5

Transcribing Talk-in-Interaction

CONTENTS

The unique core activity of conversation analytic work is the careful, repeated listening to (and viewing of) recorded interaction in order to make detailed transcriptions of it, using some version of a set of conventions originally developed by Gail Jefferson. This chapter will provide an extensive discussion of methodological and practical aspects of using this style of transcribing verbal interaction.

What is involved in 'transcription'

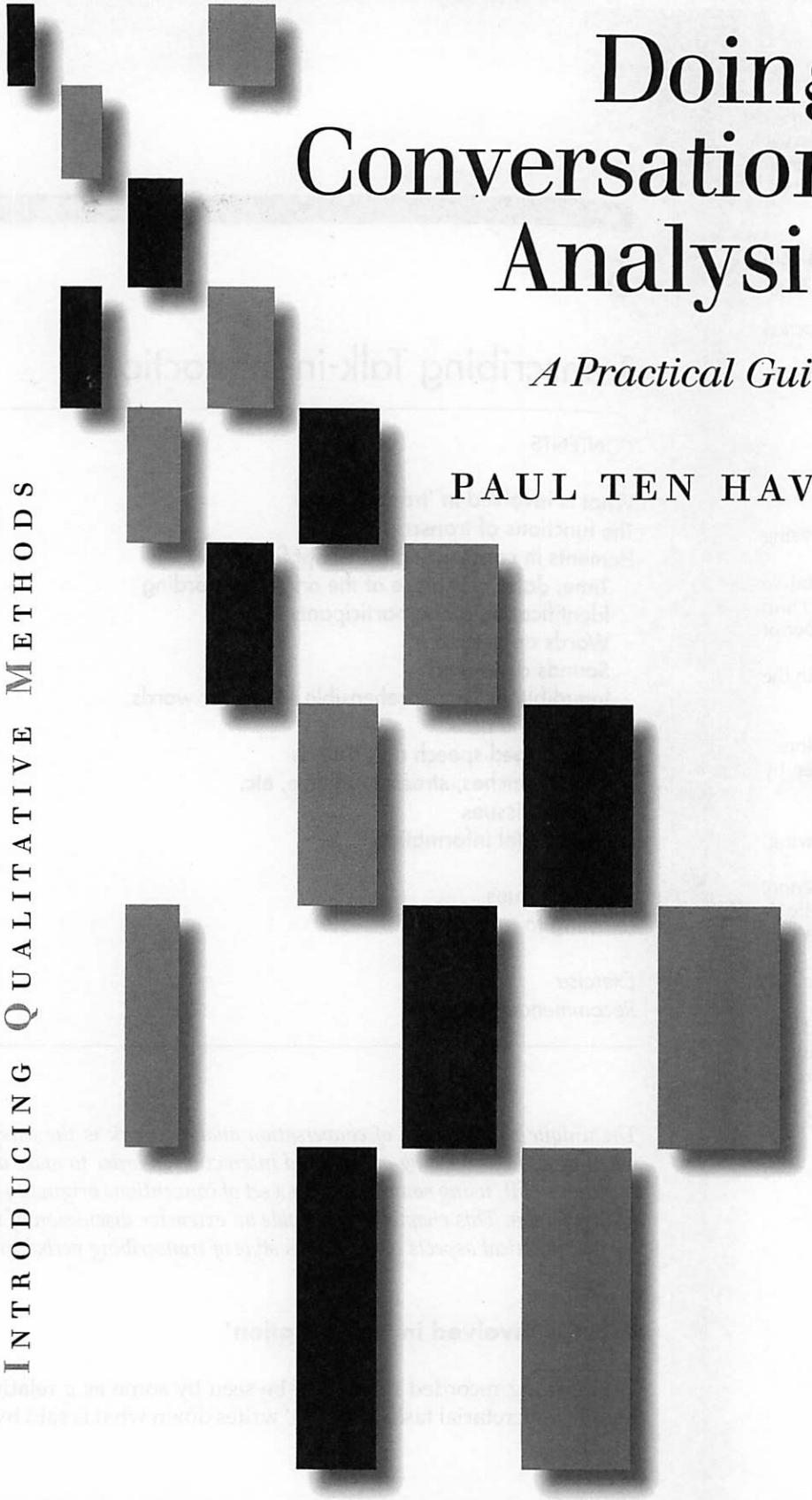
Transcribing recorded talk might be seen by some as a relatively simple matter, a secretarial task. One 'just' writes down what is said by the parties

Doing Conversation Analysis

A Practical Guide

PAUL TEN HAVE

INTRODUCING QUALITATIVE METHODS



to the interaction. In fact, this is only the starting point for a transcription adequate for a detailed analysis. But even this task may be quite difficult when the recording is not of the highest quality, when people are not articulating very well, when more than one person is talking at a time, when they are laughing, etc. A conventional secretarial transcription will tend to clean up the mess a bit, by leaving out 'noise' considered inessential, and by 'correcting' obvious mistakes. This will be the way 'verbatim protocols' of meetings or interviews will be made when one is only interested in the contents of what has been said. For analysing talk-in-interaction, however, one not only wants to write down *what* has been said, but also *how* it has been said. And that's why transcription is so important and difficult for a research tradition like CA.

For contrast, one may think of a kind of transcription that only catches the 'sounds' on the tape, as in a *phonetic* transcription, independent of the speech's meaning, or of the particular language spoken by the interactants. Such a transcript will provide a certain kind of access to *how* things were said, but *what* was said will be inaccessible to most readers. Therefore, transcription systems used in the various varieties of 'discourse analysis', largely conceived, tend to offer a practical *compromise* between the interests of faithfulness to the original, recorded sounds, and of readability of the final transcribed product. As Heritage and Atkinson write:

> [. . .] the transcripts result from and represent an attempt to get as much as possible of the actual sound and sequential positioning of talk onto the page, while at the same time making this material accessible to readers unfamiliar with systems further removed from standard orthography. (Heritage & Atkinson, 1984: 12)

In other words, a transcription might best be seen as a *translation*, made for various practical purposes, of the actually produced *speech* into a version of the standardized *language* of that particular community, with some selective indication of the actual speech production.[1]

As noted, most, if not all, transcripts used in CA, as conceived in this book, employ a more or less close variant of the transcription 'system' devised by Gail Jefferson in her work for Harvey Sacks, first, and on her own, later. Therefore, I will limit my discussion to this particular style of work. There are, of course, other systems (cf. Edwards & Lampert, 1993), but as this book is designed to help those who are beginning to work *within* the CA tradition, it is best to stay on the path that has been taken before. No transcription system is perfect, of course, since it represents the result of a series of compromises between heterogeneous considerations. It has not been 'designed' on the basis of a set of worked-out principles, but it evolved in the course of doing CA's work. It has been criticized on a number of points, including its inconsistency (O'Connell & Kowal, 1994), but as a working instrument it seems to be generally sufficient for most CA purposes. Furthermore, it can be extended with special *ad hoc*

features if required for a particular purpose (cf. Atkinson, 1984a, 1984b, for the 'transcription' of applause, for instance). The disadvantages of continuing to work with the system seem to be less than those of designing a new system that nobody else uses or knows how to read (cf. Psathas, 1995). In fact, it takes time and practice not only to learn to *make* transcriptions according to a specific system, but also to learn to *read* them. Alessandro Duranti (1997: 142) stresses 'the fact that the process of transcribing implies a process of socialization of our readers to particular transcribing needs and conventions'. Probably the best strategy is to combine the two processes of learning to make and to read transcriptions.

Although the basic system was devised by Gail Jefferson, it has become a kind of 'common language', with various dialects, so to speak. There is not *one* clearly defined, canonical way of making and formatting CA transcriptions. One can discern minor variations in the conventions that are actually used. What I will describe, then, is just another version, not *the* officially endorsed system.

The functions of transcripts

The activity of transcribing a recording can be conceived in various ways, as rendering the sounds on the tape, as describing the verbal interaction, or, as I did above, as translating 'speech' into 'language'. These conceptions have different consequences for the methodological status conferred on transcripts, as 'data', 'observations', or 'versions'.

It is often stressed that transcripts are not the 'data' of CA, but rather a convenient way to capture and present the phenomena of interest in written form. An obvious reason for using transcripts in publications is that most publication outlets until recently did not allow any other way to 'represent' the data. It is also a common experience, however, that at a first hearing/viewing of a recording, the phenomena of interest to a particular researcher are not at all obviously available. It is only after repeated listening/viewing, and quite often only after repeated efforts at transcription, that certain phenomena 'present themselves' to the ears, eyes, and minds of the tape's audience.

As various writers, including Heritage and Atkinson (1984) and Psathas and Anderson (1990), have noted, transcriptions should not be taken as a substitute for the recordings. They are selective, 'theory-laden' renderings of certain aspects of what the tape has preserved of the original interaction, produced with a particular purpose in mind, by this particular transcriptionist, with his or her special abilities and limitations. Therefore, it is generally recommended that an analyst makes his or her own transcriptions. Even if the work is tedious, and just because it is tedious, it gives one a kind of access to the 'lived reality' of the interaction that is not available in any other way. In other words, because, for making a transcription,

a researcher is forced to attend to details of the interaction that would escape the ordinary listener, transcription works as a major 'noticing device'.

> The process of transcription is an important analytical tool, providing the researcher with an understanding of, and insight into, the participants' conduct. It provides the researcher with a way of noticing, even discovering, particular events, and helps focus analytic attention on their socio-interactional organisation. (Heath & Luff, 1993: 309)

Furthermore, once made, transcripts provide the researcher with a quick access to a wide range of interactional episodes that can be inspected for comparative purposes. Transcription, therefore, allows the analyst to build an accessible data archive. As I will explain at greater length in chapter 8, this 'archival' function of transcripts can be supported by various additional techniques.

In short, making transcriptions helps one to take note of particular phenomena, it serves to build an accessible data archive, and it provides an audience with a limited but useful access to the phenomena discussed in an analysis.

Elements in constructing transcript files

Transcriptions, then, are always and necessarily selective. The system used in CA is specifically designed to reveal the *sequential* features of talk. As it has developed over the years, more and more details of the actual sequential production of talk-in-interaction have been added to the basic 'text', written in standard orthography, or a modified version of it. From its inception in the work of Harvey Sacks in the 1960s, this development has mainly been the work of Gail Jefferson, whose sensitivity and precision in the rendering of interactional details seems to be unmatched by anyone in the field.[2] Occasionally, other analysts have added particular features in which they had an interest to the system.

> [. . .] conversation analysts do *not* claim that the transcription system captures the details of a tape recording in all its particulars, *or* that a transcript should (or even could) be viewed as a literal representation of, or observationally adequate substitute for, the data under analysis. Like all transcription systems, the one used [in CA] is necessarily selective [. . .] and indeed this system is particularly concerned with capturing the sequential features of talk. (Heritage & Atkinson, 1984: 12)

In order to give some substance to my discussion of the transcription system, I quote an example of a transcript by Jefferson.[3]

Excerpt 5.1, from Jefferson, 1989: 171–2 [SBL:1:1:12:R:15–16:SO] ((telephone))

[line numbers added]

1	Maude:	I says well it's funny: Missi:z uh: ↑Schmidt ih you'd
2		think she'd help< ·hhh Well (.) Missiz Schmidt was the
3		one she: (0.2) assumed respo:nsibility for the three
4		specials.
5		(0.6)
6	Bea:	Oh↓*::. °°M-hm, °°=
7	Maude:	=Maybe:lle ↑told me this.
8	Bea:	Ah ↓hah,
9		(1.2)
10	Bea:	°Uh-hah, ° ·hh Isn't ↑her name ju:t plain Smi:th?
11		(0.7)
12	Maude:	Schmidt.h
13		(1.2)
14	Bea:	Oh I thought it was just S-m-i-t-h:.
15	Maude:	No I think it's S-c-h-m-i-d-t, something like that it's just
16		Sch↑mi↓:dt.
17		(0.3)
18	Bea:	Ah hah.

Following Psathas and Anderson (1990: 80–4), I will discuss the following kinds of information available in a transcript file.

- Time, date, and place of the original recording.
- Identification of the participants.
- Words as spoken.
- Sounds as uttered.
- Inaudible or incomprehensible sounds or words.
- Spaces/silences.
- Overlapped speech and sounds.
- Pace, stretches, stresses, volume, etc.

I'll take up these elements one by one in the same order.[4]

Time, date, and place of the original recording

As an essential part of the research *archive*, one should note the details of the recording occasion. When making a recording public, however, it is usual to use a coding system that is opaque to outsiders, as in the example above.

Identification of the participants

Participants are identified in the left column by a letter code. Psathas and Anderson (1990: 80) suggest that within the CA framework 'the respective

membership categories of the participants are not deemed relevant, except as they appear/are accomplished in the course of the interaction'. So in transcripts of 'ordinary' conversations, you see either letters or names being used (as in excerpt 5.1). For studies of institutional interactions, however, most transcribers use some sort of categorical identification. Rodney Watson, however, in a recent essay on the relation between sequential analysis and *Membership Categorization Analysis* (MCA), objected to just this practice. He notes (Watson, 1997: 51–3), for instance, that CA studies of medical interaction are in the habit of presenting their data in the format:

Excerpt 5.2

Dr: Did y'feel sick.
(0.6)
Pt: A little bit. Yes

He argues that such a presentation seems to be 'instructing' the reader to 'hear' the utterances transcribed as being produced by 'the doctor' and 'the patient', respectively, without providing or inviting an MCA of the utterances under consideration. This critique is part of a general argument that pleads for a re-involvement of MCA in the CA enterprise generally.[5] Watson suggests that in later CA 'categorical' aspects tend to recede to a background status, while sequential organization is 'foregrounded'. This, he suggests, impoverishes the analysis and may lead to a 'constructive analytic' reification and stabilization of the categories involved. Analysts would do better, he thinks, to 'combine' categorical and sequential analyses and include the interactional relevance of various categories into their analytic problematic (cf. Schegloff, 1991: 49–52). Although I sympathize with this argument in principle, I think that if one were to use, say, 'John' and 'Mary', or 'Mr Jones' and 'Mrs Peterson', while knowing they 'are' doctor and patient to each other, that would be too artificial an agnosticism. So even this minute technical detail can be seen to have analytic consequences, or at least carry analytic suggestions. As Psathas and Anderson note, it ultimately depends on the analysts' purpose, doing 'pure' or 'applied' CA, for instance, what kind of identification code is most useful.

Words as spoken

Under this heading Psathas and Anderson write that 'a first effort is directed toward capturing (in written form) the actual words as spoken'. And they continue saying: 'The assumption here is that the interactants are engaged in the use of conventional linguistic forms grounded in a common language with semantic and syntactic conventions' (Psathas and Anderson, 1990: 80–1).

This suggests that one starts the transcription by rendering the words spoken in standard orthography, which seems sensible. It does, however, already 'translate' stretches of sound into strings of discrete units, i.e. 'words', which in fact may not be audible as such on the tape. While this translation seems unavoidable, it should not be ignored that it is being used (cf. Duranti, 1997: 123–6).

The assumption of 'a common language' may be even more problematic, however. Contemporary sociolinguistics seems to undercut this kind of assumption for many if not most communicational situations. One may think of situations in which speakers with different linguistic backgrounds talk with one another in a lingua franca, or in the language of one or the other (cf. Wagner, 1996). And one should also recognize the fact that even 'within' a language there may be more or less marked linguistic variations related to various sub-communities. *Black English* is probably the most noted and best researched example of this (cf. Erickson & Shultz, 1982; Labov, 1972). Well-known phenomena like *code switching*, linguistic mockeries, *sound play*, and *language mixing* further complicate the issue.

In terms of my earlier discussion of transcripts trying to catch both the *what* and the *how* of talk-in-interaction, one can say that rendering the 'words spoken' in standard orthography involves an idealization of *speech* in terms of the standard *language* (cf. Duranti, 1997: 125). This obviously harms the purpose of rendering the *how* of actual speech. Therefore, many transcribers modify the standard orthography in order to catch some of the ways in which the actual speech practice *deviates* from the model implied in standard orthography, as Jefferson has done in the transcripts exemplified in excerpt 5.1 (cf. line 2: 'Missiz Schmidt'). Since readers are used to texts in standard orthography, however, such modifications seem to have a stronger impact on the reader's experience than the transcriber may have intended. The fact that such modifications seem to suggest stronger deviations than actually heard on the tape may picture the speakers as remarkably sloppy or stupid, and their overall speech as overly regional or 'ethnic' (Duranti, 1997: 137–44; Heritage & Atkinson, 1984: 12; Jefferson, 1983).

In other words, the dilemmas created by the effort to combine the *what* and the *how* in a readable way force one to make an explicitly reasoned choice and to stick to it if one wants to avoid confusion. Let me discuss some possibilities:

- One solution would be to use standard orthography throughout, ignoring language variation as well as everyday 'sloppiness' or 'informality', even if possibly locally relevant for the interaction. In so doing, one would lose the possibility of noting and studying many interesting phenomena. Most conversation analysts would be of the opinion that this is too close to a secretarial conception of transcription.
- A second solution would be to use standard orthography most of the

time, but use some modification to mark some specially significant 'deviations'. The problem with this is that it seems hard to decide when to choose one or the other option, when in actual fact the difference is very gradual. This method creates variations in the transcript which are not clearly related to variations in the talk.

- A third option is to use modifications continuously and as far as possible consistently. This seems to work best for the researcher's own purposes, that is, the 'noting' and 'archive' functions discussed before, but may be difficult for uninitiated readers of CA transcripts.

This discussion[6] leads me to the overall suggestion that one should adapt one's transcription style to one's purpose and audience, that one should be clear about one's method, and that one should use it consistently.

Sounds as uttered

Apart from the vocal sounds that can be interpreted as words, all other sounds that might play a role in the interaction are to be noted as well. These include vocal sounds that can be rendered as 'tch', 'pt', 'eh' or 'uh' and 'mhm' (and many variants), inhalation and exhalation, and laughter. The general idea behind this practice is that these vocalizations can have interactional meaning, for instance as a claim to a turn of speaking.

Excerpt 5.3

32	A:	Ye:s u[h huh
33	B:	[°Mm.°
34	B:	°M[m,°

From a 'language'-oriented perspective, however, this aspect of the Jefferson tradition has again been criticized as being overdone (Haegeman, 1996; O'Connell & Kowal, 1994). One argument for the inclusion of such elements is that they contribute to a 'picture' of the *rhythm* of the talk, especially when the 'transcription' is done in a manner that represents their construction out of 'syllable-like' parts, as can be done quite well with laughter (see, in particular, Jefferson, 1985b). This is what has been tried in the just quoted excerpt (5.3). Similar efforts to visually 'picture' the stream of vocal sounds influence other parts of the system.

There are, of course, sounds on the tape that are not sensibly 'transcribable' in this way, but these are mainly non-vocal sounds. These can be *de*scribed rather than *tran*scribed. Such descriptions are put within double brackets, to indicate their non-transcript status. The following excerpt (5.4), representing the start of an emergency call (CT = Call-Taker; C = Caller), has both.

Excerpt 5.4, from Zimmerman, 1992: 433 [MCE/20–10/196]

```
1  CT:  Mid-City police and fire
2       ((background noise and music on the line))
3  C:   (YA::H) Thiz iz thuh (      ) ((voice
4       is very slurred))
5       (1.5) ((loud background noise))
6  CT:  Hello:?
```

Inaudible or incomprehensible sounds or words

It is quite common that some vocal sounds are not easily transcribable because they are not comprehensible to the transcriptionist. In such cases, one can still try to guess what might have been said or to capture the sound as best as one can, and one can try to preserve the rhythm of the sound, including the number of 'beats' (syllables), the duration, and any intonational or stress patterns (see below). Such uncertain transcriptions are put within single brackets. It is also possible to note alternative hearings between which one cannot for the moment make a choice. Experience shows that one will quite often be able to 'hear' what was said when one returns to the data at a later date, or have someone else listen to the fragment. One can try to fit in one's mind various alternatives to the sound, until one of these seems to 'click', so to speak.

Spaces/silences

It is clear from many studies that pauses in speech can be very significant, although it may be unclear at first what their significance is. Pauses can occur when one party stops speaking and no one else takes the next turn, at least not immediately, in the 'rhythm' of the interaction so far (cf. Sacks et al., 1978). When the previous speaker continues speaking after such a 'break', for example by what has been called a 're-completor', it becomes a 'within-turn pause'. Or another may finally speak, 'to break the silence' as they say, possibly changing a topic which has run out; the break would then have become a 'between-turns pause'. One can also find occasions where someone has initiated an action or has given information and no uptake follows. When the projection for this uptake is clear, as with adjacency pair formats, such an 'absence' is noticeable and accountable (Schegloff, 1968). When it concerns a telling that *might* have been taken up, this is much less so (cf., for example, Ten Have, 1991a). Still another possibility is that at a moment no one is speaking, a non-vocal action takes place, which is only discernible using a video recording. In any event, noting pauses has proven to be important, but what a pause 'means' may be difficult to decide.[7]

The issue, then, is not *whether* pauses are to be noted, but *how* this will

be done. In the CA tradition a denotation given in numbers in parentheses has become the habitual method; (0.7) meaning one seventh of a second, for instance. Again, this practice has been criticized as being 'overdone' and suggesting a level of exactitude that one is not able to realize in actual practice (Haegeman, 1996; O'Connell & Kowal, 1994).

Gail Jefferson starts her paper on the length of silences in conversation with the following 'confessions':

> For most of the 18 years that I have been producing transcripts for the analysis of naturally occurring conversation, I have been timing silence in tenths of seconds. While I try to be accurate, I have not given particular attention to the phenomenon of silence *per se*, and have been content with rough timings. (So, for example, I started out using a stop-watch but in 1968 it broke and instead of replacing it I switched over to the method favoured by amateur photographers, simply mumbling 'no one thousand, one one thousand, two one thousand . . .') (Jefferson, 1989: 166)

In the paper she describes how she became gradually aware of a 'possible metric which provides for a "standard maximum" silence of approximately one second'. She reports that in a later phase of her explorations, she 'started retiming and counting the silences in some face-to-face, multiparty conversations', and she adds:

> And given that there was now good reason to be as accurate as possible with the timings, I bought a digital stopwatch, now timing the silences both 'photographer' fashion and by clock. The timings are fairly consistent, within a tolerance of about a tenth of a second, but still rough. (Jefferson, 1989: 182)

In their earlier mentioned paper on transcription practices, Psathas and Anderson (1990: 82, 86–90) also provide an extensive discussion of the timing of pauses. As they write:

> The systematic attention to and notation of silence, gap, or lapse as a timed and visually displayed unit is a standard practice in the Jeffersonian Transcription System. [. . .] The methodological maxim operative in the timing of these phenomena is that the transcriptionist strives to be internally consistent rather than to arrive at a standardized (clock time) demarcation [. . .]. The reason for this is that transcriptionists strive for a rendering that is as close as possible to the experience of those actually participating in the interaction. The transcriptionists' close and repeated listening to the interaction enables her/him to perceive the relative differences in the spaces (pauses, gaps, silence) that occur. (Psathas & Anderson, 1990: 87)

The general idea seems to be that by closely attending to the pace of the talk, the transcriber can catch the local significance of pauses by an informal method of 'rhythm-sensitive timing'. This surfaces in the following quote, which comes after the one above.

> [. . .] a consistently used mnemonic and silently uttered 'metronomic beat', such as the phrase 'one one thousandth', which has five distinct beats, each equivalent to two tenths of a second, can serve as a 'self-standardized' measurement device. The transcriptionist can count this off whenever a space in the talk occurs and thereby achieve a consistently applicable estimate of the length of each silence. (Psathas & Anderson, 1990: 87)[8]

On the basis of this 'relativist' method of timing, the authors warn:

> [. . .] readers are cautioned not to interpret these timings in an overly precise fashion; not to attempt to compare, across different analyst's transcripts, the occurrence of timings of different length. (Psathas & Anderson, 1990: 87)

In my own practice, I have used both methods of timing. I must say that I feel more comfortable timing with a stopwatch. But even there, I use repeated timings, because one still has the problem of 'catching' the exact onset and finish 'points'. By closely monitoring one's timing activities, the closeness of the correspondence of one's 'clickings' with what one hears, one can observe which of the timings are better than others.

In recent times, a third method of timing pauses has become a real possibility. This is based on computer software (such as 'SoundEdit' and 'CoolEdit') that can produce a visual display of digitized sounds. The idea is that one can just 'see' when the sound level is low, and use the horizontal time axis to 'measure' such periods. At the moment, I have neither the experience, nor the technical competence to explicate and evaluate this method any further. Nor have I read any reports from conversation analysts who have used it.

In sum, there are three methods of timing pauses:

- using an informal beat count as a proximate measure of 'rhythm-sensitive' pause length;
- timing with a stopwatch, to approximate a clock-time pause length;
- reading the acoustic pause length from a computer display.

As with other dilemmas in research, a practical solution will have to be found in the light of one's technical possibilities and analytic purposes. In any case, one should strive for a consistent and explicit method. A combination may offer the best results, because, in my opinion, both computer-based measurement and stopwatch timing should be seen as technical supports for the basic activity of hearing what is happening on the tape, rather than displacing it.

Overlapped speech and sounds

An essential feature of CA transcription is that it requires the transcriber to take careful note of phenomena of 'overlapping' speech. These phenomena

are most significant in terms of speaker transition, competition for the floor, etc., in short the operation of the turn-taking system, as analysed in the classic paper by Sacks, Schegloff, and Jefferson. The basic idea of that paper (originally published in 1974, but used here in the 1978 version) is that turns-at-talk in ordinary conversation are constructed in the actual course of speaking, using locally recognizable 'units' (TCUs, for turn constructional unit; cf. Schegloff, 1996a) as their 'building blocks'. During the production of any TCU, the current speaker will be treated as the 'owner' of the turn, but as soon as it is finished, another speaker might come in, unless special measures are taken to prevent this. This 'moment' is therefore called a 'transition relevance place' or TRP. Many CA studies have paid close attention to the management of these 'moments' or 'places' in the flow of talk, demonstrating the enormous interactional importance of a whole range of phenomena related to them (cf. Jefferson, 1973, 1986; Ochs et al., 1996). It is essential, therefore, that one tries to capture the details of turn management as closely as possible in one's transcripts.

Excerpt 1.1, repeated below, is an excellent illustration of the usefulness of careful transcription of overlaps:

Excerpt 1.1, from Heritage, 1984a: 236 [NB:VII:2]

```
E:  Oh honey that was a lovely luncheon I shoulda ca:lled you s:soo[:ner but I:]l:[lo:ved it.
M:                                                    [((f)) Oh::: ]  [(  )
E:  It w's just deli:ghtfu[:l. ]
M:                         [Well]=
M:  I w's gla[d     you] (came).]
E:            ['nd yer f:] friends] 're so da:rli:ng,=
M:  =Oh:::[: it  w'z]
E:        [e-that P]a:t isn'she a do:[:ll? ]
M:                                   [iYe]h isn't she pretty,
          (.)
E:  Oh: she's a beautiful girl.=
M:  =Yeh I think she's a pretty gir[l.=
E:                                 [En' that Reinam'n::
          (.)
E:  She SCA:RES me.
```

The special 'interweaving' of the assessments that characterizes this fragment would be completely lost if the lines were just typed one below the other, without marking the overlap starts and stops.

Or take a look at the next example:

Excerpt 5.5, from Jefferson, 1989: 172 [Fr: USI:2:R:2:SO] (face-to-face)

[line numbers added]

```
1  Carol:  Victor
2  Vic:    Ye:h?
3  Carol:  Come here for a minute.
4                (1.0)
5  Vic:    You come he[: r e .      ] please?]
6  Carol:                [↑You can] come b]a:ck=
7  Vic:    =I ↑have to go to the ba:th↓room.=
8  Carol:  =°Oh:.°
```

As Psathas and Anderson write:

> [. . .] it is possible to display (1) where the overlap began; (2) with which other speaker the speech/sound was overlapped; (3) when the overlap ended (though this becomes difficult to do in transcription and is often not precisely noted); and (4) what the speech/sounds were within the overlapped segment, for both parties. (Psathas & Anderson, 1990: 82)

In the earliest CA publications, the place where a 'second' overlapped a 'first' was marked with double slashes ('//'), but this device is not used anymore, being replaced by the square bracket system already demonstrated above. The latter device allows for a much more mnemonic display of the overlap, especially when the closing bracket is in fact used, as it in excerpt 1.1 above.

In some cases, Jefferson 'stretches' the display of one of the overlapped parts by using extra spaces, in order to provide a clearer 'picture' of the fact that the two parts in actual speech took about the same time, although the transcription in itself is of unequal length, as in line 5 in excerpt 5.5, and in the detail from excerpt 1.1 quoted below.

Excerpt 1.1, detail

```
M:  I w's gla[d      you] (came).]
E:            ['nd yer f:] friends] 're so da:rli:ng,=
```

This practice, like some of Jefferson's others, has been criticized for its inexactitude, but I like its suggestiveness.

Pace, stretches, stresses, volume, etc.

Under this rubric, Psathas and Anderson collect a number of conventions that further elaborate the *process* rather than the *content* of talk. This includes:

- 'latching', when one spate of talk directly follows another, with no gap between the two, indicated by an equals sign, as in lines 6 and 7 in excerpt 5.1;
- cut-off of a word in a markedly abrupt fashion, marked by a dash at the end of the word, as in line 12, excerpt 4.3;
- stretching, of words and other sounds, indicated by full colons after the stretched syllable, letter or sound, with the number of colons suggesting the length of the stretch, as in lines 1, 3, 6 and others in excerpt 5.1;
- stress, the (part of) a word or other sound that is stressed is underlined or printed in *italic*, as can be seen in the excerpts in this and other chapters;
- volume, markedly loud parts are printed in CAPITAL letters, while softly spoken words are enclosed by degree signs, as in line 10, or double degree signs for very soft, as in line 6, both in excerpt 5.1;
- intonation is marked by a special use of punctuation, with a question mark signalling rising intonation at the end, a period a downward, 'closing' intonation, and a comma a non-final intermediate one; in addition to this punctuation for intonation practice, some transcribers use arrows to indicate marked rising or falling of intonation; check excerpts 5.1 and 5.5 for examples.

Some of these conventions have been seen as problematic by some people. For all conventions, there is a problem of 'quantity': how much should an element be stretched or stressed in order to be noted as such, how steep should the rise in intonation be, how loud or soft the word marked as such, etc.? Here the remarks about the relativism of the enterprise, as made before concerning the timing of pauses, are again relevant. Intonation and stress, for instance, are rather subtle aspects of speech. Some transcribers mark intonation and stress only when they deviate from 'expected' variations, for instance when someone stresses a word that would not ordinarily be stressed in such a sentence; a 'natural' stress and intonation pattern would then not be marked in the transcript at all. This strategy is similar to the one noted earlier, to use standard orthography except for marked deviations. But in both of these cases, such practices introduce a 'normative' element in the transcription process that may be inevitable, but that others might like to minimize.

As an illustration of this problem, consider the following, not uncommon experiences. In a Spanish text, the name of the painter Goya does not have an accent, although the first syllable is ordinarily stressed and the second not. When someone in a Dutch conversation uses that name in a Spanish manner, should one transcribe it as 'Goya'? Or consider the case of a Dutch speaker who does not know the rules of Spanish pronunciation very well, and who stresses the second syllable, as in French, 'Goy↑a'. In any case, the decision on whether to transcribe one or another syllable as stressed is at least partly based on normative considerations. But if one

fails to make such indications, some imaginable later utterances, such as corrections, 'Goy↑a? oh, you mean ↑Goya', might become unintelligible. In short, making useful transcriptions requires both care and compromise!

Formatting issues

In the general literature on transcription, a distinction is often made between different ways in which the transcript can be arranged *visually* on the page. Edwards (in Edwards & Lampert, 1993: 10–12) differentiates what she calls 'vertical', 'column', and 'partiture' systems. The Jefferson system is a 'vertical' one, in that the utterances by different speakers are printed one below the other in the order in which they were spoken. In the column system each speaker has his or her column, which suggests essential differences between the parties. The 'partiture format is highly efficient for capturing stretches of an interaction that involve many simultaneous utterances or actions', writes Edwards (Edwards & Lampert, 1993: 11). It is structured as in a musical score, with each speaker having his or her own line, which indeed facilitates the precise noting of overlaps.[9] One could say that the way overlaps are displayed in the CA system is a 'borrowed' element from a 'partiture' system into a 'vertical' one.

In fact, when I started making transcripts, I began using a partiture system. But I soon took up the Jeffersonian system, for two reasons. The first was that I did not want to create a huge and inessential difference between my transcripts and those of other CA researchers, for reasons of 'membership' as well as readability. And the second reason was that it proved to be less easy to refer to particular turns in a transcript when one uses a partiture system, because it allows for more than one turn on a line. In any case, what I am discussing here is the system that is used in CA, but this comparison raises some further issues in the formatting of transcripts that are also noted by Psathas and Anderson.

In a 'vertical' system like the one used in CA, you have to decide what kind of 'unit' you will put on a line, or, to formulate it differently, when to start a new line. The simplest solution would be to continue putting transcription text on a line until you reach the right margin, or the current speaker stops and another starts. This seems generally to be the way Jefferson works, as in excerpt 5.1.

There are good reasons to make lines in a transcript fairly short and change to a next line earlier than at the usual right margin. A major one is that the number of characters one can put on a line varies from one word processing format to another, depending on font type and size. One could, of course, use rather narrow margins for a transcript, and produce short lines in a 'mechanical' fashion, just on the basis of who speaks and how the margins are set. One could also argue, however, for a strategy in which one would try to have a line's content to display some kind of 'unit'. Psathas and Anderson (1990: 85–6) mention several possibilities:

- 'breath units', what 'the speaker could produce in one breath';
- 'phrasal or clausal units', 'distinct or partial phrases or clauses as semantic/grammatical units';
- 'turn constructional turn completion units','turns or turn constructional components which may or may not become "turns" as a result of the next speaker beginning or not beginning to talk'.

They conclude their consideration of this issue with the following remarks:

> It should be clear that the transcriptionist's choices in breaking speech and action into line units with numerical indicators may be based on any number of analytic considerations. We should also note that the same data, when analyzed for different purposes, may be re-transcribed with different line-by-line divisions. Although there is no single 'best' linear representation, we urge the reader to consider the ways in which the choices regarding the line-by-line production of a transcript may affect the analysis. We would certainly caution readers of transcripts not to take the number of lines in a transcript to be an indicator of the temporal length of the transcript. (Psathas & Anderson, 1990: 86)

In my own experience, this issue of 'line units' became relevant when I was experimenting with the possibilities for using a software program supporting 'qualitative data analysis' for CA purposes (Ten Have, 1991c; I will return to these possibilities in chapter 8). That program, The Ethnograph, has lines as its basic units that one can 'code'; therefore, I had to limit lines to elements that might deserve an independent 'code'. But even without such restrictions, it might be easier to limit lines to some basic units, like TCUs, in order to facilitate clear referencing using line numbers.

A special issue in the 'line formatting' aspect of transcript is the role of pauses. Look again at excerpt 5.1, partially repeated below:

Excerpt 5.1 detail

1	Maude:	I says well it's funny: Missi:z uh: ↑Schmidt ih you'd
2		think she'd help< ·hhh Well (.) Missiz Schmidt was the
3		one she: (0.2) assumed respo:nsibility for the three
4		specials.
5		(0.6)
6	Bea:	Oh↓*::. ° °M-hm, ° °=
7	Maude:	=Maybe:lle ↑told me this.
8	Bea:	Ah ↓hah,
9		(1.2)
10	Bea:	°Uh-hah, ° ·hh Isn't ↑her name ju:t plain Smi:th?
11		(0.7)
12	Maude:	Schmidt.h

You will see that there is a (0.2) pause in line 3, after which talk by the same

speaker, Maude, continues. On line 5, there is a (0.6) pause, which has been given its own line, so to speak. After that pause, Bea takes the next turn. The first pause, then, is formatted as an *intra-turn* pause, while the second is an *inter-turn* one. But now look at the (1.2) pause in line 9. It also has its own line, but after it, the previous speaker, Bea, continues. So here the two rules for starting at a new line (at turn change and on reaching the right margin) do not suffice. On some occasions of within-turn silence, then, Jefferson notes that silence on a new line, followed by subsequent speech by the same speaker on another new line, while in other cases, a similar silence is included on the same line as the preceding speech, with more speech being typed at the line after the silence.

One could think that this is an arbitrary matter, maybe based on visual, aesthetic preferences. But I agree with Psathas and Anderson (1990: 88–90) that the format in these cases carries an analytic suggestion. In a sense, then, a formatting choice in this matter reflects what a pause is considered to 'be' interactionally'. In fact, in scanning some extended transcripts by Jefferson, the pauses that she formated as being *intra-turn* seem generally to be small and occur either between a 'starter' (like an in-breath or a 'well') and the turn 'itself', or within a TCU which, at the place where the pause occurs, is evidently not finished. The point of all this is

> that the transcript may and does incorporate some analysis, as it is being produced by the transcriptionist. The seemingly simple matter of how interaction is presented in a line-by-line format should be carefully considered when interpretations of interactional phenomena are based on the 'display conventions' rather than the 'actualities' of phenomena. (Psathas & Anderson, 1990: 89–90)

A transcript, then, is, as Psathas and Anderson (1990: 90) say, a 'post-hearing/seeing depiction', 'a *constructed* version of the actualities and particularities of the interaction'. They note seven properties of transcripts that should be taken into account, seven ways in which the experience of reading a transcript is bound to be different from the experience of the lived interaction that is being depicted. Most of these have been discussed in the preceding sections. Transcribing recorded talk is a necessary but rather imperfect, instrumental task within the CA enterprise. It is essential to work hard on it, and to recognize inevitable limitations.

Adding visual information

The CA conventions, as discussed above, evolved from the transcription of vocal sounds available in audio recordings of both telephone and face-to-face interaction. In later periods, researchers working on different kinds of materials have added conventions to depict the particular phenomena that they wanted to investigate. Max Atkinson's earlier-mentioned studies of political oratory, focusing on *applause*, for instance, necessitated a

careful 'transcription' of applause in close connection to the speech to which it was a reaction (cf. Atkinson, 1984a, 1984b, 1985). In a similar fashion, researchers working with *video* materials, like Charles Goodwin (1979, 1981, 1986, 1987, 1994b, 1996, etc.; C. Goodwin & M.H. Goodwin, 1996), Marjorie Harness Goodwin (1995), Christian Heath (1986, 1988; Heath & Luff, 1996) and Christoph Meier (1997), to name but a few, have developed and used methods to add information on visual phenomena to a transcript of vocal actions. As noted before, this book deals mainly with the analysis of auditory materials, as another volume in the series will treat video, but I think a few notes and references concerning the depiction of visuals are in order (Duranti, 1997: 144–54; C. Goodwin, 1981: 46–53; Heath, 1986: ix-xiv, 1–24; 1997; Heath & Luff, 1993; Jordan & Henderson, 1995; Meier, 1997: 41–8).

The basic procedure used in CA studies based on video recordings has been to start with a detailed transcription of the vocal part of the interaction, and add description or symbolic depictions of the visual activities, like gaze, gesture, posture, and others, to the 'time line' provided by the transcript, either above or below each line. In order to clarify the 'location' of activities during vocal pauses, these may be indicated by 'dashes' ('-'), each for a tenth of a second.

> The researcher should at least indicate the onset and completion of particular movements [. . .]. It is also useful to indicate any critical junctures within the development of a particular movement. Movements are represented by a continuous line, although in fact a whole range of ad hoc signs and symbols are often used to represent particular aspects of movement. (Heath & Luff, 1993: 317)

In order to make it clear how this works for one aspect of visual conduct, let me quote from Charles Goodwin's explication of his system for the notation of gaze, on which his early work (1979, 1981) was focused.

> Gaze will be transcribed as follows: The gaze of the speaker will be marked above the utterance and that of the recipient(s) below it. A line indicates that the party being marked is gazing toward the other. The precise place where gaze reaches the other is marked with a capital X tied to a specific place within the talk with a bracket. [. . .] The movement bringing one's party gaze to the other is marked with dots, whereas the movement withdrawing gaze is indicated with commas. (C. Goodwin, 1981: 52)

This system is exemplified in the following:

Excerpt 5.6, from C. Goodwin, 1981: 52

```
A:  _____,,,,,,,
    We went down t o- (0.2) When we went back . . .
                        [
B:              . . X_____
```

In addition to these transcriptions, some aspects of the original video can be made available in print by still pictures, either in the form of *drawings* made after the original video picture (as in Heath, 1986, 1988; McIlvenny, 1995) or so-called *frame grabs*, digitized frames taken from a video tape (as in C. Goodwin, 1994b; M.H. Goodwin, 1995; Meier, 1997; Suchman, 1992). The drawing technique has the advantage of preservation of the anonymity of the participants, while the digitized frame pictures allow for the addition of explicative symbols like arrows (who speaks to whom) and initials (as in Meier, 1997). When using electronic publishing, on CD-ROM or the World Wide Web (see discussion in chapter 8), even more sophisticated techniques are possible, including 'sound bites', and even 'video clips'.[10] It seems clear that the possibilities of these technologies will be more fully used in the near future.

Translation

The methodological literature of CA hardly ever discusses problems of translation, but for anyone who has to present to an audience which is not familiar with the language used by the participants, translating such materials is a difficult task (see Duranti, 1997: 154–60). These difficulties are reflected in the various ways in which translations are presented, or not, in actual CA publications. I have seen publications in which:

- the materials are only presented in translation into the language of the publication;
- the materials are presented in translation into the language of the publication in the body of the text, with the original transcript given in an appendix (cf. Bergmann, 1992);
- the materials are presented in translation into the language of the publication in the body of the text, with the original transcript given immediately below it, as a separate block of text (cf. Houtkoop-Steenstra, 1991), or the other way around, first the original and then the translation (Ten Have, 1991b);
- the materials are presented in the original language, with a translation into the language of the publication immediately below it, line by line (cf. excerpt 4.3);
- the materials are presented in the original language, but with first a morpheme-by-morpheme 'gloss', and then a 'translation' into the language of the publication immediately below it, line by line (cf. excerpt 5.7 below).

In my view, only the last three options are acceptable, since I think that one should provide the readers with as much information on the actual, original interaction as possible. The difference between the last two options seems to be related to the 'distance' between the two languages

involved. When the two are not too different, as in the cases of Dutch/German or Dutch/English, one can catch a lot of the original interaction in an almost word-for-word translation. In such a case, the translator has to balance the two interests of, on the one hand, rendering the original talk as faithfully as possible and, on the other, of producing a translation that seems 'natural' in the destination language. When the two language systems are very different, however, as in the pairs Finnish/English or Japanese/English, these dilemmas may be impossible to solve in an acceptable manner, because various means of expression natural in one system are simply absent in the other. In such case, the researcher has to provide both morpheme-by-morpheme glosses and a free translation. Below is an example of this method, with Finnish being the original language. The second lines do not only provide 'words', but also grammatical information, which is explained in an appendix to the paper from which the example has been taken (Sorjonen, 1996: 326–7).

Excerpt 5.7, from Sorjonen, 1996: 281–2

1 S: *Hy*[*vä juttu*]
good thing
Goo:d]
[

2 T: [>.*h* ↑Arto *ei pääse*] tule-e Se on vähä
[1nameM NEG be able to come-ILL it is a little
[>.h ↑Arto can't make it. That's a bit of a

3 *tyhmä-ä*<.
stupid-PAR
nuisance.

This example amply illustrates the difficulties involved in translating orally produced materials, I think. And when the interests of the analysis are directed at specific *linguistic* phenomena, such as word order or particles, this three-line format might also be sensibly used in situations where the two languages are more similar than Finnish and English.

Practical issues

Making transcriptions is extremely time-consuming. One should, therefore, carefully consider which parts of the available recordings should be transcribed and in what kind of detail. This, of course, will depend on the particular research interests one has. If you are interested in, say, the overall structure of telephone conversations, it makes sense to make complete transcriptions of a limited set of examples. But if you are after a

particular kind of interactional feature that happens now and then in a large corpus of video tapes, you should rather transcribe only those relevant episodes.

Especially when working with video tapes, it makes good sense to start with making an inventory of the tape's 'content', a 'content log' (C. Goodwin, 1994a; Jordan & Henderson, 1995: 43; Suchman & Trigg, 1991). Listening to the tape, or viewing it, one makes summary descriptions of what happens, adding notes on especially interesting happenings, non-recorded contextual particulars, etc., using either the counter numbers or time stamps (on video) as an index. In this way, one gets an overview of what is available, which allows a relatively quick retrieval of episodes to consider for more detailed consideration and/or transcription.

Douglas W. Maynard (personal communication) suggests still another sensible strategy for large projects:

> If you've collected a large number of very long recordings in some setting, it may be extremely inefficient to do detailed transcriptions of the entire corpus. Instead, you may be interested in 'advice-giving', or 'diagnostic news delivery', or the 'medical exam' (in doctor–patient interaction, for instance), and only do the detailed transcripts of those aspects of the interaction while writing standard transcripts of the rest. This way you can read at least rough versions of the entire interview and know what happened before and after focal episodes, while having the details on just those episodes themselves.

My general suggestion for making transcripts is to do it in 'rounds': start with putting down *what* has probably been said in standard orthography, and add the various details concerning the *how*'s later, one type after the other. One can, of course, make a note of remarkable details in earlier rounds than the ones in which one concentrates on a certain type, but it proves a good practice to focus on particular kinds of phenomena one after the other, for instance 'intonation', 'pauses', etc.

The ideal play-back machine would allow one to listen repeatedly to the same fragment, leaving the machine in 'play', while pushing the 'rewind' button for the required period to return to the start of the fragment that one wants to focus on. A foot pedal is very 'handy' for this type of work and many transcribers prefer to use headphones for concentration.

Originally, transcripts were produced on the typewriter, but nowadays word processors are used. There are various kinds of difficulties that one encounters in adapting the typewriter-based transcription conventions to this new platform. One concerns the production of specific transcript symbols, while another has to do with aligning simultaneous speech. These are discussed below. But using a wordprocessor also has advantages, of course, over using a typewriter, and some of these that are relevant to presenting transcriptions will also be noted.

In most cases, the signs and characters available on the QWERTY keyboard can be produced without any difficulty using a word processor. As noted, you can use either 'underline' or 'italic' to indicate stress, which are

TABLE 5.1 The production of transcript symbols

Symbol	Example	WordPerfect 5.1*	WordPerfect 6.1, 7	MS-Word 7
degree sign	°soft°	Alt-248	Ctrl-W, (6,36)	Input>Symbol
high point	·hh	Alt-250	Ctrl-W, (6,32)	Input>Symbol
up arrow	↑high key↑	Alt-24	Ctrl-W, (6,23)	Input>Symbol
down arrow	↓low key↓	Alt-25	Ctrl-W, (6, 24)	Input>Symbol
? and, combined	mild rise¿	Shft-F8, 4,5,1:,?	'overstrike':,?	not available

*Using the Numeric Key Pad.

available as standard facilities. Underlined letters or words tend to be more easily distinguishable than *italic* ones. The extra symbols, however, may be difficult, especially when you convert a text to another format, or even to another font.

Table 5.1 summarizes some suggestions which are based on my own experiences, which are, of course, limited.

When transcribing episodes in which one participant's talk overlaps with that of another, indicated by the use of square brackets, it helps to align the portions of simultaneous speech as precisely as possible. This creates special difficulties with modern word processors, which tend to use 'proportional fonts' (also called 'variable-pitch fonts'). With such fonts, the horizontal space a letter is accorded on the line varies with its size, 'w' getting more than 'l', etc., and with the number of letters in relation to the length of the line. Consequently, the exact place that a point of overlap starts or finishes can vary when something is added or when the margins are changed, or when a different font is chosen. As a solution one can try using a 'fixed-pitch' or 'monospaced' font, but it may require a bit of experimenting with one's word processor's fonts as well as one's printer. An alternative method is suggested by Charles Goodwin (1994a), who puts a TAB before the bracket and adjusts the TAB stop using the 'Ruler Bar'.[11]

Another suggestion of his (cf. C. Goodwin, 1994a) is that it can be useful to use a word processor's table feature to type the transcripts. One can define columns of different width for different purposes such as 'line number', 'time', 'arrows', 'speaker', 'utterance', and 'notes'. A 'landscape' format may be helpful so that each row can be longer than usual. The 'notes' column may be used to add 'observations' on hard to transcribe details, such as tone of voice or – in the case of video tapes – visual aspects. Alternatively, or in an additional column, one might add 'analytic' comments, pointing out remarkable phenomena that deserve attention in a later phase, etc. In presentations or publications, such non-transcript columns can be deleted and the table lines can be hidden (by changing the preferences for line display in the layout menu to 'none').

Learning to transcribe

Looking at any Jefferson transcript may make one wary of ever trying to make transcripts oneself, but – as many have suggested – making transcripts is an essential part of the craft of CA. Some may be more talented at the job than others, but all can learn to make useful transcripts. Probably the best setting is one of 'friendly supervision'. If you are working in a group, you might select a 5- to 10-minute fragment, provide all participants with a copy on tape, and have each make a transcript individually, using the same set of conventions. Then, go over the transcripts collectively, comparing the transcripts, listening to the tape, trying various 'hearings' of parts on which there is disagreement, try to reach collective agreement, and preserve alternative solutions if no agreement is available. This exercise could be repeated until each has acquired a minimal level of competence and confidence. If one is working individually, one might try to find an experienced transcriber willing to go over one's transcript together with the tape, and discuss any problems.

EXERCISE

Make a transcription of a 10-minute fragment of recorded verbal interaction, using the various suggestions of this chapter, including the last section. Keep notes of the process of making, discussing, and revising the transcript and write a report in which you confront your working experiences with the observations and suggestions of this chapter.

Recommended reading

Duranti, A. (1997) 'Transcription: from writing to digitized images'. In his: *Linguistic anthropology*. Cambridge: Cambridge University Press: 122–61

Jefferson, G. (1985b) 'An exercise in the transcription and analysis of laughter'. In: T.A. van Dijk, ed., *Handbook of discourse analysis*, vol. III. London: Academic Press: 25–34

Jefferson, G. (1989) 'Preliminary notes on a possible metric which provides for a "standard maximum" silence of approximately one second in conversation'. In: D. Roger & P. Bull, eds, *Conversation: an interdisciplinary perspective*. Clevedon: Multilingual Matters: 166–96

Jefferson, G. (1996) 'A case of transcriptional stereotyping', *Journal of Pragmatics* 26: 159–70

Psathas, G., & T. Anderson (1990) 'The "practices" of transcription in conversation analysis', *Semiotica* 78: 75–99

Notes

1. This conception of transcription as translating 'speech' into 'language' has been inspired by a presentation by Jan Blommaert at the 4th IWIA Symposium on 'Oral Communication in Organizations', University of Antwerp, 17 October 1997; see also Duranti (1997: 122–61).
2. Cf. Jefferson (1985b) for her progressively refined rendering of laughter and its gains. For more extended discussion of transcription practices, comparing approaches from CA, social psychology and linguistics, see Section 3: Transcription Procedures, in Roger & Bull (1989), and especially Psathas & Anderson (1990).
3. Transcription conventions are summarized in appendix A; a full version of Gail Jefferson's own description can be found in an appendix to the paper from which this excerpt was taken (cf. Jefferson, 1989: 193–6). That paper contains a large number of instructive examples of Jefferson transcripts.
4. My discussion owes much to the one provided by Psathas and Anderson, but my preferences and suggestions differ from theirs on a number of points, to be noted as I go along.
5. See Silverman (1998: chaps 5, 7) for a general discussion of Membership Categorization and the MCA/CA relationship.
6. I have only provided a brief and selective summary of what might be said on these issues; cf. Duranti (1997: 122–61), Edwards & Lampert (1993), Haegeman (1996: 87–110), Jefferson (1983). It should be stressed that some criticisms of the Jeffersonian practices refer to the argument that one wants to facilitate computer searches in large collections of data. As I will argue later (in chapter 7), this does not seem to be a sensible practice for CA and therefore is not a relevant argument concerning transcription within the CA tradition.
7. You might want to check excerpt 5.1 and think about the local significance of the various pauses there.
8. I was puzzled at first by the notion that 'one one thousandth' would have 'five distinct beats', but apparently the 'th' at the end should be pronounced as a full 'beat'.
9. The best known example of such a system is called HIAT (*Halbinterpretative Arbeitstranskriptionen*, which means 'semi-interpretive working transcriptions') and has been developed by Konrad Ehlich and Jochen Rehbein in Germany; see Ehlich's contribution to the 1993 Edwards & Lampert volume.
10. Leslie Jarmon has distributed her PhD on CD-ROM using such technologies, while Michaela C. Goll and Christoph Meier have done so in their on-line paper (1997), discussed in chapter 4 (see http://www.unigiessen.de/~g312).
11. Consult your word processor's 'Help' for how to use TAB-settings and the Ruler Bar.