and validity.

Reliability. Questions of interviewer reliability in interview research are frequently raised. Yet in contrast to sociolinguistic research, transcriber reliability is rarely mentioned for social science interviews. Technically regarded, it is an easy check to have two persons independently type the same passage of a taped interview, and then have a computer program list and count the number of words that differ between the two transcriptions, thus providing a quantified reliability check.

The interpretational character of transcription is evident from the two transcripts of the same tape recording in Table 9.1. The words that are different in the two transcriptions are italicized. The transcriptions were made by two psychologists who were instructed to transcribe as accurately as possible. Still, the transcribers adopted different styles: Transcriber A appears to write more verbatim, includes more words, and seems to guess more than transcriber B, who records only what is clear and distinct, and who also produces a more coherent written style. The most marked discrepancy between the two is rendering the interviewer's question as "because you don't get grades?" versus "of course you don't like grades?" It thereby becomes ambiguous what the subject's answer—"Yes, I think that's true . . ."—refers to.

The quality of transcriptions can be improved by clear instructions about the procedures and purposes of the transcriptions, preferably accompanied by a reliability check. Yet even with detailed typing

TABLE 9.1 Two Transcriptions of the Same Interview Passage

Transcription A:

I: And are you also saying *because* you don't *get* grades? *Is that true?*

S: Yes, I think that's true because if I got grades I would work toward the grade as opposed to working toward . . . *umm,* expanding what I know, or, pushing *a limit* back *in myself or, something* . . . contributing new ideas . . .

Transcription B:

I: And are you also saying *that of course* you don't *like* grades?

S: Yes, I think that's true, because if I got grades I would work toward the grade as opposed to working toward expanding what I know or pushing *those limits* back . . . (tape unclear) contributing new ideas.

instructions it may be difficult for two transcribers to reach full agreement on what was said. Listening again to the tape might show that some of the differences are due to poor recording quality and mishearing. Other differences, which are of interest from an inter-relational perspective, may not be unequivocally solved, as for example: Where does a sentence end? Where is there a pause? How long is a silence before it becomes a pause in a conversation? Does a specific pause belong to the subject or to the interviewer? And if the emotional aspects of the conversation are included, for instance "tense voice," "giggling," "nervous laughter," and so on, the intersubjective reliability of the transcription could develop into a research project of its own.

Validity. Ascertaining the validity of the interview transcripts is more complex than assuring their reliability. The issue of what a valid transcription is may be exemplified by two different transcriptions of a story told by a 7-year-old Afro-American pupil (see Table 9.2). The two transcriptions are from a segment of a longer story from a classroom exercise, transcribed by two different researchers and discussed by Mishler (1991). Transcript A is a verbatim rendering of the oral form of the story; the school teacher found the whole story disconnected and rambling, not living up to acceptable criteria of coherence and language use. Transcript B is an idealized realization of the same story passage, retranscribed into a poetic form by a researcher familiar with the linguistic practices of black oral style.

TABLE 9.2 Two Transcriptions of Leona's Story of Her Puppy

Transcription A:

. . . and then my puppy came / he was asleep / and he was—he was /
he tried to get up / and he ripped my pants / and he dropped the oatmeal—
all over him / and / my father came / and he said
. . .

Transcription B:

an' then my puppy came
   he was asleep
   he tried to get up
   an' he ripped my pants
   an' he dropped the oatmeal all over him
an' my father came
   an' he said
. . .

SOURCE: From Mishler (1991).

Here the story appears as a literary tour de force, yielding a remarkable narrative. Neither transcription is more objective than the other; they are, rather, different written constructions from the same oral passage: "Different transcripts are constructions of different worlds, each designed to fit our particular theoretical assumptions and to allow us to explore their implications" (Mishler, 1991, p. 271).

Transcribing involves translating from an oral language, with its own set of rules, to a written language with another set of rules. Transcripts are not copies or representations of some original reality, they are interpretative constructions that are useful tools for given purposes. Transcripts are decontextualized conversations, they are abstractions, as topographical maps are abstractions from the original landscape from which they are derived. Maps emphasize some aspects of the countryside and omit others, the selection of features depending on the intended use. Maps of the same topographical area for purposes of driving, aviation, agriculture, and mining will tend to be rather different. An objective map representing, for example, the island of Greenland does not exist: The shape depends on the selected mode of projection from a curved to a flat plane, which again depends on the intended use of the map.

Correspondingly, the question "What is the correct transcription?" cannot be answered—there is no true, objective transformation from the oral to the written mode. A more constructive question is: "What is a useful transcription for my research purposes?" Thus verbatim descriptions are necessary for linguistic analyses; the inclusion of pauses, repetitions, and tone of voice are relevant for psychological interpretations of, for example, level of anxiety or the meaning of denials. Transforming the conversation into a literary style facilitates communication of the meaning of the subject's stories to readers.

Oral and Written Language

By neglecting issues of transcription, the interview researcher's road to hell becomes paved with transcripts. The interview is an evolving conversation between two people. The transcriptions are frozen in time and abstracted from their base in a social interaction. The lived face-to-face conversation becomes fixated into transcripts. A transcript is a transgression, a transformation of one narrative mode—oral discourse—into another narrative mode—written discourse. To *trans*scribe means to *trans*form, to change from one form to another. Attempts at verbatim interview transcriptions produce hybrids, artificial constructs that are adequate to neither the lived oral conversation nor the formal style of written texts. Transcriptions are translations from one language to another; what is said in the hermeneutical tradition of translators also pertains to transcribers: *traduire traittori*—translators are traitors.

The different rhetorical forms of oral and written language are frequently overlooked during the transcription of social science interviews; one exception is Poland (1995). Recognizing the socially constructed nature of the transcript, he discusses in detail procedures for increasing the trustworthiness of transcripts and thus enhancing rigor in qualitative research. Sociolinguistics and ethnomethodology have brought the differences between oral and written language into focus (Ong, 1982; Tannen, 1990; Tedlock, 1983). In a historical linguistic study, in particular of Homer's work, Ong outlines the thought and expression of a primarily oral culture as being close to the human life world, situational, empathic and participatory, additive, aggregative, agonistic, and redundant. In contrast, a written

culture is characterized by analytic, abstract, and objectively distanced forms of thought and expression.

Interview transcriptions are often boring to read, ennui ensues in face of the repetitions, the incomplete sentences, and the many digressions. The apparently incoherent statements may be coherent within the context of a living conversation, with vocal intonation, facial expressions, and body language supporting, giving nuances to, or even contradicting what is said. Such discrepancies between what is said and the accompanying bodily expressions are deliberately used in some forms of comical and ironical statements.

The problems with interview transcripts are due less to the technicalities of transcription than to the inherent differences between an oral and a written mode of discourse. Transcripts are decontextualized conversations. If one accepts as a main premise of interpretation that meaning depends on context, then transcripts in isolation make an impoverished basis for interpretation. An interview takes place in a context, of which the spatial, temporal, and social dimensions are immediately given to the participants in the face-to-face conversation, but not to the out-of-context reader of the transcript. In contrast to a taped interview, a novel will report the immediate context of a conversation, including nonverbal communication to the extent the author finds it relevant for the story he or she wants to tell. Similar considerations hold for journalistic interviews.

The transcriptions are detemporalized; a living, ongoing conversation is frozen into a written text. The words of the conversation, fleeting as the steps of an improvised dance, are fixated into static written words, open to repeated public inspections. The words of the transcripts take on a solidity that was not intended in the immediate conversational context. The flow of conversation, with its open horizon of directions and meanings to be followed up, is replaced by the fixated, stable written text.

In a conversation we normally have immediate access to the meaning of what the other says. When analyzing the interviews, the tape recording, and in particular the ensuing transcript, tends to become an opaque screen between the researcher and the original situation. Attention is drawn to the formal recorded language and the empathically experienced, lived meanings of the original conversation fade away; the dried pale flowers in the herbarium replace the fresh

colorful flowers of the field. The transcripts become a kind of funda-
mental verbal data for interview research, rather than a means to evoke
and revive the personal interaction of the interview situation.

The rather interpretative basis of the transcripts is often forgotten
in the analysis, where the transcripts tend to become a rock-bottom
basis for the ensuing interpretations. Ignorance of the many technical
and theoretical issues of transforming conversations into texts may be
due to a neglect in social science of the linguistic medium of interview
research. Social scientists are today naive users of the language that
their professional practice and research rests on. Although most social
science programs today require courses in statistical analysis of quan-
titative data, even a rudimentary introduction to linguistic analysis of
linguistic, qualitative data is a rarity.

"Not being able to rely on a conception of a stable, universal,
noncontextual, and transparent relation between representation and
reality, and between language and meaning, confronts researchers
with serious and difficult theoretical and methodological problems"
(Mishler, 1991, p. 278). Neglecting linguistic complexities during
transcription from an oral to a written language may be related to a
philosophy of naive realism, with an implicit constancy hypothesis of
some real meaning nuggets remaining constant by their transfer from
one context to another. In contrast, postmodern conceptions of
knowledge emphasize the contextuality of meaning with an intrinsic
relation of meaning and form, and focus on the very ruptures of
communication, the breaks of meaning. The nuances and the differ-
ences, the transformations and discontinuities of meaning become the
very pores of knowledge. Postmodern approaches to knowledge do
not solve the many technical and theoretical issues of transcription.
The emphasis on the linguistic constitution of reality, on the contex-
tuality of meaning, and on knowledge as arising from the transitions
and breaks, however, involves a sensitivity to and a focus on the often
overlooked transcription stage of interview research.

Transcribing Interviews

Transcribing the interviews from an oral to a written mode struc-
tures the interview conversations in a form amenable for closer
analysis. Structuring the material into texts facilitates an overview and

is in itself a beginning analysis. The amount and form of transcribing
depends on such factors as the nature of the material and the purpose
of the investigation, the time and money available, and—not to be
forgotten—the availability of a reliable and patient typist. Transcrip-
tion from tape to text involves a series of technical and interpretational
issues for which, again, there are few standard rules, but rather a series
of choices to be made.

It is a useful exercise for interviewers to type one or more pilot
interviews themselves. This will sensitize them to the importance of
the acoustic quality of the recording, to paying attention to asking
clear audible questions and getting equally clear answers in the
interview situation. The transcribing experience will also make inter-
viewers aware of some of the many decisions involved in transforming
oral speech to written texts, and it will give an impression of the time
and effort the transcription of an interview requires.

Typing. The time needed to transcribe an interview will depend on
the quality of the recording, the typing experience of the transcriber,
and the demands for detail and exactitude. Transcribing large amounts
of interview material is often a tiresome and stressing job; the stress
can be reduced by securing recordings of high acoustic quality.

For the interviews in the grading study, an experienced secretary
took about 5 hours to type verbatim an interview of 1 hour. A 1-hour
interview results in 20 to 25 single-spaced pages, depending on the
amount of speech and how it is set up in typing.

Who Should Transcribe? In most studies the tapes are transcribed
by a secretary, who is likely to be more efficient at typing than the
researcher. Investigators who emphasize the modes of communication
and linguistic style may choose to do their own transcribing in order
to secure the many details relevant to their specific analysis. Some have
a typist do a first transcription of all the interviews in a study; then
after reading them through, the researcher goes back and retypes those
interviews, or those parts of the interviews, that will be subjected to
intensive analysis.

Style. There is one basic rule in transcription—state explicitly in
the report how the transcriptions were made. This should preferably
be based on written instructions to the transcribers. If there are several

transcribers for the interviews of a single study, care should be taken that they use the same procedures for typing. If this is not done, cross-comparisons among the interviews will be difficult to make.

Although there is no standard form or code for transcription of research interviews, there are some standard choices to be made. They involve such issues as: Should the statements be transcribed verbatim and word by word, including the often frequent repetitions, or should the interview be transformed into a more formal, written style? Should the entire interview be reproduced verbatim, or should the transcriber condense and summarize some of the parts that have little relevant information? Should pauses, emphases in intonation, and emotional expressions like laughter and sighing be included? And if pauses are to be included, how much detail should be indicated?

There are no correct, standard answers to such questions; the answers will depend on the intended use of the transcript. One possible guideline for editing, doing justice to the interviewees, is to imagine how they themselves would have wanted to formulate their statements in writing. The transcriber then on behalf of the subjects translates their oral style into a written form in harmony with the specific subjects' general modes of expression. The extent of detail in a transcription will depend on its use; regarding pauses, for example, it may be sufficient for some purposes simply to note "a short pause" or "a long pause," whereas for detailed sociolinguistic analyses the length of a pause will be indicated in milliseconds.

Decisions concerning style of transcription depend on the audience for which a text is intended. For the investigator, as an aid in remembering the interviews? For the interview subjects, to confirm that their views are adequately rendered in the interview and possibly also as an invitation to expand upon what they have said? For a research group that will make extensive analyses of the interviews, or for critical colleagues who want to check the basis on which the researcher draws his or her conclusions? Or for general readers who want some concrete illustrations from the interviews?

The decisions about style of transcribing depend on the use of the transcriptions. If they are to give some general impressions of the subjects' views, rephrasing and condensing of statements may be in order. Also, if the analysis is to be in a form that categorizes or condenses the general meaning of what is said, a certain amount of

editing of the transcription may be desirable. If, however, the transcriptions are to serve as material for sociolinguistic or psychological analysis, they need to be in a detailed, verbatim form. Even the many "hm"s of an ordinary conversation, disturbing when reading a transcript, can be relevant for later analysis: for example, whether the "hm"s of the interviewer selectively follow, and thus reinforce, special types of answers by the subject. And, if psychological interpretations are to be made, the emotional tone of the conversation should also be included. Here the very pauses, repetitions, and so forth may yield important material for interpretation.

In Jacobsen's (1981) study of the university socialization of students of Danish and of medicine to their respective professional cultures, the interviews were transcribed verbatim, including the many "hm"s, "ain't it true," and the like. Jacobsen counted the use of such fillers by the students of Danish and of medicine, respectively, and found a markedly more frequent use of "ain't it true" by the students of Danish. He interpreted this, together with other indications, as being in line with the culture of the humanities, in which there is an emphasis on dialogue with attempts to obtain consensual validation of interpretations, involving appeals to the others, such as "ain't it true." In contrast, the medical profession is more characterized by lectures as monologues authoritatively stating nondebatable truths.

The issue of how detailed a transcription should be is also illustrated by an interview sequence on competition for grades, which in Denmark is a negative behavior that many pupils hesitate to admit to:

Interviewer: Does it influence the relationship between the pupils that the grades are there?

Pupil: No, no—no, one does not look down on anyone who gets bad grades, that is not done. I do not believe that: well, it may be that there are some who do it, but I don't.

Interviewer: Does that mean there is no competition in the class?

Pupil: That's right. There is none.

At face value, this pupil says that one does not look down on pupils with low grades and confirms the interviewer's interpretation that there is no competition for grades in the class. A critical reading may lead to the opposite conclusion—the boy himself introduces the

phenomenon of looking down on pupils with bad grades, first denies that it occurs, then repeats the denials with three "no"s and four "not"s in the few lines of his statement. This many denials of looking down on other pupils might, with the quantitative increases, suddenly lead to a qualitative change for the reader, and the statement come to mean the opposite of what was manifestly said. If the above interview statement had not been transcribed verbatim, but rephrased into a briefer form such as "One does not look down on others with low grades nor compete for grades," the reinterpretation of the manifest meaning of the statement into its opposite could not have taken place. The effect of multiple negations canceling each other out is used in literature, in *Hamlet,* for example:

Hamlet: Madam, how like you this play?
Queen: The lady doth protest too much, methinks. (*Hamlet,* act III, scene 2)

Ethics. Transcription involves ethical issues. The interviews may treat sensitive topics in which it is important to protect the *confidentiality* of the subject and of persons and institutions mentioned in the interview. Along with the necessary and simpler but sometimes forgotten tasks goes the need for secure storage of tapes and transcripts, and of erasing the tapes when they are no longer of use. In sensitive cases, it may be advantageous as early as the transcription stage to mask the identities of the interviewed subjects, as well as events and persons in the interviews that might be easily recognized. This is particularly important if a larger research group is involved and several persons will therefore have access to the transcripts.

Some subjects may experience a shock as a *consequence* of reading their own interviews. The verbatim transcribed oral language may appear as incoherent and confused speech, even as indicating a lower level of intellectual functioning. The subjects may become offended and refuse any further cooperation and any use of what they have said. If the transcripts are to be sent back to the interviewees, rendering them in a more fluent written style might be considered from the start. And if not, consider accompanying the transcripts with information about the natural differences between oral and written language styles. Be mindful that the *publication* of incoherent and repetitive verbatim

interview transcripts may involve an unethical stigmatization of specific persons or groups of people.

Those teachers in the grading study who had expressed interest received a draft of the book chapter in which their statements were discussed. A teacher of Danish, who had been quoted extensively, called and asked me to omit or rephrase his statements in the book. The rather off-the-cuff verbatim quotes from his interview showed a very poor Danish used by a teacher of Danish, which he found penible in his profession. At that time I was little aware of the different rules for oral and written language and believed that a verbatim transcription of the interviews was the most loyal and objective transcription. I did, however, respect his request and changed his quotes into a correct written form, which also made them more readable.

Computer Tools for Interview Analysis

During the past decade, computer programs have been developed to facilitate the analysis of interview transcripts. They replace the time-demanding cut-and-paste approach to analysis of often hundreds of pages of paper with "electronic scissors." The programs are aids for structuring the interview material for further analysis; the task and responsibility of interpretation still rest with the researcher.

The computer programs serve as textbase managers, storing the often extensive interview transcripts, and allow for a multitude of analytic operations (for overviews, see Tesch, 1990; Weitzman & Miles, 1995; Miles & Huberman's, 1994, appendix gives a short introduction to choosing among computer programs for qualitative analysis). The programs allow for such operations as writing memos, writing reflections on the interviews for later analyses, coding, searching for key words, doing word counts, and making graphic displays. Some of the programs allow for on-screen coding and note taking while reading the transcripts.

The most common form of computer analysis today is coding, or categorization, of the interview statements. The researcher reads through the transcripts and categorizes the relevant passages; then with code-and-retrieve programs the coded passages can be retrieved and inspected again, with options of recoding and of combining codes.