



Transcription as a Crucial Step of Data Analysis

Sabine Kowal and Daniel C. O'Connell

For the first decade of the twenty-first century, the Google Ngram viewer (Michel et al., 2011) shows a sudden increment in books and articles in both English and German on the topic of *qualitative research methods*. This development is clearly related to the large number of scientific disciplines – including communication, economics, education, health care services, linguistics, marketing, psychology and social work – which involve qualitative research. One upshot of this surge has been that the use of transcripts for such research can now be ‘taken for granted’ (Dresing and Pehl, 2010: 731; our translation); hence, the necessity to engage the complexity of transcripts as ‘artefacts in need of thoughtful consideration’ becomes all the more urgent. At the same time, Harris (2010: 4) has warned against a certain methodological ‘incoherence’ on the part of contemporary linguistics consequent upon ‘a failure to recognize the nature of the disparity between oral and written communication’. His comment assumes a special importance with regard to the faithful

representation of oral communication in written transcripts.

A large portion of this complexity – and incoherence – is traceable to the heterogeneity of purposes served by transcription and the consequent variable standards across disciplines. For example, Langer (2010: 520; our translation) has set the bar fairly low: ‘In educational research projects, detailed notation systems are for the most part bypassed by reason of the specific *status quaestionis* and in order to foster simplicity and readability.’ Frost (2011: 101) has emphasized a more detailed approach to the transcription of interviews (see Roulston, Chapter 20, this volume) used in psychology: ‘The transcription of interviews is carried out in multiple rounds.’ These rounds might begin with a rough transcript, including the words uttered and other features such as pauses or laughing, followed by another round wherein shorter pauses, fillers and false starts are added. An even more exacting and detailed approach to transcription in the linguistic field of pragmatic research is to be found in Schmidt and

Wörner's EXMARaLDA, 'a system for the computer-assisted creation and analysis of spoken language corpora'. According to the authors, such corpora focus on

linguistic behaviour on *different linguistic levels*. It is usually not sufficient to simply record the syntactic and lexical properties of speech, because para-linguistic phenomena (like laughing or pauses) and suprasegmental characteristics (like intonation or voice quality) may play an equally important role in the analysis. The data structure must therefore also be able to accommodate and distinguish descriptions on different linguistic levels. (2009: 567)

An explicit concern about various levels of detail in transcribing is built into the German *Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)* developed by a group of linguists (Seltling et al., 2009) and into the English-language adaptation of *GAT 2* by Couper-Kuhlen and Barth-Weingarten (2011). These authors have themselves translated the acronym GAT as 'discourse and conversation-analytic transcription system' (2). It distinguishes 'three levels of delicacy: minimal, basic and refined transcript versions' (353).

In view of such diversity of research purposes and the concomitant transcription requirements, we wish to limit ourselves in the following to a consideration of what we think of as basic assumptions and principles needed for an informed use of transcription, with an emphasis on qualitative research in the social sciences and, more specifically, on dialogical interaction. Such an approach is in accord with Aufenanger's (2006: 111) recommendation that the choice of transcription methods be *appropriate for the specific purposes of a given research project*. Such adaptation also serves the purpose of avoiding superfluous and/or unanalysable transcripts.

The appropriate use of transcription entails an awareness of problems related to the tasks of both the transcriber and the reader of the transcript – conceptualized as *language users* who bring their own habits, competencies and limitations to these tasks. In addition, the relevance of transcription for both qualitative and quantitative data analyses should be

noted, especially in view of an increasing interest in bridging the gap between qualitative and quantitative methods (e.g. Flick et al., 2004; Kelle and Erzberger, 2004).

In the following, examples are, unless otherwise noted, taken from our own psychological research on dialogical interaction in a variety of settings. These corpora include transcripts of audio recordings of English-language TV interviews (O'Connell and Kowal, 2005) and feature movies (O'Connell and Kowal, 2012).

TRANSCRIPTION – A UNIVERSALLY INDISPENSABLE STEP IN RESEARCH

According to Peez (2002: 24; our translation), 'all social scientists doing qualitative research must ... carefully attend to the phase of setting down the verbal research material in writing by means of transcription'. In addition, there are applied contexts such as courtrooms and medical offices where records of spoken data are important. In order to deal with these otherwise ephemeral and elusive materials in an orderly manner, transcripts must be derived. And yet, the research community must face the vast complexity involved in this transfer to the written mode, especially when multimedia dialogical interaction is involved. The putative close correspondence between the spoken discourse and the written record thereof must be examined. Presently it is widely acknowledged that the written record cannot be accepted uncritically as a reliable source of analyses accurately reflecting the mental, social, affective and cultural components of both individual and group performance. For example, Chafe (1995: 61) has commented in the very last sentence of his chapter on transcribing, 'Perhaps the spoken corpora of the future ... should be packaged with a legal requirement that users listen as well as look' (see also Harris, 2010).

In other words, transcription is both an inevitable and problematic step in the qualitative (and quantitative) analysis of data consisting of spoken discourse. There is in fact no

transcription notation system capable of providing to the researcher a completely accurate and comprehensive narrative of the original performance: all transcription is in principle *selective* and entails the inevitable risk of systematic *bias* of one kind or another. Nonetheless, this risk can be countered by making decisions on the basis of reasoned choices rather than arbitrary, non-reflective ones. Consequently, both basic and applied researchers in the social sciences must approach transcription with a very critical eye (and ear).

Our critical remarks should therefore be understood as a sort of consciousness raising regarding the intrinsic methodological limitations of transcription and the consequent cautiousness that should be exerted in interpreting transcripts. Such cautiousness would also demand, quite in accord with Chafe's (1995) recommendation mentioned above, that the interpretation of transcripts should always be verified by a return to the audio and video recordings.

In light of the complex behaviours and contexts of dialogical interaction, we have chosen in this chapter to limit our more detailed discussion of transcription to the words spoken (the *verbal* component), to the way in which they are spoken (the *prosodic* component), and to whatever non-verbal vocal behaviour accompanies the words (the *paralinguistic* component). These three components are clearly the most frequently relied upon in qualitative analyses of spoken discourse. In addition, we have included a section on the transcription of turn-taking in the transcription systems presented below. Readers interested in the transcription of extralinguistic behaviour may turn to the readings we recommend below (Jenks, 2011; Kreuz and Riordan, 2011).

BASIC TERMINOLOGY FOR TRANSCRIPTION

Transcription

The generic term *transcription* here refers to any graphic representation of selective aspects

of verbal, prosodic and paralinguistic behaviour; in other words, we limit our overview of transcription to vocal behaviour. Such representation presupposes a unique performance and is typically not meant as a script for a further performance. The selected aspects are by necessity represented sequentially because real time is involved. There is in principle a wide range of detail involved in the transcription of these various aspects. This range has been illustrated by Chafe (1995: 56ff.) by means of a short utterance which he has transcribed in seven steps, by adding more prosodic detail at each step, starting with the verbal utterance presented in step 1, transcribed in standard orthography:

(1) the other thing you can do is (56)

and ending with step 7:

(7) ... (0.3) the **óther** thing you can **dò= i=s:** (58)

The additional steps have added the following prosodic notations:

- acute (*óther*) and grave (*dò*) accents for pitch prominence;
- boldface type (**other**, **do** and **is**) for greater loudness;
- equal signs (=) for the lengthening of the preceding vowel;
- spacing between *do=* and *i=s* for an even stronger accentuation;
- measured pause duration (0.3) in seconds in parentheses;
- a colon for level pitch (*i=s:*)

It has become a commonplace now to emphasize that the choice among the behavioural aspects to be included in transcription of verbal interaction cannot be determined independently of the purposes of transcription; but the choice is also dependent upon the competencies of the transcriber. The most basic part of any transcript always remains the verbal component. Chafe's prosodic transcription of step 7 above demands a large measure of linguistic competency on the part of the transcriber; it is also appropriate only when a given research project calls for the

representation of details regarding *how* a verbal utterance has been produced.

Description

Transcription is to be distinguished from *description*. The latter is useful as a supplement to denote paralinguistic or extralinguistic behaviours as well as non-linguistic activities observed in dialogical interaction.

Thus, a given instance of the paralinguistic behaviour of laughter may be transcribed, as shown in *Example 1* from a TV interview of Bill Clinton (BC; O'Connell and Kowal, 2005: 289):

Example 1

BC HE HA HA HA HE

Or it may be simply described as *laughter*. The description eliminates the notions of sequentiality, temporality, numerosity and the specificity of phonemes which are represented in the transcribed version of the laughter. More specifically, in *Example 1*, the sequential priority of the first occurrence of HE is to be noted along with the implication that the sequence occurs in measurable real time, involves five separable segments, and designates the phonemes specifically as HE and HA. As a simple notation of an event, the description of laughter is devoid of all these details.

Extralinguistic communicative behaviour includes *non-vocal* bodily movements (e.g. hand gestures and gaze) occurring during a verbal interaction. Both speakers and listeners may engage in extralinguistic behaviours. They are typically described rather than transcribed in qualitative research.

In some dialogical interactions, talking is not the primary activity of *all* the participants. A participant may initiate a verbal response or react to a verbal request with a *non-linguistic* activity. *Example 2* of a dialogical interaction where non-linguistic activities initiate brief verbal responses is taken from the movie *Bonnie and Clyde*

(Beatty and Penn, 1967) as presented in O'Connell and Kowal (2012: 115). A police officer is silently presenting to a witness photographs of potential suspects in a grocery robbery, while the injured witness is lying in a hospital bed. The non-linguistic activity of the police officer is described in brackets:

Example 2

Police Officer	[presentation of photo]
Witness	no
Police Officer	[presentation of photo]
Witness	huh-uh [as negation]
Police Officer	[presentation of photo]
Witness	no ...

Coding

Transcription is also to be distinguished from *coding*, which refers to the classification of events in discrete categories and the labelling of these categories. An example can be found in Bull and Mayer (1993: 655) who have classified the reactions of British politicians in response to interviewer questions into three categories: *replies*, *non-replies* and *answers by implication*. Note that coding is logically dependent on previous transcription and entails a further theoretical orientation as foundation for its categorizations.

Transcript

A *transcript* is the result of the activity of *transcribing* performed by a single person or by several persons, sometimes by the researchers themselves, sometimes by personnel not otherwise involved in the research. Some researchers have emphasized that transcribing and the analysis of transcripts should be done by the same persons (see, e.g., Dittmar, 2009: 59f.; ten Have, 2007: 95; Lapadat and Linsay, 1999, as cited in Tilley, 2003: 751). In fact, Chafe (1995: 61)

has bluntly stated, ‘One cannot fully understand data unless one has been in on it from the beginning.’

Notations

The set of signs used to represent selective aspects of the behaviour of participants involved in a verbal communication, that is the tools for transcribing, are referred to as *notations*. A *transcription system* is the sum of all the notation signs plus the conventions for arranging the signs sequentially on paper or screen and the methods used to assess the various behavioural aspects. With respect to the assessment of prosodic parameters, namely duration, pitch and loudness, basically two methods may be distinguished: with reliance upon the perceptual reliability of the transcriber(s) and with supplementation by instrumental measurement.

Transcribers and Transcript Users

In addition, there are two personal roles involved in transcription: that of the *transcriber* and that of the *transcript user*. From a psychological perspective, both roles demand extremely complex processing. Riessman (1993; cited in Frost, 2011: 101) has emphasized the importance of the transcript user when choices have to be made about how to design transcripts: these choices ‘have serious implications for how a reader will understand the narrative’. In fact, Du Bois (1991: 77ff.) has even made ‘transcription design principles’ a basis for his *Discourse Transcription (DT)* system in an effort to accommodate the needs of a large variety of users.

In the following section, we will discuss some common problems that have been shown in empirical research to have an impact specifically on the transcriber’s job of faithfully representing selective aspects of spoken discourse and that should be considered when training novice transcribers.

THE TRANSCRIBER, AS A LANGUAGE USER, IS ‘OFTEN QUITE UNRELIABLE’ (MACWHINNEY AND SNOW, 1990: 457)

Results of Transcription Research

A number of psycholinguistic studies have indicated that the production of transcripts from audio and video recordings by use of various notation systems is a quite demanding task. This fact has to be taken into account, especially in applied contexts where a transcript can have important consequences. For example, Walker (1986: 209) has reported the case from a court transcript in which the spoken phrase ‘male in extremis’ had been changed in transcription to ‘male, an extremist’.

O’Connell and Kowal (1994) have analysed four types of changes in the verbal component of transcripts in German corpora of spoken discourse (including parliamentary speeches, interviews and an informal conversation): deletions, additions, substitutions and relocations, including linguistic units ranging in size from phonemes to sentences. They found that changes were quite common, occurring on average every 13 syllables. Deletions were most frequent (42%), followed by additions (34.3%), substitutions (18.1%) and relocations (5.6%). Among the most frequent deletions were the function words *und*, *auch*, *also* (and, also, well) and the filler *äh* (uh); among the most frequent additions were corrections of elisions typical for spoken discourse, for example *is* was changed to *ist* and *n* was changed to *ein* or *eine* (a, masculine or feminine) indicating that transcribers were either deliberately or inadvertently ‘introducing alterations from characteristically spoken discourse to properly written discourse’ (132). These deletions and additions may be disregarded for some transcription purposes, but in other cases they may even constitute an open violation of the explicitly formulated transcription rules.

Dresing and Pehl (2011: 14; our translation), in their second transcription rule, have stipulated: ‘Word contractions are not transcribed

but instead are moved in the direction of standard German orthography.’ Their motivation in this regard is to simplify transcription rules. *Example 3* provides an English-language transcript from the movie *Bonnie and Clyde* (Beatty and Penn, 1967) in standard English orthography and, for comparative purposes, in our own transcript notation (adapted from O’Connell and Kowal, 2012). The reader may note that our transcript notation coincides with what will be described below as *literary transcription*:

Example 3

<u>Standard English</u>	Clyde	since it does not look like you are going to invite me inside
	Bonnie	ah you would steal the dining room table if I did
<u>Our Notation</u>	Clyde	since it don’t look like you’re goin’ to invite me inside
	Bonnie	ah you’d steal the dinin’ room table if I did

In the present instance, the difference made by using our own notation amounts to three syllables (*does not* → *don’t*; *you are* → *you’re*; *you would* → *you’d*); our version is also more faithful to the acoustic realization as articulated by the actors and reflects familiar spoken English usage.

More recent data from Chiari’s (2007) Italian corpora are couched in a similar set of four categories of changes as in O’Connell and Kowal (1994). Of interest is Chiari’s ‘most striking finding’ regarding:

the amount of repair that does not rely of [*sic*] linguistic form but on creative unconscious reconstruction made by the transcriber, that generally tends to preserve utterance meaning. The transcriber attributes intentions and beliefs to the voice heard, and tends to filter inevitably the spoken sounds re-interpreting them in a way that is always both grammatical and meaningful. (2007: 10)

An example of such a ‘repair’ is provided by Chiari herself in the following substitution: ‘*rendere flessibile il patto* (“make an agreement flexible”) > *rendere possibile il patto* (“make an agreement possible”)’ (5). In this case, it is easy for the transcriber thus to pass over the import of the semantic difference by reason of the sound similarity of the two words *flessibile* and *possibile*. Chiari has concluded that the uncovering of changes (*errors* in her terminology) made by transcribers in research reports should be used as a teaching tool in the training of transcribers. We too consider this an important pedagogical device.

It should be emphasized that the changes (or errors) observed in transcripts are not primarily due to careless transcription but to the fact that transcribing is a highly unusual way of using language, often quite conflictual with respect to both one’s everyday habits of spoken language use and one’s schooling regarding proper written usage. More specifically, in everyday spoken discourse, listeners must seek out the *gist* of a message for their own purposes rather than attending carefully to the individual words, whereas in transcription the sequencing of sounds articulated by a speaker must be assessed as objectively and as accurately as possible. But what finds its way into a written transcript is not simply a matter of careful listening; it also involves decision processes which derive from implicit theories, goals and convictions. For example, Tilley (2003) has dedicated an entire article to the problems one inexperienced transcriber has had with the task of transcribing focus group interviews including five participants and two interviewers; in this instance, the source of the difficulty was clearly the quality of the recording and in particular the simultaneous speaking of participants.

In *Example 4*, taken from the movie *A Month at the Lake* (Fox and Irvin, 1995), a combination of rapid articulation and poor acoustic quality in the original recording led the transcribers (ourselves: Transcriber’s Best Guess) to an absurdly irrelevant guess which could be disambiguated only by persistent

repeated listening sessions on the part of both transcribers together (Original Recording):

Example 4

<u>Transcriber's Best Guess</u>	Miss Bentley	I've always some (?)cakes and cheese
<u>Original Recording</u>	Miss Bentley	Miss Beaumont is hardly antique

The task of transcribing may contradict over-learned habits regarding the use of well-formed structure in written language. Our own transcription research has shown that untrained transcribers frequently use self-instruction which contradicts the experimental instructions in order to produce correct written language use (O'Connell and Kowal, 1994: 129).

The Need to Train Transcribers

The consequence of all of this is that the task of producing even transcripts limited to the words spoken necessitates some training in order to avoid the transcripts becoming more a self-revelation of the transcriber than a record of the interlocutors' spoken discourse. In addition, such training should involve reliability checks by way of having several transcribers work independently on the same excerpt of spoken discourse and then comparing their transcripts so as to verify both their validity and reliability.

VARIOUS COMPONENTS OF VOCAL BEHAVIOUR

The Verbal Component

Typically, in transcripts of spoken dialogue, the words spoken constitute the core units of a transcript. Although this sounds like a straightforward task, in fact it is not. Before even turning to the different ways of putting spoken words on paper or on the screen, the question arises: *What is considered by the transcriber to be a word?*

Our own research mentioned above (O'Connell and Kowal, 1994) as well as Chiari's (2007) has shown that transcribers tend to delete parts of utterances which they may either fail to hear or hear but not consider as words to be noted in a transcript. These would include primarily fillers such as *um* and *uh*, repetitions of words which are not syntactically integrated (e.g. *the the child*), other varieties of haltingly produced spontaneous speech (e.g. *after he uh because he*), and also a variety of interjections of both the conventional (e.g. *gee*) and the non-conventional type (e.g. *oosh*). In recent linguistic, psycholinguistic and sociolinguistic research on dialogical interaction, the functional importance of such segments is typically acknowledged and consequently they are carefully noted in transcripts. But for the transcriber him- or herself, these elements may constitute a source of confusion for several reasons, among them the following:

- They typically do not occur in well-formed written text.
- They may be considered flaws in 'good rhetoric' and therefore not worthy to be written down.
- In spontaneous, casual dialogue, they are often articulated rapidly and at a lower pitch than the surrounding speech and are therefore difficult to hear.
- Their sequential occurrence in a chain of words may not be easy to ascertain perceptually.
- Orthographic representation may be difficult.

In other words, the inexperienced transcriber may either use his or her everyday habits of filtering them out without noticing that he or she is doing so, or consider them 'bad speech' and therefore deliberately exclude them from a transcript. Such exclusion, however, may lead to the loss of information crucial for purposes of interpretation. The various ways of transcribing the verbal component are sequenced in the following four paragraphs incrementally according to their approximation of phonetic accuracy.

Standard Orthography

The words can be represented in *standard orthography*, that is in the spelling given

to them in a standard dictionary of the language. Deviations from standard pronunciation by a speaker are thereby lost. Optional variations in orthography, for example British *-our* and American *-or*, should be used appropriately and consistently. *Example 5*, taken from the movie *African Queen* (Spiegel, Woolf and Huston, 1951), is presented here in both standard orthography and in our own notation (adapted from O'Connell and Kowal, 2012):

Example 5

<u>Standard Orthography</u>	Charlie	oh Miss it is not your property
<u>Our Notation</u>	Charlie	oh miss it ain't your proputy

In *Example 5*, the standardization includes a replacement of the contraction *ain't* by *is not*, the spelling of *proputy* changed to *property*, and the initial capitalization of the address term *Miss*.

Literary Transcription

Another mode of transcribing the verbal component is by way of *literary transcription*. It constitutes part of Ehlich's (1993) originally German transcription system referred to by the acronym *HIAT* (*Halb interpretative Arbeitstranskriptionen*). A literary transcription of the words spoken takes account of deviations in pronunciation whereas *standard orthography* does not. According to Ehlich (1993: 126), this method allows for 'systematic departures from the standard orthography rendering of an item but in a manner that is meaningful to someone familiar with the orthographic system as a whole'. *Example 6*, from the movie *African Queen* (Spiegel, Woolf and Huston, 1951), provides a literary transcription of Charlie Allnut's response to Rose Sayer's comment that her brother, the reverend, has been killed by soldiers (adapted from O'Connell and Kowal, 2012). For comparative purposes, we have also included a version notated in standard orthography:

Example 6

<u>Literary Transcription</u>	Charlie	oh well now ain't that awful if they'd up 'n shoot a reverend couldn't do 'em a bit of harm then
<u>Standard Orthography</u>	Charlie	oh well now is that not awful if they would up and shoot a reverend could not do them a bit of harm then

Eye Dialect

The method of transcribing words in *eye dialect* is used especially in conversation analysis. It entails an even greater amount of deviation from standard orthography in the attempt to represent in a pseudo-phonetic way how words have actually been pronounced. In the following example of eye dialect, taken from Schegloff (1984: 288), only the words spoken are included; underlining and punctuation are deliberately left out. We have added a version notated in standard orthography for comparative purposes:

Example 7

<u>Eye Dialect</u>	Curt:	I heard Little wz makin um was makin frames'n sendin 'm t'California
<u>Standard Orthography</u>	Curt:	I heard Little was making um was making frames and sending them to California

The difference between literary transcription and eye dialect is a matter of degree; *Examples 6* and *7* illustrate this relativity quite well. But the difference between eye dialect and standard orthography is considerable. It is of interest that the *eye dialect* method has been criticized for its poor readability, inconsistency and wrong phonetics (Edwards, 1992: 368). In addition, Gumperz and Berenz (1993: 96f.) have argued that 'eye dialect tends to trivialize participants' utterances by conjuring up

pejorative stereotypes'. Our readers may wish to ask themselves whether they themselves experience the eye dialect in *Example 7* as trivializing Curt's utterance.

Phonetic Transcription

Phonetic transcriptions by means of the International Phonetic Alphabet (IPA) entail a written representation of phonetic categories sequentially realized in a corpus of spoken discourse. According to Ehlich (1993: 125), 'phonetic transcriptions aim at one-to-one relationships between (a) graphemes and (b) phonetic units and other characteristics of the spoken language'. Although the IPA is well suited for detailed transcripts used by linguists, it is seldom used by social scientists in qualitative research. The reason for this unpopularity of the IPA is partly the onerous training required for transcribing and reading it. Its complexity also makes its use subject to frequent errors.

The Prosodic Component

This component specifies *how* the words are spoken in terms of the characteristics of pitch, loudness and duration. But it should be noted that the terms *emphasis* and *stress* both subsume one or more of these characteristics indiscriminately. The characteristics are also referred to as suprasegmentals insofar as the sequential segments are supplemented by the additional notation of diacritical marks. This is typically done by adding discrete graphic units (e.g. the question mark in *Example 8* below), by super- or subimposing diacritical marks (e.g. the acute and grave accents in Chafe (1995: 58) given above and the underlining in *Example 12* below), or by changing the sequential segments themselves (e.g. RACHEL ROBERTS in *Example 9* below). While the verbal component is typically assessed by listening to an audio recording repeatedly, the various suprasegmental characteristics are more difficult to assess perceptually due to the limitations of the human auditory system. Some researchers have insisted on the perceptual assessment of these

characteristics on the ground that they aim at transcribing what the participants themselves perceive in a dialogical interaction. Others have pointed to the necessity of using instrumental measurement precisely because of the unreliability of the human ear and the correlative difficulty of transcribing from the perspective of the participants in a conversation.

Pitch

In basic transcripts, notation of the prosodic component may be relevant to disambiguate syntactic features of an utterance. *Example 8*, from the movie *Houseboat* (Rose and Shavelson, 1958), occurs in a scene where the father is talking to his young son, who is fishing. A question mark is used here as a prosodic notation of raised pitch in order to identify the utterance as a question rather than as an imperative (adapted from O'Connell and Kowal, 2012):

Example 8

Father catch anything?

As Kowal and O'Connell (2003: 100) have shown in an analysis of five German-language and three English-language transcription systems, rising intonation has been notated in these systems in several different ways: as +, as ´, as †, or as ?. We have chosen the question mark because it is the common notation sign for written text and therefore is the easiest for inexperienced transcribers and transcript readers to use.

Loudness

Another prosodic notation is related to variations in loudness. Unfortunately, the concept of *stress* is frequently made synonymous in the archival literature with the concept of loudness; but there are many ways of accomplishing stress other than loudness, for example the very opposite of loudness, namely whispering. *Example 9*, taken from Atkinson and Heritage (1984: xii), uses capital letters to indicate 'an utterance, or part thereof, that is spoken much louder than the surrounding talk':

Example 9

Announcer: an the winner: ↓iz:s (1.4)
RACHEL ROBERTS for Y↑ANKS

Duration

Still another basic prosodic characteristic that is included in all current transcription systems is the temporal organization of utterances. It includes both *ontime*, that is the duration of utterances uninterrupted by pauses, and *offtime*, the duration of pauses. Whereas *pause duration* is consistently considered in current transcription systems, variation in *ontime* is only occasionally included. The following example is taken from the movie *Unforgiven* (Eastwood, 1992). It is part of a conversation between Delilah, a prostitute, and Bill Munny, a gunman. In order to emphasize the importance of pause durations, in *Examples 10a* and *10b* we provide our own transcription without and with pause notation (see O'Connell and Kowal, 2012: 125). All pauses were measured instrumentally to a cut-off point of 0.10 seconds by use of the PRAAT software, because the research of O'Connell and Kowal (2008: 105f., for a summary) has shown that the perceptual assessment of pauses may lack both reliability and validity. Pause duration in *Example 10b* is given in parentheses:

Example 10a

Delilah ... your friends they been takin'
advances on the payment
Bill advances
Delilah free ones
Bill free ones
Delilah Alice and Silky been givin' them free
ones
Bill oo I see
Delilah would you like a free one
Bill no I I guess not ...

Example 10b

Delilah ... your friends (1.40) they been
takin' advances on the payment
(1.78)

Bill advances
(2.64)
Delilah free ones
(1.46)
Bill free ones
(1.31)
Delilah Alice and Silky been givin' them free ones
(2.12)
Bill oo I see
(4.33)
Delilah would you like a free one
(8.77)
Bill no I (1.38) I guess not ...

Notation of pauses in *Example 10b* discloses the unusually slow pace of this conversational interaction as well as the thoughtful reflections identified by the long pauses. Note that turn-taking pauses between speakers are on a separate line, whereas pauses within the turn of a speaker are on the same line as his or her words. The assumption is that the former are shared by both participants, whereas the latter may be ascribed to the current speaker. In addition, these conventions facilitate the separate analyses of the two types of pauses.

The Paralinguistic Component

Vocal features occurring during speaking but not as part of the linguistic system are referred to as paralinguistic. They include audible breathing, crying, aspiration and laughter. Paralinguistic features may entail separate segments, or they may occur as suprasegmental additions to verbal segments. In both cases, they are not easy to transcribe and are therefore typically described in or omitted from transcripts in qualitative research.

In our own studies of laughter in TV interviews with Hillary and Bill Clinton, respectively (O'Connell and Kowal, 2004; 2005), and in the film *The Third Man* (O'Connell and Kowal, 2006), we have developed the following notation conventions: so-called *ha-ha laughter* was transcribed by an approximation to the number and phonetic constitution of laughter syllables; so-called *overlaid laughter*,

that is laughter that occurs as overlay on spoken-word syllables, was transcribed by underlining those parts of an utterance which were produced laughingly and with occasional alteration and/or addition of syllables. *Example 11* is taken from an interview of Bill Clinton (BC) by Charlie Rose (CR) on the occasion of the publication of Clinton's memoir *My Life* (O'Connell and Kowal, 2005: 286):

Example 11

- CR well there was also this you were gettin' beat up so bad at home that you were anxious to get to the office
- BC that's right I said that uh yeah that's ri-hi-hi-hight HU HU HU HU I probably was more attentive to my work for several mo-honths just because I didn't want to have to attend to anything else

NOTATION SYSTEMS FOR TRANSCRIPTION

In the spirit of the critical approach to transcription in this chapter, Chafe (1995: 55) has stated 'that any transcription system is a theory of what is significant about language' (see also Ochs, 1979), and we might add: about paralinguistic, extralinguistic and non-linguistic components of communicative behaviour. With this basic fact in mind, we will briefly present a selection of transcription systems.

Among the most common transcription systems in use today are the Jeffersonian *Transcript Notation*, developed in the context of conversation analysis (CA) (Atkinson and Heritage, 1984; see Toerien, Chapter 22, this volume); the *Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)*, developed by Selting et al. (2009) and translated and adapted for English by Couper-Kuhlen and Barth-Weingarten (2011); *Discourse Transcription (DT)*, developed by DuBois et al. (1993); and *HIAT* (Ehlich, 1993), an acronym for the German *Halbinterpretative Arbeitstranskriptionen*. For a more detailed summary of the Jeffersonian *Transcript Notation*, *DT* and *HIAT* see O'Connell and Kowal (2009).

The Jeffersonian Transcript Notation

According to ten Have (2007: 95), 'the basic system was devised by Gail Jefferson', but 'there is not *one* clearly defined, canonical way of making and formatting CA transcriptions'. The canonical reference to the system is Atkinson and Heritage (1984: ix–xvi). The main purpose of Transcript Notation is to represent the sequential characteristics of spoken interaction. It allows for the notation of the words spoken, the sounds uttered, overlaps in speaking of two or more participants, and various prosodic features (e.g. pauses, tempo, stress and volume). In addition, it may be used to transcribe laughter (e.g. Jefferson, 1979), applause and a variety of extralinguistic behaviours (e.g. gaze direction). The following example from Schegloff (1984: 288) is identical with *Example 7* above but includes suprasegmental notation signs (underlining):

Example 12

Curt: I heard Little wz makin um, was makin frames 'n sendin 'm t' Californ^{ia}.

According to CA transcript notation (Atkinson and Heritage, 1984: xif.), underlining of segments indicates emphasis, a comma indicates continuing intonation and a period (full stop) indicates a stopping fall in tone.

Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)

GAT 2 has been developed over a period of more than 10 years, originally for the analysis of German-language data, to be used in particular in the context of conversation- and discourse-analytic research (see Toerien, Chapter 22, and Willig Chapter 23, this volume). Its emphasis is on the notation of 'the wording and prosody of natural everyday talk-in-interaction' and it is of interest for both 'the compilation of working transcripts ... for research purposes and for

transcripts in linguistic publications' (Couper-Kuhlen and Barth-Weingarten, 2011: 2). Its main asset for qualitative research is the fact that it is 'easily accessible for novices to transcription' (3) because it offers rules for the production of a minimal transcript 'sufficient for a range of purposes in the social sciences (such as content analysis in interviews)' (7). Couper-Kuhlen and Barth-Weingarten have provided a detailed account for notating the following characteristics in a *minimal* transcript: segments and wording, sequential structure (e.g. overlaps and simultaneous speech, pausing), other segmental transcription conventions (e.g. hesitation markers, laughter), non-verbal vocal actions and events (e.g. sniffs, sighs), and intelligibility (e.g. assumed or uncertain wording). All characteristics are documented with numerous examples. In case more detailed notation is necessary, the researcher may turn to the *basic* (see 18ff.) or to the *fine* transcript (see 25ff.).

Discourse Transcription (DT)

Du Bois et al. (1993: 45) have defined DT 'as the process of creating a written representation of a speech event so as to make it accessible to discourse research'. They have developed DT in a top-down manner on the basis of transcription design principles with the goal of developing a system that consists of good, accessible, robust, economical and adaptable notation conventions. The system uses standard orthography for the verbal component and most of the notations represent suprasegmental characteristics. For the sake of notation adaptable to different research purposes, Du Bois et al.'s goal of adaptability implies: '*Allow for seamless transition between degrees of delicacy*' (94). An example of a rather narrow transcript is given below (Du Bois, 1991: 77):

Example 13

L: . . . But `they never `figured ^out what he had?

The double period represents a short pause, the grave (`) and caret (^) accents represent

the secondary and primary accent, and the question mark represents appeal.

HIAT

The acronym *HIAT* may be translated into English as *semi-interpretative working transcription*. The term *interpretative* is meant to emphasize the transcriber's role in structuring the spoken corpus by way of both segmentation and commentary. Peculiar to Ehlich's (1993: 125) notation system is the arrangement of speakers' contributions in '*score notation*' analogous to musical score: 'Semiotic events arrayed horizontally on a line follow each other in time, whereas events on the same vertical axis represent simultaneous acoustic events' (129). *Example 14* provides part of Ehlich's (1993: 130) longer example of score notation:

Example 14

Mi: ... bottom. Pardon? Hewers.
In: Uh/hewers – did you use that term, too? Hewers.

TRANSCRIBING TURN-TAKING

All four transcription systems presented above include notations for the sequential organization of successive turns in dialogical interaction. In order to emphasize the similarities and differences among the systems, we present below a brief fictitious example of turn-taking between two participants (A and B), notated in standard orthography and without prosodic notation, but transcribed according to the different systems. Basically, there are three different modalities of turn-taking that might be noted: (1) with a measurable pause between two turns, (2) without a pause between turns (referred to as latching), and (3) with overlapping speech. *Examples 15a–d* include all three varieties in the sequential order indicated by the numbers above:

Example 15a (Jeffersonian Transcript Notation)

A was it good
(0.5)
B I don't know=
A =come on [tell me] more
B [that's all]

Example 15b (GAT 2)

A was it good
B (0.5) I don't know=
A =come on [tell me] more
B [that's all]

Example 15c (DT)

A was it good
B .. I don't know
A (0) come on [tell me] more
B [that's all]

Example 15d (HIAT)

A was it good come on tell me more
B I don't know that's all

Turn-taking serves well to exemplify the complexity of spoken dialogue and the urgent importance of transcription appropriate for the specific purposes of a given research project. The omission of any explicit preference in *Example 15* for *a*, *b*, *c* or *d* can be considered our vote against standardization without reference to purpose.

NEW TECHNOLOGIES AND PERSPECTIVES

In their presentation of EXMARaLDA, Schmidt and Wörner (2009) have counted the following among the 'main objectives' of this computer-assisted system for research in corpus-based pragmatics: 'to pave the way for long term archiving and reuse of costly and valuable language resources (e.g. to ensure the compatibility of corpora with existing or emerging standards for digital archiving)

(566). But Hartung (2006), in the context of qualitative methods in media research, has pointed out potential problems with digitalization of data in view of the formidable changes that continue to characterize this development of technology: 'It is precisely the enormous rapidity of technical progress which makes it difficult to say anything about the future and the further development of digital data formats and the corresponding hardware. For long-term archiving it is therefore not at all simple to make the right decisions' (476; our translation). At the same time, Hartung has emphasized that empirical data in the social sciences are typically analysed only within a given project and not kept in long-term archives.

With regard to future technological perspectives of transcription, it is our position that software remains a research *tool*; the finality of a research project is antecedent to and independent of the software itself. Transcribing 'accurately and unambiguously' (MacWhinney and Wagner, 2010: 156) still remains a property of the human transcriber, not of software of any kind. However, knowledge of the various capacities of available software may indeed determine for researchers what projects can prudently be engaged.

FURTHER READING

The rationale for our selection of recommended readings is as follows. Recency in such a rapidly developing field is obviously important; the earliest of our recommendations appeared within the last two decades. But breadth of treatment is another requirement. Edwards and Lampert (1993) have engaged the field of transcription quite generally, but provide further details by the authors of the transcription systems DT and *HIAT* mentioned above. Jenks (2011) has deliberately truncated his reference list for the sake of inexperienced students, has provided examples, and has referred to the Jeffersonian Transcript Notation, to DT and to *GAT 2*. Finally, Kreuz and Riordan (2011) have

provided a concise and critical treatment of various transcription systems and have also included brief references to the transcription of child language, signed language and the language of cognitively impaired individuals.

Edwards, Jane A. and Lampert, Martin D. (eds) (1993) *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale, NJ: Lawrence Erlbaum.

Jenks, Christopher J. (2011) *Transcribing Talk and Interaction*. Amsterdam: John Benjamins.

Kreuz, Roger J. and Riordan, Monica A. (2011) 'The transcription of face-to-face interaction', in Wolfgang Bublitz and Neal R. Norrick (eds), *Foundations of Pragmatics*. Berlin: De Gruyter Mouton. pp. 657–79.

REFERENCES

Atkinson, J. Maxwell and Heritage, John (1984) 'Transcript notation', in J. Maxwell Atkinson and John Heritage (eds), *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press. pp. ix–xvi.

Aufenanger, Stefan (2006) 'Interview', in Ruth Ayaß and Jörg Bergmann (eds), *Qualitative Methoden in der Medienforschung*. Reinbek bei Hamburg: Rowohlt. pp. 97–114.

Beatty, Warren (Producer) and Penn, Arthur (Director) (1967) *Bonnie and Clyde*. USA: Warner Bros.

Bull, Peter and Mayer, Kate (1993) 'How not to answer questions in political interviews', *Political Psychology*, 14 (4): 651–66.

Chafe, Wallace (1995) 'Adequacy, user-friendliness, and practicality in transcribing', in Geoffrey Leech et al. (eds), *Spoken English on Computer: Transcription, Mark-up and Application*. Harlow: Longman. pp. 54–61.

Chiari, Isabella (2007) 'Transcribing speech: Errors in corpora and experimental settings', in Matthew Davies et al. (eds), *Proceedings of Corpus Linguistics 2007*: <http://ucrel.lancs.ac.uk/publications/CL2007/> (retrieved 10 October 2011).

Couper-Kuhlen, Elizabeth and Barth-Weingarten, Dagmar (2011) 'A system for transcribing talk-in-interaction: GAT 2'. *Gesprächsforschung – Online – Zeitschrift zur verbalen Interaktion*, 12: 1–51: www.gespraechsforschung-ozs.de (retrieved 31 October 2011).

Dittmar, Norbert (2009) *Transkription: Ein Leitfaden mit Aufgaben für Studenten, Forscher und Laien*, 3rd edition. Wiesbaden: VS Verlag für Sozialwissenschaften (1st edition, 2002).

Dresing, Thorsten and Pehl, Thorsten (2010) 'Transkription', in Günter Mey and Katja Mruck (eds), *Handbuch Qualitative Forschung in der Psychologie*. Wiesbaden: VS Verlag für Sozialwissenschaften. pp. 723–33.

Dresing, Thorsten and Pehl, Thorsten (2011) *Praxisbuch Transkription: Regelsysteme, Software und praktische Anleitungen für qualitative ForscherInnen*, 2nd edition. Marburg: www.audiotranskription.de/praxisbuch (retrieved 13 October 2011).

Du Bois, John W. (1991) 'Transcription design principles for spoken discourse research', *Pragmatics*, 1 (1): 71–106.

Du Bois, John W., Schuetze-Coburn, Stephan, Cumming, Susanna and Paolino, Danae (1993) 'Outline of discourse transcription', in Jane A. Edwards and Martin D. Lampert (eds), *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale, NJ: Lawrence Erlbaum. pp. 45–89.

Eastwood, Clint (Producer and Director) (1992) *Unforgiven*. USA: Warner Bros.

Edwards, Jane A. (1992) 'Transcription of discourse', in W. Bright (ed.), *International Encyclopedia of Linguistics*, 4 vols. Oxford: Oxford University Press. pp. 367–70.

Edwards, Jane A. and Lampert, Martin D. (eds) (1993) *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale, NJ: Lawrence Erlbaum.

Ehlich, Konrad (1993) 'HIAT: A transcription system for discourse data', in Jane A. Edwards and Martin D. Lampert (eds), *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale, NJ: Lawrence Erlbaum. pp. 123–48.

Flick, Uwe, von Kardoff, Ernst and Steinke, Ines (2004) 'What is qualitative research? An introduction to the field', in Uwe Flick et al. (eds), *A Companion to Qualitative Research*. London: Sage. pp. 3–11.

Fox, Robert (Producer) and Irvin, John (Director) (1995) *A Month by the Lake*. UK/USA: Miramax.

Frost, Nollaig (2011) *Qualitative Research Methods in Psychology: Combining Core Approaches*. Maidenhead: Open University Press.

Gumperz, John J. and Berenz, Norine (1993) 'Transcribing conversational exchanges', in Jane A. Edwards and Martin D. Lampert (eds), *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale, NJ: Lawrence Erlbaum. pp. 91–121.

Harris, Roy (2010) *Integrationism: A Very Brief Introduction*: www.royharrisonline.com/integrational_linguistics (retrieved 13 October 2011).

Hartung, Martin (2006) 'Datenaufbereitung, Transkription, Präsentation', in Ruth Ayaß and Jörg R. Bergmann (eds), *Qualitative Methoden der Medienforschung*. Reinbek bei Hamburg: Rowohlt. pp. 475–88.

- Jefferson, Gail (1979) 'A technique for inviting laughter and its subsequent acceptance/declination', in George Psathas (ed.), *Everyday Language: Studies in Ethnomethodology*. New York: Irvington. pp. 79–96.
- Jenks, Christopher J. (2011) *Transcribing Talk and Interaction*. Amsterdam: John Benjamins.
- Kelle, Udo and Erzberger, Christian (2004) 'Qualitative and quantitative methods: Not in opposition', in Uwe Flick et al. (eds), *A Companion to Qualitative Research*. London: Sage. pp. 172–7.
- Kowal, Sabine and O'Connell, Daniel C. (2003) 'Datenerhebung und Transkription', in Gert Rickheit et al. (eds), *Psycholinguistik/Psycholinguistics: Ein internationales Handbuch/An International Handbook*. Berlin: Walter de Gruyter. pp. 92–106.
- Kreuz, Roger J. and Riordan, Monica A. (2011) 'The transcription of face-to-face interaction', in Wolfgang Bublitz and Neal R. Norrick (eds), *Foundations of Pragmatics*. Berlin: De Gruyter Mouton. pp. 657–79.
- Langer, Antje (2010) 'Transkribieren – Grundlagen und Regeln', in Barbara Friebertshäuser et al. (eds), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft*. Weinheim: Juventa. pp. 515–26.
- MacWhinney, Brian and Snow, Catherine (1990) 'The Child Language Data Exchange System: An update', *Journal of Child Language*, 19: 457–72.
- MacWhinney, Brian and Wagner, Johannes (2010) 'Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository'. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 11: 154–73: www.gespraechsforschung-ozs.de (retrieved 12 October 2011).
- Michel, Jean-Baptiste et al. (2011) 'Quantitative analysis of culture using millions of digitized books', *Science*, 331: 176–82.
- Ochs, Elinor (1979) 'Transcription as theory', in Elinor Ochs and Bambi B. Schieffelin (eds), *Developmental Pragmatics*. New York: Academic Press. pp. 43–72.
- O'Connell, Daniel C. and Kowal, Sabine (1994) 'The transcriber as language user', in Guillermo Bartelt (ed.), *The Dynamics of Language Processes: Essays in Honor of Hans W. Dechert*. Tübingen: Gunter Narr. pp. 119–42.
- O'Connell, Daniel C. and Kowal, Sabine (2004) 'Hillary Clinton's laughter in media interviews'. *Pragmatics*, 14 (4): 463–78.
- O'Connell, Daniel C. and Kowal, Sabine (2005) 'Laughter in Bill Clinton's *My Life* (2004) interviews'. *Pragmatics*, 15 (2/3): 275–99.
- O'Connell, Daniel C. and Kowal, Sabine (2006) 'Laughter in the film *The Third Man*'. *Pragmatics*, 16, 2/3: 305–27.
- O'Connell, Daniel C. and Kowal, Sabine (2008) *Communicating with One Another: Toward a Psychology of Spontaneous Spoken Discourse*. New York: Springer.
- O'Connell, Daniel C. and Kowal, Sabine (2009) 'Transcription systems for spoken discourse', in Sigurd D'hondt et al. (eds), *The Pragmatics of Interaction*. Amsterdam: John Benjamins. pp. 240–54.
- O'Connell, Daniel C. and Kowal, Sabine (2012) *Dialogical Genres: Empractical and Conversational Listening and Speaking*. New York: Springer.
- Peez, Georg (2002) 'Rezension zu: Norbert Dittmar, Transkription. Ein Leitfaden mit Aufgaben für Studenten, Forscher und Laien. Opladen: Leske + Budrich 2002. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 3: 24–8: www.gespraechsforschung-ozs.de (retrieved 20 October 2011).
- Rose, Jack (Producer) and Shavelson, Melville (Director) (1958) *Houseboat*. USA: Paramount.
- Schegloff, Emanuel A. (1984) 'On some gestures' relation to talk', in J. Maxwell Atkinson and John Heritage (eds), *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press. pp. 266–96.
- Schmidt, Thomas and Wörner, Kai (2009) 'EXMARaLDA – Creating, analysing and sharing spoken language corpora for pragmatic research', *Pragmatics*, 19 (4): 565–82.
- Selting, Margret et al. (2009) 'Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)', *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 10: 353–402: www.gespraechsforschung-ozs.de (retrieved 31 October 2011).
- Spiegel, Sam and Woolf, John (Producers) and Huston, John (Director) (1951) *African Queen*. UK: Shepperton Studios.
- ten Have, Paul (2007) *Doing Conversation Analysis*, 2nd edition. Los Angeles: Sage.
- Tilley, Susan A. (2003) '"Challenging" research practices: Turning a critical lens on the work of transcription', *Qualitative Inquiry*, 9 (5): 750–73.
- Walker, Anne D. (1986) 'The verbatim record: The myth and the reality', in Sue Fisher and Alexandra D. Todd (eds), *Discourse and Institutional Authority: Medicine, Education, and Law*. Norwood, NJ: Ablex. pp. 205–22.