References and Suggested Readings


Transcribing and Transcription

Felicia Roberts
Purdue University

Transcribing is the process of representing, in written form, some stretch of lived activity. The resulting transcription provides a document that is easily perused and examined, and in a variety of institutional settings it serves as the official record of the actual proceedings. Such governmental and commercial transcripts are generally perceived as impersonal or unbiased renderings and are intended primarily as references to activity. In communication research, however, it is understood that transcribing is an analytic process, since, in actuality, a transcriber is always selecting and distilling the complexities of speech and action. Or, in the case of rendering original handwritten documents, the researcher loses the artistry of the hand that produced the original document. Transcripts are, therefore, abstract versions of verbal, vocal, bodily, and spatial activities; they embody the transcriber’s stance toward the aims of recording and studying a communication event.

Origin and Function

While researchers who study naturally occurring interaction can agree on the need to represent the participants’ talk and action as carefully as possible, choices of what and how to transcribe are driven by philosophical, theoretical, and methodological orientations to the materials being handled (Ochs 1979). For example, if the researcher is concerned primarily with the thematic content of a narrative – with what is being described or what
the general purpose of the narrative seems to be – then transcribing the verbal content may be sufficient for constructing an argument about that aspect of the discourse. If, however, the researcher is concerned with how a narrative unfolds as an interactive achievement, then the contributions (vocal and nonvocal) of co-present speakers will be necessary as well.

Although the representation of actual events in graphemic and pictorial form can be traced back millennia, the use of transcription for communication research is a relatively modern phenomenon. A variety of systems have developed for the purpose of representing spoken languages. The publication of an international phonetic alphabet (IPA) in 1888 was the first major step in allowing researchers to share materials across language groups without recourse to native orthographic systems. However, the IPA requires a level of phonetic training that is not necessarily relevant for all types of communication research. Therefore, orthographic representation of spoken language is the most widely accepted approach in the field. Furthermore, with the advent of analog and digital recording, researchers rarely try to transcribe human communication directly onto paper; the research process generally begins with electronic recordings, which are then transcribed.

In producing a transcript for analysis, researchers rely on their auditory and visual perceptual skills to render from recordings some stretch of lived interaction. As noted, this is a selective process, based on beliefs about the nature of communication and the aims of the study. Researchers make choices about the kinds of detail they will attend to in interaction and the way in which they will organize those details on the page. Whatever researchers’ orientations to their materials may be, all persons engaged in the transcription process confront the paradox that communication is both simultaneous and sequential; participants in interaction move simultaneously through time such that, while some aspects of their activity may co-occur, other aspects follow sequentially. Researchers attempt to capture this layering of simultaneity and linearity by the way that transcripts are structured on the two-dimensional page. Although both the vertical and the horizontal dimensions can represent the passage of time, the horizontal is generally reserved for an individual speaker’s contribution (biased in a left to right reading format) while the vertical (top to bottom) will represent the change of speakers over time. Conjoining these levels is accomplished with a variety of symbols to show when overlapping or simultaneous talk (or action) occurs.

**EXAMPLE**

In the following (adapted from Roberts & Robinson 2004, 397–407), the beginning of the overlap is captured with the left bracket (\[). Other symbols that capture paralinguistic features are presented in Table 1.

**Example Transcript**

Mara: Have you been there yet?
Tina: Yeah. It’s really nice.\[ =Like I \[go:
Mara: \[It’s like a hotel.
A particularly daunting task for the person transcribing is managing “the recurrent interpenetration of verbal and visual communication” (Duranti 1997, 146). A variety of approaches have been developed for transcribing the coordinated vocal and nonvocal activities of interactants. Some conceive of the transcript as a musical score; others use parallel columns; still others integrate the visual and spatial information parenthetically into the verbal stream or in the lines just below the representation of the verbal stream. The approach used is dependent on the researchers’ interests, their analysis of what is relevant to the participants, and the constraints of the media they are using to format their transcription.

### TRANSCRIPTION METHODS

For transcribing linguistic aspects of interaction, several approaches have developed in the decades since the advent of recording technologies. Although approaches to transcribing talk vary considerably in their details, they can be divided into two general categories: those that privilege supra-segmental aspects of language (such as intonation and accent units, or perceptions of rhythmic synchrony) and those that privilege speaker transition and the sequential features of talk-in-interaction. This latter category is the foundation of most contemporary transcription systems since all discourse, whether monologic or multiparty, proceeds through time and therefore develops sequentially. In the case of multiparty conversation, speaker transition will also always be relevant.

The sequentially grounded form of transcription relies, explicitly or implicitly, on a model of conversation (empirically derived) that accounts for the construction and allocation of turns (Sacks et al. 1974). In this approach to segmenting discourse, the core organizing principle is conceived of as speaker transition, with silences providing the fulcrum that moves the transcript forward. In a system for transcription initially developed by Gail Jefferson (Sacks et al. 1974), the representation of both segmental features (sounds/words) and supra-segmental features of talk (such as lengthening, loudness, and other aspects of voice quality) are included, but the transcript is not organized around these supra-segmental features. The range and nature of details in this “Jeffersonian” tradition (e.g., from tokens of laughter to micro-pauses) has allowed research on conversation to move from analysis of surface-level semantic content to the analysis of underlying mechanics and structures of coordinated action. Some of these discoveries might not have been otherwise recoverable from transcription of words alone.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question mark (?)</td>
<td>Rising intonation</td>
</tr>
<tr>
<td>Period (.)</td>
<td>Falling intonation</td>
</tr>
<tr>
<td>Equal sign (=)</td>
<td>Talk that is quickly latched</td>
</tr>
<tr>
<td>Colon (:)</td>
<td>Sound that is stretched out</td>
</tr>
</tbody>
</table>

A table showing some transcription symbols and what they indicate.
Transcription systems that are intentionally designed around supra-segmental features are often crafted on the basis of a sequential understanding of activities, but they bring an additional interest in the music or rhythm of the speech and its relationship to communication. One prominent approach addresses intonation and action units (also known as “tone groups” or “tone units”). These are conceived of as “units of considerable cognitive significance” (Chafe 1993, 33), which fall under one prosodic contour. In this approach, the tonal unit has a functional significance that may transcend syntactic and semantic boundaries. Such units are discerned by the transcriber’s perception of cues such as phrase-final lengthening, phrase-initial pitch resetting, inter- or intra-utterance pauses, and so on; generally these cues are objectively available in the acoustic signal. Another elaborated, though less widely used, approach is one that captures the rhythmic synchrony in human communication. Here, rhythm is not something objectively measurable in the acoustic information, but is perceived by the transcriber from patterns of acoustic cues. Therefore, “rhythmic transcriptions require the special skill of analytic hearing – a property they share with narrow phonetic transcriptions” (Auer et al. 1999, 35).

RELIABILITY

Because the process of transcription is essentially one of perception, selection, and representation, questions can arise over the reliability of transcription practices. Researchers engaged in projects that require transcription must continually work in conjunction with the recorded data to ensure a close relationship between the recordings and the renderings. In addition, researchers often rely on a consensus process, comparing and contrasting their transcriptions with the help of colleagues. This ensures some level of agreement across the perceptions of multiple listeners. Finally, structured assessments of reliability are possible, which can ensure that transcribers are sufficiently converging on the perception and representation of phenomena. This provides some confidence in both the transcription process and its product (Roberts & Robinson 2004).

Although no transcript is ever truly “finished” (in the sense that refinements or new angles of interest can emerge), at some point the researcher makes the transcript accessible to a wider audience for presentation of findings in public meetings or in publications. This “final step” sometimes leads to simplification of the transcript for general consumption or, for those working in languages not known to their audience, the transcription becomes more complex due to the need for translation. Translation poses an additional challenge in that the transcribed activity is now an additional level removed from the original recording. Researchers can provide just the translation, or present the original and the translation in parallel columns, or, in the interest of providing the greatest access to the participants’ actual talk, one common practice is to provide, in three parallel horizontal lines, first a transcription of the original talk (usually in romanized script), then a word-by-word translation underneath, including the identification of grammatical morphemes, and finally a third line which provides an idiomatic translation into the language of the presentation. Following is an example of talk transcribed by Moerman (1988, 125) for a study of Lue-Thai interaction. The Lue-Thai orthography has been simplified here for ease of presentation:
In the future, there is always the possibility that machines will be able to handle the conversion from acoustic signal to typewritten transcript. The value of such technology will be debated and its use will vary according to the analytic orientations of the researchers. Perhaps more promising, and useful for presenting and sharing findings, is the proliferation of sophisticated presentation software as well as the possibility of including electronic media (copies of audio/video recordings) in serialized print publications (e.g., *Journal of Communication*, vol. 52, 2002). More widespread is the incorporation of URL addresses to link print readers to audiovisual media through the Internet. Having access to researchers’ recordings of lived events allows audiences to experience the data, and keeps the transcript closely linked to that data.

SEE ALSO: ► Action-Implicative Discourse Analysis ► Conversation Analysis ► Gaze in Interaction ► Interactional Sociolinguistics ► Microethnography ► Reliability ► Voice, Prosody, and Laughter

**References and Suggested Readings**


